Investigating French Interference in Algerian Students' English-Arabic/Arabic-English Translations of Collocations

Souhaila Messaoudi

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

The University of Leeds

Faculty of Arts

School of languages, Cultures and Societies

January, 2023

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

Dedication

To myself

Acknowledgements

First and foremost, I want to express my gratitude to Allah Almighty for granting me the strength, knowledge, ability, and opportunity to complete this task. Without His blessings, this achievement would not have been possible.

It has been a pleasure working on this thesis with the support and guidance of a multitude of people. In appreciation of the patient guidance and positive encouragement provided by Dr Claire Brierley and Professor James Dickins, my research supervisors, I would like to express my sincere gratitude to them. Neither developing the theoretical nor the practical parts of this interdisciplinary study would have been possible without their expertise which was shared at various stages of this work.

Further, I would like to extend my deepest appreciation and gratitude to Prof. Janet Watson for reading my first draft of this thesis, which helped me finalise this work.

Special thanks go to Mrs Karen Priestley, Doctoral College Officer, for her cooperation and administrative help throughout the PhD journey.

I would like to thank my family for their undying love and support. You were there for me when I most needed support.

Finally, I owe a special gratitude to my UK friends: Farhana, Karishma, Adaeze Ohuoba, and Dr Sookhoe, who were there for me in my darkest moments. I am and will always be highly indebted for your help.

Abstract

This study investigates the interference of French as Second Language (SL) on English as a Foreign Language (FL) among 89 Algerian translation students in their English-Arabic/Arabic-English translations of collocational false friends (*adjective+noun* collocations) which is rarely explored in literature.

It investigates how the 89 Algerian students: (i) render into Arabic English collocations involving adjectives which are themselves false friends with corresponding French adjectives; and (ii) translate Arabic collocations, the nodes of which are synonymous nouns in Arabic to English nouns that are false friends with French nouns. Both the English nouns and the English adjectives, are obtained from a compiled list of false friends between English and French (Thody and Evans, 1985).

The research adopts a mixed-methods approach in which both a self-reporting questionnaire, adapted from other studies (Magno, 2009; Ahmed, 2012), and a 30-item translation exercise consisting of two parts, involving English and Arabic collocations, have been used. A two-stage collocation extraction process was used to extract both frequent and exclusive general English and Arabic collocations. The first extraction involved the analysis of the chosen 30 false friends to find their best collocate among the top ten collocates in both the Log-Likelihood (LLR) (Dunning, 1993) and Log Dice (LD) (Rychlý, 2008) score lists in the corpus linguistics toolkit Sketch Engine. The second stage was employed when agreement could not be reached between LLR and LD. In this case, the bilingual lists were used to highlight collocates that belong to any shared semantic category between collocates in English and French. This process produced (i): twenty English collocations, the nodes of which are adjectival false friends with French; and (ii) ten Arabic collocations whose focal nodes represent nouns in Arabic, synonymous with English nouns which are false friends of French nouns.

The results demonstrate that the participants adopted eight distinct strategies in the rendition of English collocations into Arabic and vice versa, the most frequent being literal translation, which, in turn, revealed instances of French interference. Interference occurred more frequently when translating from English to Arabic than vice versa, as evidenced by the better performance in translating Arabic collocations into English than in translating English collocations into Arabic. Moreover, while the reasons for Algerian students' collocational errors were largely lexical, in some cases they were grammatical, underscoring how collocation demonstrates the inseparability of lexis and grammar even if they do not contribute equally to lexical cores (Gabrielatos, 2019).

Table of contents

Dedication ... **iii**
Acknowledgements ... **v**
Abstract ... **vi**
Table of contents ... **vii**
List of Tables ... **xiv**
List of Figures ... **xvi**
List of Abbreviations ... **xviii**
Transliteration System for Standard Arabic ... **xx**
Chapter One Introduction ... **1**
1.1 Aims of the study ... 2
1.2 Research Questions ... 3
1.2.1 Sub-Questions ... 3
1.3 Instruments of the study ... 3
1.4 Methodology ... 4
1.5 Participants in the study ... 5
1.6 Originality and Objectives ... 6
1.7 Introducing the research corpora and data analysis tools ... 6
1.7.1 Concordance lines ... 7
1.7.2 Word Sketch ... 7
1.7.3 Bilingual word sketch ... 7
1.7.4 Statistical Package for Social Sciences (SPSS) ... 8
1.8 Microsoft Excel ... 8
1.9 Overview of the structure of the thesis ... 8
2 Chapter two Theoretical background ... **11**
2.1 Basic Approaches to Collocations ... 11
2.1.1 Psychological Approach ... 12
2.1.2 Phraseological Approach ... 13
2.1.3 Statistical (Corpus) Approach ... 14
2.2 Key Figures in the Study of Collocation and Related Areas ... 15
2.2.1 Palmer ... 15
2.2.2 Firth ... 16
2.2.3 Sinclair and the Neo-Firthians ... 16
2.2.4 The Phraseologists ... 16

2.3 Distinguishing Collocations from Related Phenomena 17
2.3.1 Collocations vs. Formulaic Sequences 18
2.3.2 Collocations and Formulaic Sequences vs. Multiword Expressions 19
2.3.3 Multiword Expressions vs. Compounds 22
2.3.4 Multiword Expressions and Compounds vs. Phrasal Verbs 23
2.3.5 Multiword Expressions, Compounds and Phrasal Verbs vs. Idioms 24
2.3.6 Multiword Expressions, Compounds and Phrasal Verbs Idioms vs. Proverbs 25
2.4 Works Considering both Approaches 26
2.4.1 Nesselhauf (2003, 2004, 2005) 26
2.4.2 Gyllstad (2007) 27
2.4.3 Dukali (2018) 28
2.5 Principles Relevant to the Acceptability of Collocations 30
2.5.1 Frequency 31
2.5.2 Adjacency 32
2.5.3 Semanticity 32
2.5.4 Grammaticality 33
2.5.5 Conventionality 34
2.6 Review of collocation use by Arabic- speaking learners 35
2.6.1 Zughoul and Abdul-Fattah, 2001 35
2.6.2 Ibrahim, 2003 35
2.6.3 Shehata, 2009 38
2.6.4 Boussalia ,2010 38
2.6.5 Dweik and Abu Shakra, 2010 39
2.6.6 Alsakran, 2011 40
2.6.7 Ahmed, 2012 40
2.6.8 Shammas, 2013 41
2.6.9 El-Dakhs, 2015 41
2.6.10 Hammadi, 2015 42
2.6.11 Jabak, Abdullah and Mustapha, 2016 43
2.6.12 Almaktary, 2017 43
2.6.13 Dukali, 2018 44
2.7 Relevance of the reviewed collocation studies 45
2.8 Conclusion 45

3 Chapter three Collocational false friends and translation strategies 47

3.1 Strategy vs. procedure vs. technique 47

3.2 Non-included techniques 52

3.2.1 Transposition 52

3.2.2 Compensation 53

3.2.3 Inversion 53

3.2.4 Illustration 54

3.3 Adopted strategies 54

3.3.1 Literal translation 54

3.3.2 Omission 55

3.3.3 Synonymy 55

3.3.4 Equivalence 56

3.3.5 Approximation 56

3.3.6 Substitution 56

3.3.7 Reduction 57

3.3.8 Paraphrasing 57

3.4 Collocational False Friends 58

3.5 Language transfer 61

3.5.1 Cognates 64

3.5.2 Types of Cognates 65

3.5.3 False Friends 65

3.5.4 Types of False Friends 66

3.6 English and French Learners and False Friends 69

3.7 False friends in Research investigating L2 interference in L3 learning 70

3.7.1 Martins, and Nunes, 2013 71

3.7.2 Boumali, 2010 72

3.7.3 Sadouni, 2016 72

3.8 False Friends in research related to Translation Studies 73

3.8.1 Hilu, 2018 73

3.8.2 Cela Gutiérrez, 2018 74

3.8.3 Anwar A. H. Al-Athwary 2021 74

3.8.4 Roca-Varela 2015 75

3.8.5 Soglasnova, 2018 76

3.9 Relevance and gaps in the reviewed studies 77

3.10 Conclusion 78

4 Chapter four Research methodology and procedures 79
4.1 Categorisation of False Friends 79
4.2 Collocations Automatic Extraction 82
4.3 Collocations dimensions 83
4.3.1 Raw Frequency 83
4.3.2 Exclusivity 84
4.3.3 Directionality 84
4.3.4 Compactness or proximity 85
4.4 Association measures 85
4.4.1 T- Score 86
4.4.2 Mutual Information 86
4.4.3 Log Likelihood 87
4.4.4 Log Dice 87
4.5 Outcome of the first step analysis 90
4.5.1 Heinous crime 91
4.5.2 Eventual winner 93
4.5.3 خداع بصري xidāʕ baṣarī 95
4.6 The outcome of the second step analysis 97
4.6.1 Sensible person 97
4.6.2 Jolly lady 99
4.6.3 Petulant child 100
4.6.4 Fastidious person 101
4.6.5 Rentable space 102
4.7 Results of the two-stage Collocation Extraction 103
4.8 Research paradigms 105
4.9 Mixed-methods approach 106
4.10 Research instruments design 108
4.10.1 Discrimination 108
4.10.2 Validity 109
4.10.3 Reliability 110
4.11 The setting of the study 111
4.12 The sampling technique 112
4.13 Piloting of the questionnaire and the test 113
4.14 Research ethics 114
4.15 Data collection procedures 115

4.16 Error analysis 115
4.17 Classifications of the answers based on the acceptability scale..... 116
4.17.1 Good collocations and good translation..... 117
4.17.2 Acceptable collocation and acceptable translation 118
4.17.3 Acceptable translation but unacceptable collocation 118
4.17.4 Acceptable collocation but unacceptable translation 118
4.17.5 Neither acceptable collocation nor acceptable translation..... 119
4.18 Scoring procedure..... 119
4.19 Item Analysis 120
4.20 Item Analysis Procedures 121
4.21 Conclusion..... 123
5 Chapter five results and discussion..... 124
5.1 Results of Demographic Information Questions 124
5.1.1 The Pilot Study..... 124
5.1.2 The Main Study..... 125
5.2 Results of the three main questions raised in this thesis..... 128
5.2.1 How was the overall performance by Algerian Masters' students affected by their adopted strategies in the translation task? 128
5.2.2 Did the Students Perform Equally in translating English Collocations into Arabic and Arabic collocations into English (both Tests)? 132
5.2.3 How does Algerian EFL learners' L2 affect their proficiency in translating English collocations into Arabic and Arabic collocations into English?..... 135
5.3 Results of the Raised Sub-questions in this Thesis..... 136
5.3.1 Is there a significant statistical difference between the performance of males and females in producing collocations in English?..... 137
5.3.2 Did the Students from the Different Institutions Perform Equally in the Translation Test? 138
5.3.3 Is there a statistically significant difference in the participants' scores based on their amount of exposure to English? 140
Is there a statistically significant difference in the participants scores based on their self-perceived level in Arabic, French and English? 142
5.4 Discussion of the three main research questions of this thesis 144

5.4.1 Discussion of the results of the first question of the study regarding overall performance of the students and how it was affected by the translation strategies they used 144

5.4.2 Discussion of the second question related to the performance of the Students in translating English Collocations into Arabic and Arabic collocations into English 154

5.4.3 Discussion of the third question in relation to how Algerian EFL learners' L2 affects their proficiency in translating English collocations into Arabic and Arabic collocations into English 155

5.5 Discussion of the four Sub-questions of this thesis 157

5.5.1 Discussion of the first sub-question related to how gender affected the translation test scores 157

5.5.2 Discussion of the second sub-question in relation with the difference in the participants scores based on the faculties in which they are enrolled 158

5.5.3 Discussion of the third sub-question in relation to how the performance of the students was affected by their exposure to English activities 160

5.5.4 Discussion of the fourth sub-question regarding differences in the participants' scores based on their self-perception of their levels in English, Arabic, and French 164

5.6 Conclusion 166

6 Chapter six Conclusion 168

6.1 Findings and implications 168

6.2 Challenges and reflections 171

6.3 Contributions 172

6.4 Limitations 174

6.5 Recommendations and suggestions for future studies 176

References 179

Appendix A: the questionnaire in English 190

Appendix B : the Questionnaire in Arabic 193

Appendix C: Translation Test 195

Appendix D:The automatic extraction of collocations by Log likelihood and Log Dice agreement 197

Appendix E: Applying Dickins' 2020 compositionality model to the collocations used in this thesis 226

مزاج متعكر 226

خداع بصري 226

توجهات سياسية 226

عقبة كؤود 226
احترافية عالية 226
مهارة فائقة 226
حجة دامغة 226
أعذار واهية **226**
Appendix F: Normality tests 229
Appendix G: Distribution of translations for each collocation across the four groups 230

List of Tables

Table 1: Vinay and Darbelnet's main translation strategies..... 49

Table 2 Vinay and Darbelnet's complementary translation strategies..... 50

Table 3: Baker's translation strategies for non-equivalence..... 51

Table 4: English collocations extracted by the agreement of Log Likelihood scores and Log Dice 91

Table 5: Top ten collocates for 'heinous' sorted by Log Dice 92

Table 6: Top ten collocates for 'heinous' sorted by Log Likelihood 92

Table 7: Top ten collocates for 'eventual' sorted by Log Dice 94

Table 8:Top ten collocates for 'eventual' sorted by Log Likelihood 94

Table 9: Arabic collocations extracted by agreement of log Likelihood and Log Dice..... 95

Table 10: Top ten collocates for the noun خداع xidāʕ sorted by Log Dice 96

Table 11: Top ten collocates for the noun خداع xidāʕ sorted by Log likelihood..... 96

Table 12: English collocations extracted using agreement of Log Dice and Log Likelihood and bilingual lists..... 104

Table 13: Example of score discrimination on the exposure construct between three participants 109

Table 14: Summary of the classification of the students correct and incorrect responses based on the acceptability scale 117

Table 15:Classification and examples of produced errors..... 119

Table 16: Adjectival collocations according to their easiness index 122

Table 17: Average ages and the percentage of male and female participants in the pilot study..... 125

Table 18: Average ages and percentages of male and female participants 125

Table 19: The first language spoken by the participants..... 126

Table 20: Participants taught by native English teacher..... 126

Table 21: Attended training in Algeria 126

Table 22: Attending training in native English-speaking countries 127

Table 23: Languages used for surfing 127

Table 24: Types of undergraduate degree 127

Table 25: Total scores for the translation test..... 129

Table 26: Classification of the students correct and incorrect responses at each collocation based on the acceptability scale 130

Table 27: Classification of produced errors..... 131

Table 28: Frequency of strategies used 131

Table 29 :Scores for translating Arabic collocations 132

Table 30: Scores for translating English collocations 133

Table 31: Paired samples statistics 134

Table 32: Mean scores for males and females 137

Table 33: Tests of between-subjects effects for one-way ANOVA (for institution) 138

Table 34: Mean scores for the translation test across the different institutions 139

Table 35:Cohen's d 140

Table 36: Categorical responses for self-perceived amount of exposure to English 141

Table 37: Tests of between-subjects effects for one-way ANOVA (for watching English TV) 142

Table 38: Tests of between-subjects effects for one-way ANOVA (for Chatting Online) 142

Table 39: Tests of between-subjects effects for one-way ANOVA (for Surfing in English) 142

Table 40: Categorical responses for self-perceived level in Arabic, French and English 143

Table 41: Tests of between-subjects effects for one-way ANOVA (self-perceived level in Arabic) 143

Table 42: Tests of between-subjects effects for one-way ANOVA (for self-perceived level in French) 144

Table 43: Tests of between-subjects effects for one-way ANOVA (for self-perceived level in English) 144

Table 44: Summary of the classification of the students' correct and incorrect responses based on the acceptability scale 145

Table 45: Summary percentages of students' answers to Questions 4,6,8,10 and 15 across the different institutions 160

List of Figures

Figure 1 Proposed semantic relationship between collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs (from Dickins 2020, p. 61) 18

Figure 2: Typology of multiword expressions MWEs according to semantic independence of constituents from (Dickins 2020, p. 37) 22

Figure 3: Principles for collocation acceptability 30

Figure 4: translation strategies recognised in this thesis, in relation to Vinay and Darbelnet 52

Figure 5: Relationship between lexical transfer, false friends, and collocations 64

Figure 6: The basis for categorisation of weak and strong false friends..... 68

Figure 7: The list of the twenty false friends and their semantic categorisation..... 81

Figure 8: The list of the ten false friends, their semantic categorisation, and some of their Arabic semantic equivalents 82

Figure 9 :Relationship between the collocational dimensions and the association measures 89

Figure 10: Association measures along the continuum of frequency/ exclusivity 89

Figure 11: Bilingual word sketch for 'sensible'/ *sensible* in English and French 97

Figure 12: Bilingual word sketch for 'jolly'/ *joli* in English and French 99

Figure 13: Bilingual word sketch for petulant/*pétulant* in English and French 100

Figure 14: Bilingual word sketch for 'fastidious'/*fastidieux* in English and French 101

Figure 15 Bilingual word sketch for 'rentable'/*rentable* in English and French 102

Figure 16: Classification of collocations based on difficulty level..... 123

Figure 17: Mean percentages of the participants' total scores for the translation test 129

Figure 18 Mean percentages of the participants' scores for translating Arabic collocations into English 133

Figure 19 Mean percentages of the participants' scores for translating English collocations into Arabic 134

Figure 20: L2 interference errors..... 135

Figure 21: Distribution of translations of 'sensible person' across the three different institutions 136

Figure 22 The distribution of the total translation scores across the different institutions..... 140

Figure 23: Strategies adopted by the participants in the translation task 146

Figure 24: Mean percentages of the participants based on the time spent watching English shows..... 161

Figure 25: Mean percentages of the participants based on the time spent chatting online with native English speakers..... 162

Figure 26: Mean percentages of the participants based on the time spent surfing the internet in English 163

Figure 27: Mean percentages of the participants based on their self-perceived level in Arabic 164

Figure 28: Mean percentages of the participants based on their self-perceived level in English 165

Figure 29: Mean percentages of the participants based on their self-perceived level in French..... 166

List of Abbreviations

ArTenTen 12	Arabic Web Corpus 2012
BNC	British National Corpus
CCED	Collins COBUILD English Dictionary
DALF	Diplôme Approfondi de Langue Française
DELF	Diplôme d'Études en Langue Française
EFL	English as a foreign language
EnTenTen15	English Web Corpus 2015
ESL	English as a Second Language
FL	Foreign Language
FrTenTen 12	French Web Corpus 2012
IELTS	International English Language Testing System
L1	First Language
L2/SL	Second Language
L3	Third Language
LD	Log Dice
LLD	Log Likelihood
MI	Mutual Information
OALD	Oxford Advanced Learner's Dictionary
OCD	Oxford Collocational Dictionary
POS	Part of Speech Tagging
TL	Target Language
TOEFL	Test of English as a Foreign Language
TSC	T-Score

Transliteration System for Standard Arabic

Arabic letter	ا	ب	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ك	ل	م	ن	ه	و	ي
Trans-literation	ʾ *or* ā	b	t	ṯ	j	ḥ	x	d	ḏ	r	z	s	š	ṣ	ḍ	ṭ	ẓ	ʿ	ġ	f	q	k	l	m	n	h	w *or* ū	y *or* ī

The *ḥarakāt, fatḥa, kasra* and *ḍamma* are transliterated as *a, i, u*.

A *šadda* results in a geminate (consonant written twice), except in the case of the article, which is written with "sun letters" assimilated (*aš-šams*).

An *alif* marking [aː] is transliterated as *ā*.

tāʾ marbūṭa (ة) as word-final *-a* or *-t*.

ʾalif maqṣūra (ى) appears as *ā*, rendering it indistinguishable from *alif*.

Long vowels [iː] and [uː] are transliterated as *ī* and *ū*.

The Nisba suffix appears as *-iyy*. Nunation is ignored in transliteration.

A hyphen - is used to separate morphological elements, notably the article and prepositions (cf. http://en.wikipedia.org/wiki/DIN_31635).

Chapter One Introduction

This chapter provides an overview of this thesis, which investigates Second Language (SL) transfer in Third Language (L3) learning manifested in the translation of collocational false friends[1] by Algerian first-year master's students. In exploring the translation of collocational false friends, this research falls at the intersection of three different sub-disciplines within applied linguistics, namely language transfer, corpus linguistics, and translation. More specifically, it is inspired by the previous literature on false friends, collocations, and translation strategies.

This research identifies a hybrid approach toward defining collocations (chapter two); and it creates two typologies based on the emerging patterns of the data used and the previously existing ones for both false friends and translation strategies (chapter three). The hybrid statistical-linguistic approach adopted in this thesis for collocations is operationalised not only through applying Dickins' semantic model of collocations and other related phenomena but also through the process of collocation extraction using corpus linguistics techniques and verifying the acceptability of these collocations linguistically using two native speakers' judgements (chapter four).

The way lexical transfer is defined in the literature implies that knowing all the linguistic features of words in the source language influences the production of words in the target language (TL).

Words that exhibit linguistic similarity at all levels of meaning and form in two different languages can be considered cognates/true friends. While similarities are perceived to be helpful for language learners and translators, learners have to be aware, that lexical similarities between languages can be problematic as well. If two words are cognates, their meanings or forms are the same, while if they are not, a negative transfer may occur. In this respect, Ringbom (1987, 2001) distinguishes between formal and semantic negative lexical transfer. The former relates to a circumstance in which the words' form (spelling) is transferred, as in false friends,

---

1 'Collocational false friends' occur when true cognates in different languages appear as collocational nodes and attracting different collocates that render these cognates neither translation equivalents nor partial translation equivalents. Kovář, Baisa, and Jakubíček (2016, p. 346) state that "We cannot expect that all collocations from one relation will translate only to collocations in one particular relation in the target language, as different things are expressed by different means in different languages". This thesis extends the definition of collocational false friends, in contrast to how it has been used in the literature, to include false friends being used as nodes connecting them to their collocates within the semantic and syntactic associations and therefore being non-translational equivalents.

whilst the latter refers to the semantics and lexical connections of the SL word being transmitted to the TL form, as in calques and collocations.

Translating from one language to another is not an easy task due to the existence of asymmetries between different languages at various levels such as lexis, morphology, syntax and semantics as well as culture. Collocation and false friends are two areas that prominently illustrate asymmetries between languages. Translating collocations involving false friends is then an opportunity that can provide more insights into how differences among languages manifest. They may equally lead to the identification of students' susceptibilities to produce false collocational translations. It is, therefore, the claim of this study that collocational false friends are a better tool for capturing non-positive lexical transfer in translation than false friends alone or collocations alone.

Firth coined the term 'collocates' and described the concept using the phrase "you shall know a word by the company it keeps" (1957, p. 179). Collocations play a huge role in learning a new language. It is argued that knowledge of collocations plays a role in embellishing the language and developing accuracy in using language structures. Thus, the knowledge of collocations is of great importance for both language learners and translation students. One of the biggest challenges for both new language learners and students majoring in translation is rendering collocations into the target language. This thesis centres around the difficulties that Algerian students may face when rendering, into Arabic, English collocations involving either adjectives or nouns that are themselves 'false friends' with French adjectives and nouns. The study thus deals not only with translation between two languages (Arabic and English) but also with interference from a third language (French).

This introductory chapter will briefly describe the aims, instruments of study, participants, objectives and originality, research questions, methodology along with linguistic tools and data analysis tools. In the last section of this chapter, we will describe the structure of the thesis in detail.

1.1 Aims of the study

This research has both main and secondary aims. There are three main aims of this thesis, the last of which has four secondary aims. First, the thesis aims at identifying the translation strategies that Algerian students adopt in their rendition of English-Arabic and Arabic-English collocational false friends. Second, it sheds light on the effect of French as SL in translating Arabic-English and English-Arabic collocational false friends. Third, it attempts to highlight the differences, if any, in attainment between translating Arabic-English collocations and English-Arabic collocations.

Under the third aim of this study the possible influence of different variables on students' performance, particularly exposure to the English language, gender, institution, and self-perception of the languages spoken by the participants, are taken into consideration.

1.2 Research Questions

The research questions which this thesis sets out to answer are:

1/ How is the overall performance of Algerian Masters' students affected by the strategies they adopt in the translation task?

2/ How does Algerian EFL learners' L2 affect their proficiency in translating English collocations into Arabic and Arabic collocations into English?

3/ Do the students perform equally well in translating English collocations into Arabic and Arabic collocations into English?

1.2.1 Sub-Questions

The research also attempts to answer the following sub-questions.

Is there a significant statistical difference between the performance of males and females in producing collocations in English?

Is there a significant statistical difference between the students' performance based on the institution they belong to?

Is there a statistically significant difference in the participants' scores based on their amount of exposure to English?

Is there a statistically significant difference in the participants' scores based on their self-perceived level in Arabic, French and English?

1.3 Instruments of the study

To answer the questions raised in this research, the researcher espouses a pragmatic paradigm that employs a mixed-methods approach by adopting a concurrent embedded design. The latter is particularly appropriate to my research as it involves a one-phase data collection in which the qualitative approach is embedded to seek answers to my first question while the quantitative approach is prioritized to answer the second and third questions. Both a self-reporting questionnaire (see Appendix A, translated to Arabic in B) adopted from other studies (Magno, 2009; Shehata, 2009; Ahmed 2012), and a translation test (see Appendix C) have been devised and used to collect data for this study.

The first part of the questionnaire seeks demographic information such as gender, age, the first language at home, the time at which the English language was learned, and the educational background before joining the class of Masters in translation.

In the second part, the participants are asked a combination of dichotomous questions, and Likert-scale questions with 4- and 5-point answers regarding their self-perceived level of language proficiency in Arabic, French and English. Furthermore, Open-ended questions are also often added as follow-ups through the addition of the option 'other/s' to the closed-ended questions.

The test measures the extent to which the interference of French, if there is any, influences the translation of English collocations into Arabic and vice versa and the strategies adopted in the translation. The translation test also has two parts:

1. The first part comprises English collocations involving adjectives which are false friends with corresponding French adjectives. Students are asked to render these English collocations into Arabic.

2. The second part has Arabic collocations, the nodes of which are synonymous nouns in Arabic to English nouns that are false friends with French nouns. These Arabic collocations are to be translated by the students into English.

1.4 Methodology

To carry out this research a mixed-methods approach, involving both qualitative and quantitative analyses, is to be adopted. Two research tools will be used in this study to obtain potential answers to the questions raised above.

The first tool used in this research is a questionnaire (see appendices A, and B) that serves a dual purpose. The initial step is to compile complete profiles of the subjects by gathering various demographic data. The second goal is to determine the length of exposure to English on the part of these students both inside and outside of their programme of study in a multilingual context (Arabic with its two varieties, Berber, French, and English).

The second tool used to collect data is a test (see Appendix C), the aim of which is two-fold: A/ investigating the problems that students may encounter when they translate *adjective+noun* collocations from Arabic to English; and B/ investigating the difficulties encountered by Algerian translation students when rendering English *adjective+noun* collocations into Arabic. This will entail looking at the influence of French as the adjectives and nouns used in the collocations are false friends with corresponding adjectives and nouns in French. The test will contain two main sections comprising two main tasks. The first task provides twenty English sentences each

containing underlined English *adjective+noun* English collocations to be translated into Arabic, each of these adjectives being an English false friend with a corresponding French. The second part of the translation task contains ten Arabic sentences each having an underlined *adjective+ noun* collocation in which the Arabic nouns are synonymous with English nouns that are false friends with French nouns. Students are asked to translate those *adjective+noun* collocations from Arabic to English. The collocations used in both parts of the translation test were arrived at by exploring thirty English adjectives and nouns as nodes (search words) which are false friends with their French equivalent forms. This was done using Sketch Engine, an online corpus query software tool, which allowed for searching for potential collocates for the target node. To generate a list of collocates, two functionalities from Sketch Engine were used: the collocation function (which applies statistical formulae), and bilingual word sketches.

The answers provided by the students will give some insights into the translation strategies used by the participants when rendering collocational false friends from Arabic to English and vice versa and prove/disprove the influence of French on their translations. Therefore, the test will seek potential answers to the second and third questions raised in this research. Data obtained from the test will be analysed quantitatively and qualitatively using descriptive and inferential statistics produced using the Statistical Package for the Social Sciences (SPSS) and Excel.

To identify the strategies adopted by the students in their translations, a typology involving eight translation procedures was created to fit the emerging patterns of decisions made by the participants. Similarly, based on the theoretical framework of the study, a five-point acceptability scale was developed, and the answers were categorised accordingly. After that, these answers were converted to a three-point scale for scoring purposes.

1.5 Participants in the study

A total number of 89 Master One students majoring in English translation was targeted including those involved in the pilot study. The participants were chosen randomly, based on their actual presence in their classes. The setting of this study was three main institutions: Abu AL Kassem Saadallah Algiers 2 University, the High Arab Institute of Translation, and Ahmed ben Bella Oran 2 University. From each of the above-mentioned institutions the sample of the participants taken made up approximately 40% of the population in each institution.

1.6 Originality and Objectives

Collocations have a paramount significance. Therefore, linguists have given attention to them. The right use of collocations is a clear indicator of language proficiency. Linguists typically accept that collocations must be taken into consideration when interpreting a text. Newmark (1988, p.123) has beautifully referred to the pivotal role of collocations, considering them "the nerves of the text … and lexis flesh".

This work sheds light on the difficulties of translating collocations from Arabic into English and vice versa and the translation strategies adopted by the participants. It also investigates the probable influence of French as a second language on the translation of English collocations into Arabic as a first language. The issue of translating collocations in English has been dealt with in some works with much attention being given to translating from English to Arabic. However, little has been written about collocations in Arabic. Moreover, studying the effect of the second language on translating collocations from a foreign language to the first language has never been covered to the best of our knowledge.

Thus, this work will help in filling a gap found in the literature and provide some insights into the problems regarding collocations faced by Algerian English language learners particularly those majoring in translation. The results of this research may equally help both English language and English-Arabic translation curriculum designers.

1.7 Introducing the research corpora and data analysis tools

Sketch Engine is an online corpus management tool (Kilgariff et al, 2014) that examines how words behave. It works as a web-based tool for researchers, linguists, writers, and language learners that does not require installation. Users can explore grammar patterns across any language corpus. It has a wide range of freely available corpora, allows for corpora uploads, and creates new ones as well. It has many features. Among these, this study utilized concordance lines, word sketches, and bilingual word sketches.

The three corpora used in this study are the Arabic Web Corpus 2012 (arTenTen12), the French Web corpus 2012 (FrTenTen12), and the English Web corpus (enTenTen15). The Arabic Web Corpus 2012 (arTenTen12) was crawled by SpiderLing in January 2012, and cleaned, deduplicated, and tagged by Stanford Arabic Parser in August 2015. There are 7,475,624,779 words in this corpus. The French Web corpus 2012 (FrTenTen12) was retrieved by SpiderLing in February 2012. There are over 9.8 billion words in this corpus (9,889,689,889 words). Finally, the English Web corpus (enTenTen15), retrieved by SpiderLing in November and

December of 2015, contains over 13 billion words (13,190,556,334 words). The three corpora were encoded in UTF-8. All belong to the TenTen corpus family, which is a collection of web corpora built using the same method and targeting 10+ billion words. By removing duplicate content on the internet, including slightly modified copies of the same text, and eliminating any unwanted textual information, such as advertisements, text snippets, and incomplete sentences, from the Internet, Sketch Engine software ensures the validity of this dataset.

1.7.1 Concordance lines

Sketch Engine has a concordancer with an advanced option that looks through the whole corpus and finds all the instances of the searched word or expression in context. It displays all the different forms of the word or expressions along with their immediate context. The concordance lines can be further manipulated by applying further specifications like sorting filtering or counted.

1.7.2 Word Sketch

In a word sketch, an automatically generated corpus-based account of the grammatical and collocational behaviour of a word is presented on a one-page scale (Kilgarriff et al., 2004; Baisa, Jakubíček, Kilgarriff, Kovář, and Rychlý, 2014; Kovář, Baisa, and Jakubíček, 2016). This function makes it possible to explore the collocational and grammatical network of a particular word. Typing a word in the word sketch window displays word combinations sorted by grammatical relations. This function also makes it possible to check the frequency of each collocation, or sort them by raw frequency instead of mathematical measures scores. Word sketch allows for adjustments as well such as hiding the columns of the grammatical relations that are not of interest or changing their order.

1.7.3 Bilingual word sketch

Bilingual word sketches are pairs of word sketches for two different words from different corpora. They can be generated by manually selecting node words. Users select the two words they want to compare, and they select the corpora they want to explore using the input boxes for the target language, and the target corpus (Kovář, Baisa, and Jakubíček, 2016, p. 345). Combined word sketches are therefore an efficient alternative to opening two browser windows simultaneously, with a one-word sketch in each (ibid, 2016). In the end, a merged word sketch of the two words in different languages is produced, the two-word sketches arranged next to each other with compatible relationships. It is worth noting that word sketches are not just available for parallel corpora, but can be used with any target corpus regardless of language.

1.7.4 Statistical Package for Social Sciences (SPSS)

For the purposes of data analysis, SPSS, a statistical package developed by IBM Corporation and widely used by academics and other researchers is used. A wide range of statistical tests can be conducted using this software package since it is very user-friendly. For both parametric and non-parametric statistical techniques, this software is capable of conducting comparison and correlation tests based on univariate, bivariate and multivariate analyses (Ong and Puteh, 2017, p, 18). SPSS is also used for data visualisation as its Visualisation Designer program creates a range of different graphs like boxplots. SPSS also provides solutions for data management, allowing for further specifications on the data such as selecting cases.

1.8 Microsoft Excel

Microsoft Excel is used to store, process, analyse, and present data using functions such as sorting and filtering, and conditional formatting. The students' responses have been entered into Excel and a count has been generated. Calculations have been conducted and graphs presented based on the categorization of the answers.

1.9 Overview of the structure of the thesis

This thesis is structured as follows:

Chapter One is an introductory chapter that has explained the scope of the research and the theoretical framework of the thesis. An overview of the aims of the study, its significance, research questions, related data-analysis concepts and structure is provided.

Chapter Two defines 'collocation' and highlights the boundaries between collocations and other related phenomena using Dickins' (2020) model of semantic independence for collocations constituents. It then presents a brief overview of common approaches to the definition of collocation. Following that, it argues for combining the statistical approach and the phraseological approach for identifying and producing collocations. It claims that the phenomenon of co-occurrence can be investigated by using a statistical or frequency-based approach as a starting point, backing this up with evidence from a phraseological or linguistic approach that these word combinations are native-like. The chapter reviews some related works that have adopted a hybrid approach combining the two previously stated approaches that are used in this study. Finally, the chapter reviews recent works on how Arabic-speaking learners use collocations.

A literature review of studies relevant to the theoretical background is presented in **Chapter Three**, which addresses collocations, false friends, and translation strategies as linguistic features. It also introduces a new term 'collocational false friends' and reports on previous typologies of translation strategies. Based on previous categorisations proposed by three theorists and the students' decisions in their translations of collocations, an eight-term typology of translation strategies is proposed for analysing students' translated collocations.

The chapter begins by introducing a set of translation strategies adopted for the purpose of the analysis and discussing how the terms 'strategy', 'procedure', and 'technique' are used in the literature. The second part of the chapter discusses cognates, negative transfer in language learning, and false friends between English and French, and how this can affect the translation of collocations. It accordingly introduces the notion of non-positive transfer and how false friends and collocations are merged together in this phenomenon. By reviewing previous studies that dealt with false friends, the chapter examines the relationship between false friends and collocations in the literature. A new definition of 'collocational false friends' is then proposed for the purposes of this thesis. The chapter goes on to discuss false friends, their typologies, and studies that have investigated this phenomenon in relation to third-language learning and translation within the context of second-language interference.

Chapter 4 presents a description of the research methodology. It introduces the research design and explains the necessity of a mixed-method approach for answering the thesis' main and sub-questions. In addition, it discusses data collection tools and procedures, starting from obtaining ethical approval to quantifying, to categorizing the students' responses according to the developed scale of acceptability. The chapter then introduces the adopted research paradigm which has served as an inspiration for the development of the research instruments, particularly the extraction of the collocations used in the translation test. A rigorous analysis of the two-stage automatic extraction of collocations is provided. This is done by presenting a theoretical introduction to association measures and their practical effects, followed by an introduction of the steps involved in the extraction of both the English and Arabic collocations to be used in the translation test. This is followed by description of the results found by the association measures used. Detailed information is provided about the participants in the study, the setting in which the research is conducted, the sampling technique, and piloting the questionnaire. Finally, the chapter summarises the findings of the study and discusses the acceptability scale developed and used to evaluate students' translations and collocations along with the scoring of the answers.

Chapter Five provides a thorough analysis of the collected data and the findings. Research findings are also discussed in light of their relevance to previous literature and the theoretical background. This chapter is divided into two main sections. The first section summarizes the demographic variables from the survey responses. Using quantitative and qualitative research methods, it presents the results of the three main research questions, along with the four sub-questions. The translation strategies adopted by the participants and their effects on their translation are reported on, as is the effect of French as SL on the overall performance and the difference in attainment between the English-Arabic and Arabic-English translation tasks. Inferential tests are also used to examine potential relationships between social variables and students' scores. Taking into account the results presented in the first section, the second section discusses the possible interpretations of the reported findings, which are explained and compared with the results of other similar studies.

Chapter Six concludes the study with a summary of its conclusions and implications, acknowledges its limitations and presents some suggestions for future research. It first describes its mixed-methods study results. The experience gained from the research is then discussed as well as the challenges faced. Then, limitations are considered and, finally, how the results of the thesis can be applied pedagogically and inspire future research.

2 Chapter two Theoretical background

Introduction

This chapter introduces the three most commonly used approaches in studying collocations, in order to arrive at an appropriate definition of collocation for this thesis. The first section deals with the psychological approach. The second introduces the phraseological approach. The third section introduces the statistical approach. The chapter also presents key figures and influential works in the study of collocations. It then goes on to differentiate between collocations and other related word combinations within the scope of this study and clarifies the boundaries between the term 'collocation' and other word combinations and how the term 'collocation' is used throughout the whole thesis. The subsequent section provides a rationale for the suitability of combining the phraseological and corpus-based approaches for the purposes of this study and includes a review of works that have considered both approaches. The chapter finally provides an account of some recent works about collocation use by Arabic learners.

2.1 Basic Approaches to Collocations

'Collocation' has been defined in many different ways from different perspectives. In this thesis, I will adopt the following definition. I quote this here, in order to provide a general orientation to the reader. I will, however, argue for it in more detail in subsequent sections in this chapter. According to this definition, collocation is:

> The tendency of *conventionalized lexical items* to *significantly and exclusively*[2] *re-appear* in the company of another word within specific *grammatical patterns* at a *specified proximity* in a given corpus data resulting in natural combinations.

The word 'collocation', then, refers to any co-occurrence of words which is statistically (significantly and exclusively) greater than would be predicted by their occurrence in all contexts, including all forms of formulaic expressions, multiword expressions (MWEs) and compositional phrases that may co-occur more than predicted even if they co-occur because of non-linguistic factors like real world facts

---

2 Exclusivity of occurrence is defined by Gablasova, Brezina, and McEnery (2017, p.160) as "the extent to which the two words appear solely or predominantly in each other's company, usually expressed in terms of the relationship between the number of times when they are seen together as opposed to the number of times when they are seen separately in the corpus (e.g., the Mutual Information [MI] score highlights this property)."

(an example of collocation reflecting real-world facts being the co-occurrence of the verb *sell* and the noun *house*).

This definition draws on existing definitions of 'collocation' in the literature, particularly Evert (2008), Evert and Kermes (2003); McEnery and Hardie (2012); Sinclair (1991); Nesselhauf (2005); and Pastor (2017). According to this definition, collocations conform to five principles 1. Frequency (significance and exclusivity), 2. Proximity 3. Semanticity (lexical items), 4. Grammaticality, and 5. Conventionality. While the first two principles connect to frequency-based approach, the last three relate to the phraseological approach. The inclusion of the latter can be also demonstrated through including all forms of formulaic expressions. This definition, therefore, reflects a hybrid approach of frequency and phraseology in specifying the nature of collocations. The choice made here is justified on two grounds. First, these two approaches are widely known and influential in the study on collocations, there being also some points where the two approaches overlap. Second, the two approaches are based on different principles all of which are, from my perspective, crucial to the nature of collocations. I have therefore decided to amalgamate the two approaches by taking the statistical approach as a starting point and complementing it with the phraseological approach to validate my results. Before proceeding to the explanation of these principles, I shall first introduce each approach separately and comment on its suitability for this thesis.

2.1.1 Psychological Approach

The psychological approach looks at collocations as a mental phenomenon, either in terms of storage or of retrieval and (real-time) production. According to Wulff (2008), language involves a considerable number of constructions, i.e. form/meaning combinations which demonstrate different degrees of lexical restriction, involving schematization and entrenchment. What is meant by 'schematization' here is assigning constructions to different categories depending on combinatory and semantic features. Wulff (2008) defines schematization as the degree to which constructions "are lexically specified, and how much variation they allow for in those slots that are not lexically specified" (p.17). What is meant by 'entrenchment' is conserving and storing expressions in the human mind. Entrenchment is the cornerstone of numerous cognitive approaches to language (Bybee and Hopper 2001). There are different degrees to which different constructions are represented and stored in our brains. Langacker (1987) establishes a positive relationship between the usage of language constructions and their entrenchment. He argues that the usage of some constructions facilitates the process of their storage in the mental lexicon. Therefore, the more frequently the construction is used the easier its storage becomes.

The psychological approach highlights the key idea that native speakers typically do not need much time to process and build up meaningful utterances and sentences when they speak, because they draw on readymade constructions that they store in their mental lexicon repertoire. Therefore, the process of retrieving pre-constructed phrases or so-called 'prefabricated' chunks is much easier for native speakers than for non-natives (second language learners as well as foreign language learners). This argument neither undermines nor does it eradicate the creativity of language because language, after all, is not just pre-constructed blocks to be used. Language rather involves a gamut of separate words and word combinations. Pawley and Syder (1983) claim that the number of prefabricated constructions is roughly the same as the number of single words in the mental store of language speakers. For the purposes of this research the psychological approach will not be adopted for the simple reason that no psycholinguistic experiment will be carried out in this research to decide what should be considered as collocations and what should not.

2.1.2 Phraseological Approach

The word 'phraseology' can be divided into 'phrase' and the Latin suffix 'logy' meaning 'science', thus literally meaning 'science of phrases'. 'Phraseology' has a standard non-technical sense: "The selection or arrangement of words and phrases in the expression of ideas; manner or style of expression; the particular language, terminology, or diction which characterizes a writer, work, subject, language, place, etc." (Oxford English Dictionary Online). However, it also has a technical sense, where it refers to a proposed sub-discipline within linguistics dealing with phrases (i.e. linguistic units involving more than one word). It was mainly developed in the Soviet Union and subsequently other countries of the Soviet bloc from the 1930s onwards (Kerge 2016).

Under a phraseological approach, collocations are typically defined as fuzzy phraseological units or phrasemes (word combinations) having particular grammatical frames and enjoying different degrees of fixedness. This approach can be regarded as 'significance-oriented' in that it classifies all phrasemes (word combinations) according to their syntactic variability and semantic compositionality. In this approach, all word combinations including collocations (in the sense in which this term is used in phraseology) are put along a continuum with different degrees of semantic opacity, syntactic and lexical restrictedness. This idea has been clearly articulated by Pastor (2017).

Phraseology as a (sub-)discipline is typically conceived as covering phenomena such as 'multi-word expressions', 'formulaic units' and 'phraseological units', that range from weakly idiomatic (collocations), to proverbs. Natural languages are highly phraseological (Pastor, 2017). Drawing boundaries between collocations and other

related word phenomena is at the core of the phraseological approach as it specifically deals with the analysis of phrases which can help shaping an accurate definition for collocations. These boundaries, however, seem to be vaguely drawn, which does not fit easily with other ideas regarding the relevant phenomena that are more generally accepted and applied in corpus linguistics. It is important to define clear boundaries between collocations and other word combinations to bring the phraseological approach closer to the more provable and practical definition of the statistical approach.

2.1.3 Statistical (Corpus) Approach

The statistical or distributional or frequency-based approach goes back to Firth, the father of the lexical composition approach. It claims that words' meanings are shaped by the environment in which they occur. This approach considers collocations to be lexical units co-occurring more than we would expect them to by chance within a particular span.

Firth was the first person to use 'collocation' as a technical term (although it had been used before by Palmer (1938) to refer more generally to words that come together). One of the levels of meaning for Firth is 'meaning by collocation', which stipulates that one part of a word's meaning is the likelihood of its occurrence with another word. "You shall know a word by the company it keeps" (Firth, 1957, p.11). Firth (1957) and Sinclair (1991) agree that collocations are classified according to the frequency of their occurrences. For Firth, collocations are either usual or unusual. While the former have to do with frequent combinations, the latter have to do with less frequent combinations.

Corresponding to Firth's dichotomy of usual and unusual collocations is Sinclair's dichotomy of significant and casual. The words 'dog' and 'bark' are highly likely to appear together and for this reason they are called significant collocations. Taking into consideration that Sinclair believes in Firth's lexical composition approach, grammatical words were not included in his classifications in the beginning. However, in his subsequent studies, Sinclair embraced grammar as an inseparable part of aspects of collocations. This led Sinclair to introduce another classification of collocations as 'upward' and 'downward'. The former category is associated with different grammatical word classes like adverbs, prepositions, conjunctions, and pronouns that are more recurrent than the word they collocate with. The latter comprises lexical words like verbs and nouns that are less recurrent than their collocates.

Advocates of the statistical approach (Stubbs 1996; Manning and Schuze 1999; McEnery and Hardie 2012; Baker Hardie and McEnery 2006; Evert and Kermes 2003; Evert 2008; Evert 2005) have focused on "the co-occurrence patterns observed in a

corpus data" as the defining feature of collocation (McEnery and Hardie 2012, p.123). This notion of co-occurrence patterns resulted in attributing the term 'collocation' to a great multitude of notions (n-grams, lexical bundles, multiword units). In fact collocations, are not just confined to adjacent units. A collocations group, rather, involves any two words that co-occur frequently at any fixed order with a certain proximity over a particular span.

Analyzing the co-occurrences of some words and determining a span for the analysis are not enough to tell whether their co-occurrences are significant. Significance is perceived as the mathematical evidence that the co-occurrence is not a mere coincidence (McEnery and Hardie 2012, p.125) According to Sinclair (2004) there are four crucial elements for identifying the significance of collocations: a) the total number of words in the corpus (N); b) the number of times of the node co-occurs in the corpus, on the basis that the node is the word currently under search or investigation in the text or corpus Baker, Hardie and McEnery (2006); c) the number of times the collocate of the node appears; and, 4) the number of times the node and its collocate co-occur together in the whole corpus. Different mathematical measures are used in modern corpus linguistics to measure the significance of collocations rather than the traditional method of concordance lines of Sinclair (1991). The core idea of this earlier method was to manually count the left and right collocates of the node throughout the concordance lines. The main advocate of concordance analysis are the neo-Firthians.

For the purposes of this study the statistical approach will be adopted as a starting point as it is clearer and easier to prove practically than manual counting. This can be clearly noticed through the definition of collocations in the statistical/corpus linguistic approach, which does not highlight the fuzzy boundaries between collocations and other related phenomena. Another reason for adopting the statistical approach is that this study is corpus-based and therefore the most relevant definition for collocations will be one of the statistical/corpus-based ones. This will be supplemented by the phraseological approach, which will aid the linguistic analysis by verifying the produced collocations against the boundaries distinguishing collocations and other types of phrases.

2.2 Key Figures in the Study of Collocation and Related Areas

2.2.1 Palmer

Palmer was the first person to direct attention towards the phenomenon of words coming together without establishing 'collocation' as a technical term in the field of linguistics. Palmer defines collocations as "successions of two or more words, the

meaning of which can hardly be deduced from a knowledge of their component words" (Palmer, 1938: iv).

2.2.2 Firth

For Firth, the analysis of the meaning of words has four main dimensions: orthographic, phonological, grammatical and collocational. The orthographic level involves the way a word is written. Every word is made up of a particular sequence of letters which gives it its particular form. At the phonological level each word when, according to Firth, pronounced its phonological entity is stated (cited in Ahmed, 2012, p. 17). At the grammatical level, each word has a grammatical frame that underpins its structure (verb, noun, adjective, adverb, etc.). The collocational level is another dimension used in analysing the meaning of a word. For Firth, one way of looking at the meaning of the word 'night' is its association with the word 'dark' (ibid, p.20).

2.2.3 Sinclair and the Neo-Firthians

Among the linguists who developed Firth`s approach are Halliday (1966) and Sinclair (1991). They are sometimes referred to as neo-Firthian linguists. Halliday believes that grammar does not always justify the relationship between the components of a collocation. He gives the example of *strong argument* to clarify this notion. Grammatically, there is no explanation for why 'strong' and 'argument' go together. In addition to this, this combination of words can be grammatically flexible. For example, we can say he *strongly argued* Halliday (1966. pp. 150-151).

The same idea seems to be articulated by Sinclair when he notes that grammatical frames are choices to be made, which is not the case with collocational associations. This can be shown through the likelihood of certain words recurrently appearing together. However, when it comes to grammar, we make the choice whether to choose the passive or the active voice, for example. In his subsequent studies, Sinclair, adopted a lexico-grammatical approach that embraces both grammar and lexis.

The lexical composition approach highlights the inability of grammar to account for the different structures that words fit in, and the combinatory idiosyncrasies of words. However, it has been criticized for having a circular definition for the meaning of the basic unit of collocation. Thus, just as we can think of 'dark' as constituting the collocational meaning of night, we think of 'night' to make sense of 'dark' (Ahmed 2012, p.20).

2.2.4 The Phraseologists

Among the advocates of this approach are Pawley and Syder (1983), Cowie (1998), Benson (1986), Mel'čuk (1998),Wray (2002) and, Cowie (1994; cited in Mohamed 2012) divides word combinations into more fixed combinations (idioms) and less

fixed (collocations), these two kinds of combinations constituting what are referred to as “composites”. Composites are less lengthy than a clause and have a lexical and syntactic function. The second type is formulas, which are confined to clauses and all word combinations of sentence length, having a pragmatic function. According to Maurer-Stroh (2004), composites are further classified into four groups depending on their semantic transparency and lexical and grammatical variability:

Pure idioms or frozen idioms. These do not undergo any lexical or grammatical change. They are immutable and semantically opaque, e.g. *red herring.*

Figurative idioms. These have a figurative and a literal sense and allow for slight variability as well, e.g. *green fingers*.

Restricted collocations. These have one component with literal meaning and another with one figurative meaning and there is an arbitrary combinability on one of the elements within the collocation. This restrictedness determines which element from outside the collocation would be attracted to get into the combination. An example is *dry cow*.

Open collocations. These are free combined expressions and the elements of open collocations are used in their literal senses, e.g. *thunderous applause* (Maurer-Stroh, 2004, p. 18; cited in (Mohammed 2012 p., 30).

Some phraseologists classify collocations into lexical and grammatical collocations. The former category involves open-class words like nouns, verbs, and adjectives. The latter involve functional or grammatical words like prepositions, and infinitives.

Although Benson (1986) is known to be one of the advocates of the phraseological approach, he seems to include the core notion of the Fithian approach in his definition of collocations because he adds the element of the frequent co-occurrence of words. Benson defines collocations as being recurrent loosely fixed word combinations.

When defining collocations, phraseologists focus on strict combinability, claiming that it is the demarcating feature between collocations and free combinations. Phraseologists equally stress semantic transparency and compositionality as they demarcate collocations from idioms.

2.3 Distinguishing Collocations from Related Phenomena

The relationship between collocations and other word combinations (idioms, multiword expressions, formulaic expressions, compounds, phrasal verbs, and proverbs) has proved to be blurry as the boundaries between these notions have been fuzzily drawn, even when they are drawn at all. This is due to the fact that these concepts have been multifariously defined. In the following sections, I will draw on

the ideas of various writers, but particularly Dickins (2020), who provides a proposed set of relationships for the following phenomena: collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs. The relationship between these which is proposed by Dickins can be diagrammatised as follows (from Dickins 2020, p.61).

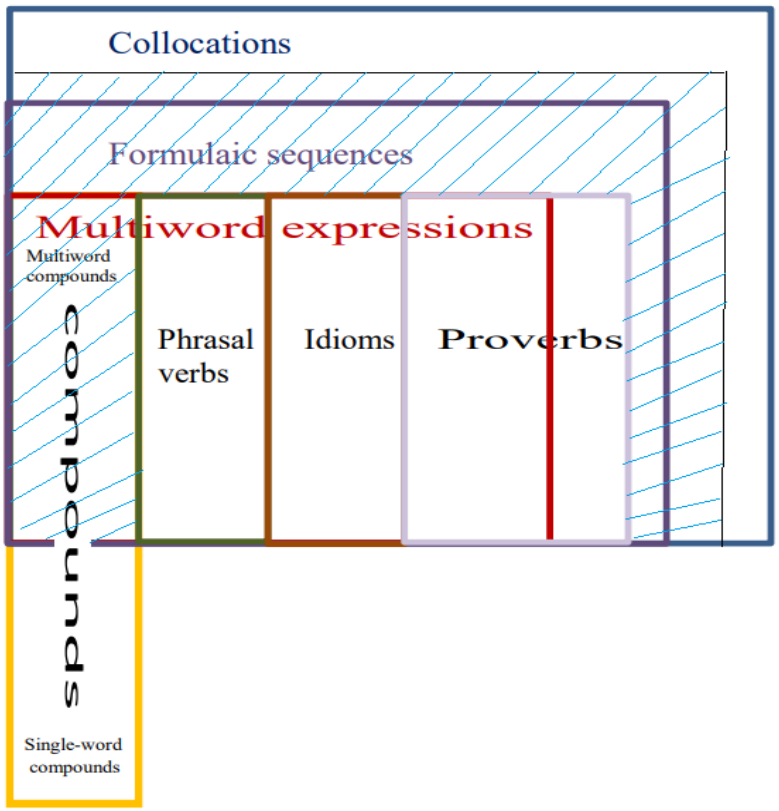

Figure 1 Proposed semantic relationship between collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs (from Dickins 2020, p. 61)

2.3.1 Collocations vs. Formulaic Sequences

Formulaic sequences, or formulaic language, have been defined in various ways. Among the most widely used definitions of formulaic sequence is that of Wray and Perkins (2000,p.1):

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

The idea of being prefabricated entails a psychological orientation which can probably only be proven by a psycholinguistic study (Dickins 2020, p. 34), and as noted above, is not adopted in this study. To overcome the restrictions of this definition, Dickins proposes a new definition in which he substitutes the psycholinguistic orientation with a corpus-based orientation, in which he states that formulaic sequences can be defined as having a higher statistical likelihood and consistency of appearance than non-formulaic collocations. Therefore, in this study, we will adopt the following definition of formulaic sequences:

> Formulaic sequences are continuous or discontinuous strings of lexical units that occur in standardised contexts and prove to have a more highly statistical and consistent appearance than non-formulaic collocations.

To illustrate this, it is highly likely that we start a letter with ‘Dear Sir/Madam’ and end it with ‘Yours faithfully/sincerely’.

Making a concordance of the word ‘wishes’ in the BNC shows that the word which is mostly associated with "wishes" in the corpus is "best". The same applies to "yours", which consistently appears with ‘sincerely’ and ‘faithfully’. The same is true for ‘ladies’ with ‘gentleman’ and ‘excuse’ and ‘me’. According to this definition, formulaic sequences are a special case of collocation.

2.3.2 Collocations and Formulaic Sequences vs. Multiword Expressions

Many attempts have been made to define multiword expression. These attempts have brought about some conflicting definitions. However, one of the commonly accepted basic criteria that is used to define multiword expressions is the dichotomy of semantic compositionality/idiomacity (Dickins, 2020; Kerge, 2016; Masini, 2005). Other researchers consider other elements in their definitions of multiword expressions like syntactic restrictedness/looseness, and pragmatic as well as statistical idiosyncracies. This addition results in narrowing the range of language constructions that are included under the umbrella term of multiword expression.

Considering the element of syntactic restrictedness/looseness, multiword expressions can range from fully frozen constructions to fully compositional and flexible combinations in terms of morpho-syntactic changes. For example, ‘kick the bucket’ is an immutable construction that neither allows morpho-syntactic variation nor permits any internal modification: i.e. we can say “He kicked the bucket” but not “the bucket is to be kicked by him” or “He kicked the bucket hard”. As far as pragmatic and statistical idiosyncrasies are concerned, the presence of two (or more) elements cannot alone make a multiword expression. They are rather considered to be defining features of formulaic expressions and collocations (Dickins, 2020, p.42).

What we remain with, therefore, in order to differentiate formulaic sequences from multiword expressions is semantic compositionality. Semantic compositionality has been defined as the situation where the overall meaning of an expression can be analysed in terms of the independent semantic contribution of its component parts and the grammatical rules used to combine them together (Wang, 2017).

One way of approaching multiword expressions is to consider how each constituent functions semantically within the expression. This can be clearly demonstrated through Dickins' analysis of multiword expressions constituents where each element of the expression either has no independent sense or has an independent sense within one context or limited contexts. What is meant by an independent sense of a word is when the word conveys a specific sense in the context in which it occurs. This may be different from the sense that it usually conveys in other more standard contexts.

According to Dickins (2020), a particular constituent word in a multiword expression may not have an independent sense. This is what he terms 'Type 1' constituent. Where none of the words in a multiword expression has an independent sense, the expression can be termed 'fully non-compositional'. A second possibility is that the particular word in question in multiword expression has an independent sense.

Under the second possibility of Multiword Expression having a word with an independent sense, there are further three sub-possibilities. The first is when a constituent has an independent sense which is just found in that particular context, giving what Dickins (2020) terms a 'Type 2' constituent. The second sub-possibility involves a constituent with an independent sense which is found in some limited contexts, giving what Dickins terms a 'Type 3' constituent. A third possibility is when a Multiword Expression constituent has an independent sense in unlimited contexts. This Dickins terms a Type 4 constituent. Dickins points out that it is possible to have, for example, a Multiword Expression which combines both Type 2 and Type 3 constituents. This category involves both a constituent which has an independent sense in that particular context and a constituent which has an independent sense in limited contexts.

An example of Type 1 constituents – and in fact a case where the Multiword Expression is fully non-compositional, can be found with pure or frozen idioms like 'kick the bucket' and 'red herring'. In these examples none of the constituents has an independent sense. Therefore, we cannot deduce the meaning of the whole expression by analysing the meaning of each constituent element. Correspondingly, we cannot make any morpho-syntactic changes to these expressions, i.e., they are neither grammatically variable nor syntactically flexible.

Examples of a Type 2 constituent, i.e. a constituent having independent sense but only in one particular context, are ‘blue’ in blueberries and ‘polar’ in ‘polar bear’ (Dickins, 2020, pp. 38-40). In this context, ‘polar’ has a different meaning from the meaning it conveys in its typical usage because ‘polar’ in this case does not mean a bear which just lives in polar regions. Rather a ‘polar bear’ is a specific species of bear, regardless of where it lives (whether in a polar region, or in a zoo, or elsewhere). The same is true for ‘blue’ in ‘blueberries’. The word ‘blue’ here, does not have its standard meaning of the ‘blue’ colour as ‘blueberries’, as can be seen from the fact that not all ‘blue berries’ are ‘blueberries’, and not all ‘blueberries’ are ‘blue berries’; e.g. unripe blueberries are green.

Unlike ‘polar’, and ‘blue’ in ‘polar bear’, and ‘blueberries’ the word ‘bear’ in ‘polar bear’ has its standard meaning which is found in unlimited other contexts. Thus, for example, ‘bear’ has the same sense in ‘brown bear’ as it does in ‘polar bear’, but also has the same sense in a phrase such as ‘that bear is dangerous’. Therefore, the word ‘bear’ in ‘polar bear’ provides a good example of a Dickins’ (2020) Type 4 constituent, i.e. a constituent of a multiword expression can be used in endless contexts to mean the same thing.

The word ‘farm’ in ‘wind farm’ (Dickins, 2020, p. 41) is an example of Dickins’ (2020) Type 3 constituent, i.e. a constituent with a specific sense, found in only limited contexts, and different from the typical sense that this word has in other contexts. In this expression, the word ‘farm’ is not used in its typical sense to mean land used to grow crops or to rear animals. Rather, it is used in this context to mean ‘array of machinery for producing energy from’ (Dickins, 2020, p. 42). The following diagram, adapted from Dickins (2020, p. 37), summarises all the cases mentioned above.

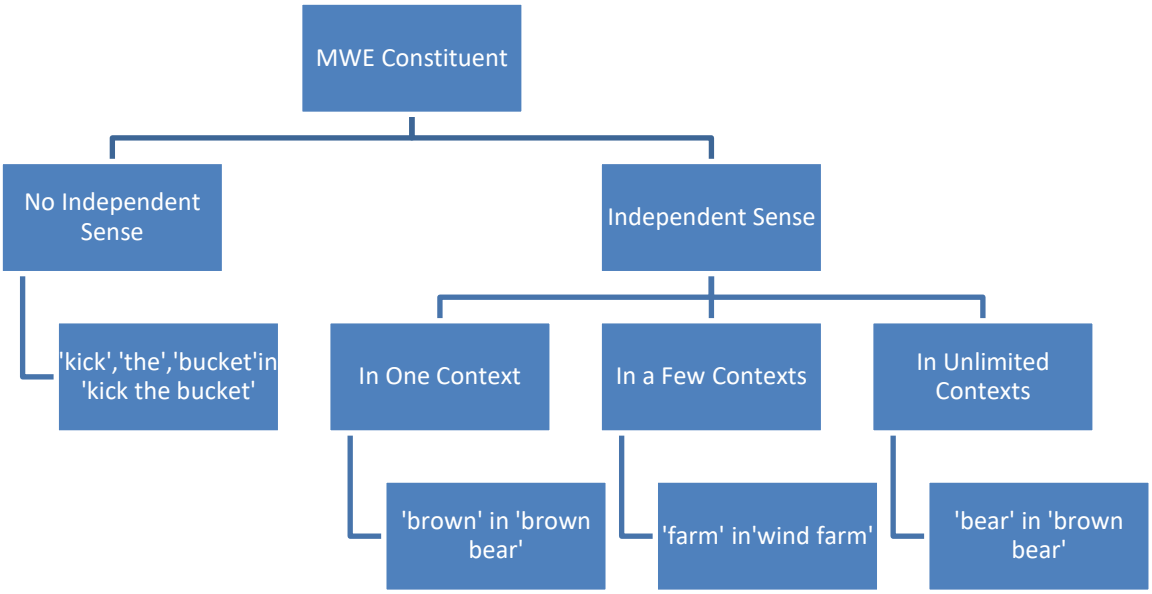

Figure 2: Typology of multiword expressions MWEs according to semantic independence of constituents from (Dickins 2020, p. 37)

2.3.3 Multiword Expressions vs. Compounds

A compound can be understood as a combination of two or more words or as a composition of two or more components that may belong to the same part of speech or may be different. 'White House', for example, consists of an adjective and a noun. Each component part of the compound can be used separately in other contexts. 'Car park' is a compound that consists of two nouns. With some exceptions, the stress falls on the first element of the compound. Although compound nominals are syntactically unchangeable, they undergo inflection for number. Therefore, compound nominals do not undergo syntactic variation and the only lexical change that they allow is inflection. Some of them allow number inflection by simply adding's' to the word end like 'car parks', to form the plural form.

Dickins highlights the difference between multiword compounds and single-word compounds. 'Blackbird' (Dickins, 2020) is a single-word compound because it consists of two free morphemes – i.e. what in other contexts can be independent words – attached together and written as one word. Multiword compounds are of many types and are a sub-type of multiword expression. None of multiword expressions, multiword compounds and single-word compounds are fully free-compositional, because in all three types not all the components have an independent sense. Dickins establishes a hyperonymy-hyponymy relationship for multiword expressions and multiword compounds: multiword compounds are a type of multiword expression.

However, single-word compounds are not to be considered a subtype of multiword expressions, simply because single-word compounds consist of one orthographic word only.

A single-word compound may have non-independent sense components, or it may have a component with an independent sense in only one context, limited contexts or unlimited contexts. ‘Ladybird’ is a single-word compound noun consisting of two components, neither of which has an independent sense. However, the whole combination, if it is analysed the same way a multiword expression is, has an independent sense.

‘Blackbird’ is an example of the second case where one constituent, ‘black’, has an independent sense but this is only found in one context. This can be seen by the fact that ‘black’ in this single-word compound does not necessarily refer to the black colour because we may have a brown blackbird (in fact female blackbirds are brown), just as we may have a white blackbird (e.g. an albino blackbird) (Dickins, 2020, p. 18). Analysing the word ‘respectful’, on the other hand, would demonstrate that the suffix ‘full’ has an independent sense in some contexts which is ‘exhibiting or having’. This sense is different from any sense which ‘full’, as a separate word has. Finally, ‘bird’ in ‘blackbird’ is to analysed like ‘bear’ in “polar bear” because it refers to specific kind of feathered animal with wings and able to fly usually, i.e. it has the sense which ‘bird’ has in unlimited other contexts.

To conclude, unlike multiword compounds, which are a type of multiword expression, single-word compounds belong to a separate group, neither being multiword expressions nor falling under collocations in general, since single-word compounds only involve a single orthographic word. However, if we consider compounds generally, we can say that they involve different word combinations, ranging from multiword expressions, formulaic expressions, to collocations and even extending to other word constructions or combinations which fall outside the scope of collocations (Dickins, 2020).

2.3.4 Multiword Expressions and Compounds vs. Phrasal Verbs

In this thesis, I shall adopt the following definition for phrasal verbs:

> Phrasal verbs may be prepositional verb constructions that involve just verbs with prepositions but they may also be verb-particle constructions which include adverbs in addition to prepositions and verbs, like ‘give up on’ (meaning ‘abandon’), ‘keep up to’ meaning ‘maintain’).

Phrasal verbs are considered to be not fully free compositional (Dickins, 2020) and therefore they are a sub-type of multiword expression separate from compounds.

Subjecting the phrasal verb “keep up” to the same analysis of semantic independence of constituents, we can deduce that neither ‘keep’ nor ‘up’ has an independent sense. The word ‘fish’ in “fish out” meaning to ‘extract’ has an independent sense in only one context because ‘fish’ here conveys the meaning of ‘extract’ or ‘find’. However, ‘out’ turns up to have an independent sense in unlimited contexts which is ‘outside [of something]’. The verb to ‘turn’ in “turn on/off/down/up” means ‘to start/stop something working’ (Dickins, 2020, pp.51-52). This is a different meaning from the usual meaning of ‘turn’ in other contexts in English.

2.3.5 Multiword Expressions, Compounds and Phrasal Verbs vs. Idioms

Idioms are of two types: decomposable and non-decomposable. They have been defined but overall in unclear terms as a “fuzzy category” (Numberg 1994) or, rather differently, as an “entity whose meaning cannot be deduced from its parts” (Mitchell 1971, p. 57) and on many occasions they are placed in a continuum along with other kinds of collocations to highlight their semantic opacity and morpho-syntactic restrictedness. Dickins (2020,p.44) defines idioms as follows:

> An idiom is a phrase that is not a compound, not a phrasal verb, is non-clausal, and is not fully free-compositional.

Dickins (2020) makes a distinction between phrasal verbs, compounds and idioms. He regards phrasal verbs and compounds as two different categories that do not intersect, which is the view adopted for the purposes of this thesis.

Idioms can be subjected to an analysis of the semantic independence of their constituents. A typical frozen idiom like “red herring” or “kick the bucket” has semantic independence only as a whole combination. However, when it comes to each constituent of these idioms, none of them has an independent sense. Conversely, ‘spill’ and ‘beans’ in the idiom “spill the beans” have independent senses because what both ‘spill’ and ‘beans’ have independent senses in this context (‘divulge’ and ‘secrets’) but what is meant by them is different from their senses in other contexts. This independence can be seen in the morpho-syntactic variability, i.e. the internal modification that this idiom allows. We can say ‘he spilled the beans’, just as we can say ‘spilling those beans’. In this case, ‘beans’ refers to secrets or confidential information, but this sense is found only in this one context. Similarly, ‘spill’, which is a polysemous verb means to give away or divulge, but only has this sense in this context.

The third example of one component having an independent sense in some contexts only is ‘mind’ meaning ‘rational faculties’, as in “lose....'s mind” (Dickins, 2020,

p.56). ‘Lose’ in this idiom has an independent sense in unlimited number of contexts. While the word ‘lose’ has slightly different senses in Cambridge Dictionary and lexico.com, all of these senses share the basic sense of having less or ceasing to have something or someone. Another example where idiom constituents with independent senses can be found in a wide variety of contexts are 'better' and 'than' in “better late than never”. Nunberg, Sag, and Wasow (1994) call idioms with components which have independent sense 'idiomatic combinations' or “idiomatically combining expressions”, while Sag et al. (2002) label them “decomposable idioms”. The idioms in which none of the constituents have an independent sense are called ‘idiomatic phrases’ by Nunberg, Sag, and Wasow (1994).

To sum up this thesis would agree with Larson (1984), Palmer (1995) and Crystal (1995), cited in Mohammed 2012) and Dickins (2020) in considering ‘collocation’ to be a generic term covering, amongst other things, idioms. When it comes to demarcating compounds from idioms Dickins suggests consulting native speakers' intuitions. Some forms can fit into both classes, compounds and idioms, like “sleeping policeman” (Dickins, 2020, p. 57). This fulfils the requirement of a compound by belonging to a specific grammatical class (nouns/nominals) besides being reckoned by native speakers as an idiom. This solution may work acceptably for compounds and idioms. Similarly, using native speakers’ intuitions to demarcate idioms and phrasal verbs may yield satisfactory results as some native speaker may classify some phrasal verbs as idioms like “take off”.

2.3.6 Multiword Expressions, Compounds and Phrasal Verbs Idioms vs. Proverbs

Language and culture have always displayed an inextricably interwoven relationship. One way of defining a particular culture is finding the ethics, and values that are encapsulated in the language of that culture. This can be particularly well achieved through looking at the proverbs in that language. Proverbs have been defined by Kuiper and Allan (1996, p. 283) as: “usually a whole sentence in length and are used as a way of morally evaluating human actions and giving advice on what to do”.

An example is “a bad penny always turns up”, referring to people who are disreputable or prodigal always returning to their former lives. In a broader sense, this proverb refers to any undesirable event that repeats itself. According to the definitions adopted in Dickins (2020), some proverbs are (a type of) multiword expression, i.e. those which are not fully free-compositional, while other proverbs are (a type of) formulaic sequences (but not multiword expressions), i.e. those proverbs which are fully free-compositional. Dickins (2020, p. 54-56) illustrates different types of proverbs as follows.

An example of a proverb which is fully free-compositional (i.e. where all the words involved have the same sense as they have in potentially unlimited other contexts) is "A good man is hard to find". Proverbs which are not fully free-compositional may be of different types. An illustration of a proverb none of whose components has an independent sense is 'one swallow does not make a summer'. By contrast, in 'forearmed' in "forewarned is forearmed", the words 'forewarned' and 'is' have the same sense which they have in potentially unlimited other contexts, while 'forearmed' has an independent sense only in this context. 'Hands' in "many hands make light work", by contrast, has an independent sense (people) in a limited number of contexts. Proverbs can also contain within them idioms.

According to Dickins 'semantic independence' model, proverbs, idioms, phrasal verb, and multiword compounds are all properly included in formulaic expressions which are in turn properly included in collocations. Therefore, collocations in their broadest sense involve all kind of formulaic expressions. Dickins' (2020) view on collocation is clearly frequency-based without setting a threshold for how frequent they should be. His analysis, however, for formulaic expressions presupposes a statistical significance for frequency in addition to conformity to syntactically flexible (continuous or discontinuous) relationships. As far as Multiword Expressions are concerned, Dickins definition draws explicitly on the notion of compositionality or semantic independence as a key definitional criterion for them.

In a narrower sense, the term 'collocation' could be used to refer only to non-formulaic collocations. In this thesis, I shall use the term 'collocation' to refer to what Dickins (2020) terms a 'statistically significant collocation' that can be analysed in terms of its elements' semantic compositionality and conform to syntactically variable relationships (syntactic coherence) which entails abiding by grammatical rules since all syntactical rules are grammatical rules. By including the two criteria semantic compositionality and syntactic relations that belong to the phraseological approach, the definition adopted in this thesis is more inclined towards a combination of both phraseological and frequency-based approaches.

2.4 Works Considering both Approaches

In the following sections, I will review a number of works which consider both the phraseological and the statistical approaches to collocations.

2.4.1 Nesselhauf (2003, 2004, 2005)

Nesselhauf regards collocations as "arbitrarily restricted lexeme combinations" (2005, p.1). Her work on collocations is primarily based on the phraseological tradition, as she uses semantic opacity for categorizing collocations into free

combination, collocations, and idioms. She draws on Howarth's (1996) investigation of *verb+noun* collocations, on the basis that this is one of the most comprehensive studies on collocations despite acknowledging the small size of his database. She also uses a corpus as a complementary method to validate her results. Nesselhauf (2004) identifies ten criteria relevant to the study of collocations: 1. frequency, 2. transparency, 3. variability, 4. grammaticality, 5. the nature of components, 6. types of components, 7. number of components, 8. whether components are separated or consecutive, 9. equality of the relationship between components, and 10. nature of the phenomenon per se. While features 5,6,7,8,9 and 10 are not covered in Dickins (2020), item 1, frequency, is the only definitional factor (i.e. part of the definition of 'collocation') in Dickins (2020). Item 2, transparency, however, can relate to Dickins (2020) notion of compositionality in the sense that fully free-compositional collocations are transparent where all the components involved have the same sense as they have in potentially unlimited other contexts. By contrast, non-fully free compositional collocations are semantically less transparent in that they can have at least one component with an independent sense in one context, in few contexts or, unlimited contexts. Additionally, variability here may relate to Dickins' (2020) notion of continuity or discontinuity in his definition of formulaic expressions. "A formulaic sequence is a collocation, whether continuous or discontinuous, which has syntactic coherence". While Dickins' model does not mention grammaticality per se, his notion of syntactic coherence may somehow be taken to refer to grammaticality given that syntax is part of grammar dealing with word order.

Nesselhauf (2003) was a corpus-based study that focused on the use of *verb+noun* collocations in the writing of German English-language learners. In order to judge the acceptability of the produced collocations, she used the following dictionaries: 1. The Oxford Advanced Learner's Dictionary (OALD, 2000), 2. Collins COBUILD English Dictionary (CCED, 1995), and 3. British National Corpus (BNC 2014), and she consulted two native speakers. The study revealed nine types of errors in producing collocations, the most common of which was the wrong choice of verb. According to Nesselhauf, the mistakes were due primarily to first language (L1) interference.

2.4.2 Gyllstad (2007)

Influenced by Nesselhauf, Gyllstad attempted to investigate Swedish learners' of English receptive knowledge of *adjective+noun* and *verb+noun* collocations. Gyllstad defines collocations as "conventionalized, recurring combinations of words". This definition draws on two main criteria, belonging to the approaches of phraseology ('conventionalised') and frequency as used in Dickins 2020 ('recurring combinations') respectively. In order to operationalize the definition, Gyllstad

developed a nine-item scale for identifying and classifying collocations. This was influenced by Nation's (2001), Nesselhauf's (2004), and Siepmann's (2005) scales of collocability. It is important to note that the above-mentioned works adopt a hybrid approach towards identifying collocations. They thus amalgamate features from the two different traditions in order to create their own scale of identification/acceptability of collocations. Gyllsad (2007) consider that there is a very large overlap between the scale proposed by Nation (2001), the approach of Nesselhauf (2004), and to a lesser extent with the questions raised by Siepman (2005) in defining collocations. Nation's (2001, 328ff) scale involves : 1. frequency of co-occurrence, 2. adjacency 3. grammatical connectedness, 4. grammatical structuredness, 5. grammatical uniqueness 6. grammatical fossilization, 7. collocational specialization 8. lexical fossilization 9. semantic opaqueness, and 10. uniqueness of meaning. Apart from frequency, which Dickins (2020)' model is centred on, grammatical structuredness may somehow relate to the definitional criterion for formulaic sequence used in Dickins' (2020) as being continuous or discontinuous having syntactic coherence.

Using Nation's scale as a starting point, Gyllsad kept the criteria of frequency and adjacency shared between Nation and Nesselhauf's and collapsed Nation's three criteria of syntactic relations into Nesselhauf's grammaticality criterion. He dismissed criterion 7, lexical specialization, which refers to the exclusivity of collocations, but preserved the last three elements of Nation: lexical fixedness, semantic opacity, and uniqueness of meaning. This last refers to whether a collocation is monosemous (has one sense) or polysemous (has more than one sense, the secondary sense(s) normally being figurative).

Gyllsad's (2007) study demonstrates that there is a positive relationship between receptive collocation knowledge and advanced learners' vocabulary size. According to Gyllsad the gap in receptive collocation knowledge between higher and lower proficiency learners has two main sources: 1 lower-proficiency learners' use of their L1 to process L2 forms; and 2. lower-proficiency learners being less exposed to the target language in comparison with higher proficiency leaners. The results also reveal that a period of 4-6 months of full-time university-level studies is not enough to yield a quantifiable increase in receptive collocation knowledge. Furthermore, there is a clear association between developing receptive collocation knowledge and learning level, overall language proficiency, and vocabulary size. This endorses language exposure as an important denominator for learning collocations.

2.4.3 Dukali (2018)

Drawing explicitly on Nesselhauf's studies, Dukali study examines the challenges faced by Libyan undergraduate English major students in using *verb+noun* and *adjective+noun* collocations. The sample of the study were fourth-year English major

students in the Department of English at Tripoli University. Data was collected through a corpus which involved the free production of a 240-word academic essay from each student produced in 45 minutes. The focus of the study was twelve verbs and twelve adjectives. The data was analysed using AntConc3.2.1 (Anthony 2007). The collocations extracted from the corpus were analysed in four different ways: 1. Eng (OCD 2009), 2. The British National Corpus (BNC), 3 Native speakers' intuitions, and 4. the acceptability survey of collocability. This last was used in order to triangulate the three previous methods. The analysis framework of collocations used in Dukali's study combines the corpus and the phraseological approaches to assess the acceptability of the collocations produced by the Libyan students. This is clearly demonstrated through the use of the British National Corpus and native speakers' intuitions, which are in turn assessed by four phraseological aspects of the language: substitutability, semantic components, grammaticality, and conventionality. Subjecting the *verb+noun* and *adjective+noun* collocations to the four above-mentioned analyses allowed the researcher to classify the collocations on a three-degree scale of acceptability: unacceptable, partially acceptable, and acceptable. Erroneous *verb+noun* and *adjective+noun* collocations were classified into three main groups: grammatical, lexical, and usage-related errors. These were broken down into sixteen subtypes of error for *verb+noun* collocations, and twelve subtypes for *adjective+noun* collocations. The results showed that grammatical errors were more frequent than lexical errors in the produced collocations. On this basis, the author recommended introducing the teaching of the whole collocational pattern rather than just its components, e.g. teaching the pre-modification and/or post-modification of nouns with elements such as articles, intensifiers, prepositions, and possessive pronouns within collocational patterns. The author's recommendation stems from Nesselhauf's (2003, p. 238) proposal to adopt a comprehensive approach in teaching collocations through teaching the entire collocational combination, and not just its constituent lexical items. The results of Al Dukali's study also reveal that register can affect word choice in producing collocations. With respect to the focus of her study, which is academic written English, non-awareness of register leads students to choose the wrong verb or adjective, notwithstanding that the resulting collocations may be acceptable in informal spoken English. Like other studies, Al Dukali's study identifies L1 interference as one of the main reasons why students produce unnatural collocations which have equivalents in colloquial Libyan Arabic. In addition to L1 interference, over-generalisation, and the use of synonyms are other reasons that resulted in producing erroneous combinations according to the author, who concludes that collocational activities should focus on these problems.

2.5 Principles Relevant to the Acceptability of Collocations

Based on the works reviewed above, I shall both define collocations and judge the produced translations for the given collocations in this thesis against five key criteria taken from both phraseology and corpus perspectives. These five criteria are believed to encapsulate the nature of collocations as a linguistic phenomenon and to help build an acceptability scale for the produced translations. I shall use the statistical approach as a point of departure and thereby include frequency and adjacency as in Nation (2001) and Gyllsad (2007) as the first criteria against which the produced translations are tested. I will include the four criteria of collocability used in Dukali's (2018) work combining semantic components and substitutability under what I term semanticity. On this basis, the following definition of acceptable collocations is proposed:

> The tendency of conventionalized lexical items to significantly and exclusively re-appear in the company of another word within specific grammatical patterns at a specified proximity in a given corpus data resulting in natural combinations.

According to this definition, producing natural word combinations requires conforming to five principles 1. frequency (significance and exclusivity), 2. proximity 3. semanticity (lexical items), 4. grammaticality, and 5. conventionality. I will discuss each of these in separate sections below.

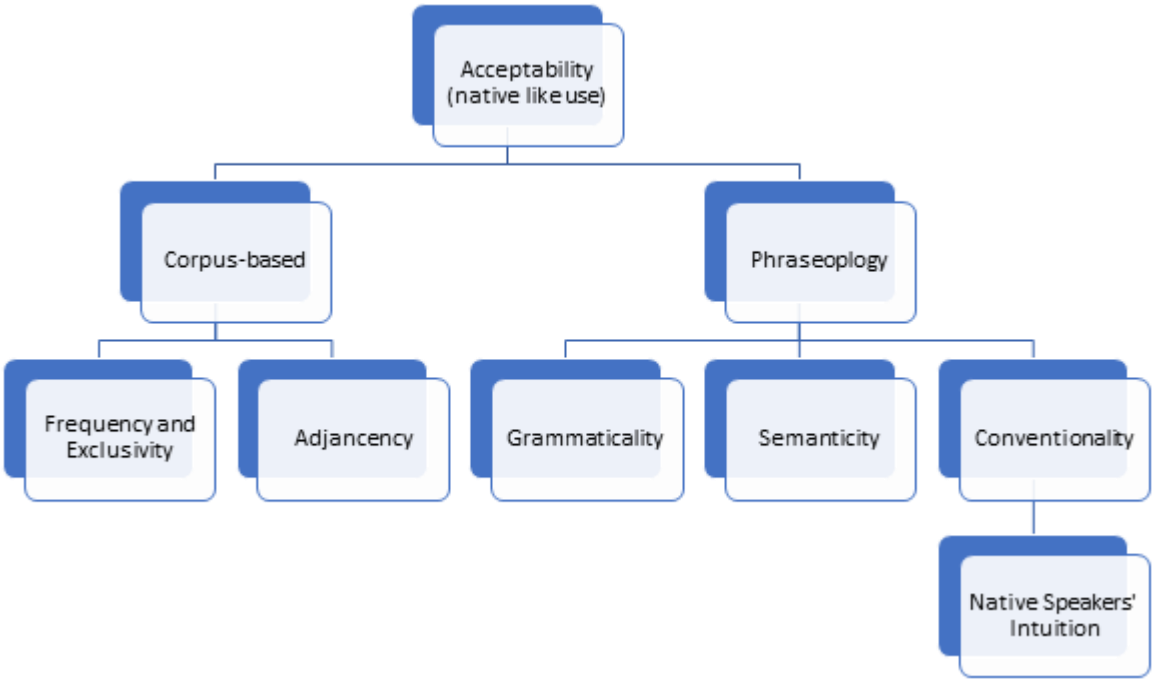

Figure 3: Principles for collocation acceptability

2.5.1 Frequency

Frequency is one of the criteria that is crucial for the operationalisation of identification and testing the collocability of collocations. Raw frequency is one of the simplest approaches for extracting and testing collocations. This approach is based on the idea that the more frequent the collocation is the more predictable in language it becomes. In corpus studies, corpora are the tool used to investigate language. In the present analysis the frequency of collocation is to be checked through one of the largest computerised corpora: English Web Corpus (enTenTen 2015).

A raw frequency-based approach search yields good results with adjacent fixed phrases. It does not, however, work with flexible combinations (Manning and Schütze1999p, 157); nor does it exclude simple function words that occur very frequently in the language. What is meant by flexible combinations is distant collocational components that co-occur within flexible grammatical patterns at different distances from each other like the distant bigram "knock" and "door" in the following sentence: "They knocked on her back door." (Manning and Schütze 1999, p. 157). Function words and the non-appropriateness for non-fixed combinations are two main areas where raw frequency searches are not effective.

Manning and Schütze (1999), suggest two ways for overcoming the limitation of the raw frequency approach in this regard. First, in relation to function words, they propose part of speech (POS) tagging of the collocates[3], i.e., all individual sets in the text should be annotated with part of speech tags. Once the tokens, i.e., words in the text, are assigned part of speech tags, collocation candidates will be filtered, and the collocation extraction process becomes easier. Second, regarding non-adjacent collocations other statistical operations need to be made. The need for developing other statistical measures not only emerges from the two above-mentioned limitations. It also emerges from the lack of clarity about how frequent a combination of words must be to qualify for a collocation, i.e., specifying a cut-off point for combinations to decide whether they are frequent enough to be collocations or not. This idea leads to what is termed “significance” of frequency. The association measure used in this study to evaluate the significance of frequency is Log Likelihood (LLD). A significantly frequent collocation may be a sign of native language use. Another related notion is “exclusivity”, which aims at measuring the strength of the relationship between the node and its collocate through measuring their dependence

---

3 Part of speech tagging is the process whereby words are labelled as a particular part of speech.

on each other. An exclusively and significantly frequent combination[4] may be an indicator of conventionalized collocation. The measure of exclusivity adopted for the purposes of our study is Log Dice. Significance of frequency coupled with exclusivity are crucial collocability dimensions that are adopted in both extracting and testing the collocability of the adjective+noun combinations in this study. On the one hand, since corpora are believed to represent language, a significantly and exclusively frequent collocation should be ideally accepted as both natural and conventionalized. However, this claim cannot be always true because of the limitedness of corpora. On the other, the opposite is always true in that a conventionalized collocation is always a significantly and exclusively frequent collocation.

2.5.2 Adjacency

Adjacency is a concept that is at the heart of the statistical approach. It refers to whether the collocation components are consecutive, or distant (i.e, they co-occur at a specific distance from each other). Adjacent words are those that co-occur immediately before/after each. Kjellmer (1996), for example, considers adjacency a crucial criterion that should be met before considering a word combination as a collocation dismissing those items co-occurring a few words apart. A node and its collocates can be either positioned right next to each other or a few words apart depending on their grammatical patterns. *Verb+noun* collocations, for instance, are commonly non-adjacent, as it is normal to find other modifiers in between, as in the case of *commit[ted]* and *crime* in *he committed a heinous crime.* The adjective *heinous* coupled with the indefinite article *a* separate the core components of the collocation which are *commi*t and *crime.* Adjacent nodes and collocates, however, can be demonstrated through *adjective+noun* patterns where the node and its collocate appear next to each other. Adjacency is represented in the statistical approach using the n-gram and window approaches. While the former better captures collocating elements over a big span, the latter identifies only adjacent n-grams. To facilitate and speed up the classification and analysis of collocations *adjective+noun* patterns are used to limit the possibilities of dealing with distant bigrams and to avoid "noisy data" Evert & Krenn (2003), which in turn should be discarded for lack of relevance.

2.5.3 Semanticity

Lexical items do not exist in isolation. They always co-occur in the company of other words belonging to the same collocational range, which is represented through the window in corpus-based approach. Therefore, each node is surrounded by different

---

4 In this thesis, exclusively and significantly frequent combinations are those that are both exclusive as evidenced by Log Dice, which is a measure of exclusivity, and significantly frequent as proven by Loglikelihood, which measures significant frequency.

collocates. Choosing the best collocate is subject to the context in which the node occurs and the meaning to be conveyed. The collocational items of a node are not randomly assigned to nodes. Meaning is not only constructed through the meaning of single words. It is, rather, a function of the relationships these words hold with each other. For this reason, the learning of new vocabulary is not confined to isolated words. It also involves the chunks and frames in which lexis fits. It needs to be accompanied by the knowledge of the semantic network that these words belong to and, therefore, the other words that collocate with them. According to Lewis (1997, p. 204) "instead of words, we consciously try to think of collocations, and to present these in expressions. Rather than trying to break things into ever smaller pieces, there is a conscious effort to see things in larger, more holistic, ways". In this regard a target-language collocation which conforms to the semanticity principle is one that is both connotatively and denotatively close to the source-language collocation. Semanticity is therefore the primary criterion against which the acceptability of collocations and translations is judged. Instances of L1 (Arabic) and L2 (French) interference can best demonstrate violation of the semanticity principles as they can show a clear deviation from the meaning of the source-language collocation.

Within semanticity I shall include the dichotomy of lexical substitutability vs. lexical fossilization as these are termed by (Nation 2001, Gyllsad 2007). Lexical substitutability refers to the flexibility of the collocational frame to subject its components to substitution, the substituted elements being replaceable by mutually interchangeable synonyms or near-synonyms. Lexical fossilization, by contrat, implies a restrictedness where lexical substitutability is not allowed. Lexical fossilization is also linked to non-compositionality (also sometimes referred to as semantic non-transparency) (Saeed, 2003; Dickins, 2020). Compositionality, as noted in section 2.4.2, is the principle that the overall meaning of a construction is determined by the meaning of its components and the rules used to combine them (Wang, 2018, p.1). This is further explained in Dickins definition for Multiword Expression; having a word with an independent sense in one lexical context, a few contexts or unlimited contexts allows for increasingly greater substitution among the components. On the other hand, a Multiword Expression none of whose constituents has an independent sense is lexically fossilised. Applying this to the translation test used in this research, we might expect that compositional source collocations will allow more candidate translations in the target language than non-compositional collocations.

2.5.4 Grammaticality

The principle of grammaticality relates to the grammatical 'frames' within which words collocate, i.e. the syntactic patterns governing lexis such as: *adjective+ noun,*

verb+noun, verb+adverb. This principle may relate to 'syntactic coherence' which occurred in Dickins' definition of formulaic expressions given that syntax is a subdivision of grammar. The focus of this study is limited to the *adjective+noun* syntactic pattern and accordingly the respondents' task was to translate only the underlined *adjective+noun* collocations. This choice was made to facilitate the assessment of the produced collocations and to limit the presence of recurrent function words such as 'and', 'for', and 'but' (Gyllstad 2007, p. 34). I say 'limit' and not 'eliminate' function words as collocations per se can be expanded through pre-modification and post-modifications to include grammatical elements such as articles, intensifiers, prepositions, etc. Therefore, in addition to determining whether two lexical items collocate together or not, there should be an awareness of whether lexical and grammatical items fit together or not. Thus, to judge the acceptability of the collocation a holistic approach should be adopted; one needs to look at the whole combination rather than only the lexical part of it. In our study, in order for the investigated collocations to conform to the grammaticality principle, their components need to occur within an appropriate grammatical frame (*adjective+noun*). This invokes the idea of the inseparability of grammar and lexis. Building grammatical awareness, then, entails knowing which lexical items occur within a grammatical structure. Collocations are also good examples showcasing the interwoven relationship of syntax and semantics in that both syntactic and semantic associations are required to produce native-like ('natural') collocations.

2.5.5 Conventionality

Conventionality is the favouring by native speakers of specific word combinations at the expense of other semantically equivalent (or nearly equivalent) alternatives. It is, then, is a principle attributing the habitual association holding between words to rules emerging from the 'cultural' properties of a given language, i.e. the way words associated together not only depends on semantics and grammar but also emerges from the set of norms of a particular speech community (Dukali, 2018, p. 12). While conventional phrases or word combinations can be significantly more frequent, not all significantly frequent word combinations are conventional. What can measure conventionality in statistical terms is rather measures of strength or exclusivity, i.e. how exclusively a word appears in the company of another. The notion of conventionality invokes the idea of "prefabricated" in Wray's definition of formulaic language and the idiom principle of Sinclair, respectively. It also corresponds to the notion of Pawley & Syder's (1983) native-like selection that may partially relate to significant frequency in Dickins (2020).The latter involves the limited selection of word combinations native speakers' use among a multitude of grammatical and semantic choices the system of the language permits. Neither semantics nor grammar

can all the time explain why a given collocate collocates with a given node. First, not all grammatical word sequences are native-like constructions. Second, the semantic arbitrariness of collocations, which often does not allow for the use of synonymous collocates shows the inadequacy of semanticity for judging the acceptability of collocations. The view that the typical co-occurrences of words are a result of their semantic properties is not always correct, and leaves the arbitrary semantic restrictedness of many collocations unexplained. The conventionality of the produced translations will be assessed by using the intuitions of native speakers. This comes as an alternative option to a more complex use of statistical comparison with other roughly synonymous collocations, which may not be conventional, as the latter would have been an enormous undertaking, beyond the scope of this thesis. From a frequency based-perspective, a conventionalised collocation is a construction that is both a significantly and exclusively frequent collocation.

2.6 Review of collocation use by Arabic- speaking learners

This section provides an overview of works that dealt with collocation use by Arabic speaking learners.

2.6.1 Zughoul and Abdul-Fattah, 2001

The study of Zughoul and Abdul-Fattah (2001) investigated the collocational knowledge of a total of 70 Jordanian students – 38 graduates and 32 undergraduates – using 16 verb collocations. Both the receptive and productive collocational proficiency of these students was measured using a productive and a receptive task. The findings demonstrated that overall students struggle with translating collocations from Arabic to English. It also showed that students performed far better in the receptive task than the productive one and undergraduates performed less well than graduates. Overall, the performance of the students was not satisfactory. The study revealed that students resorted to twelve translating strategies, with paraphrasing and avoidance (refraining from giving answers) being the most commonly adopted strategies, followed by literal translation. Overgeneralization, verbosity (using big-sounding words to 'enhance' the literary style in students' responses which altered the source SL message and idiomaticity (trying to find TL idioms for those of the SL resulting in deviant expressions from the intended message) were the least commonly observed strategies, the first two having the same frequency of occurrence (1.70%).

2.6.2 Ibrahim, 2003

This study investigates the translation of English collocations into Arabic. It assesses in practical manner the problems of translating English collocations into Arabic in dictionaries and Arabic press and seeks to identify the potential solutions that are

embodied in the translation strategies. The study begins with defining what collocation is and what is not. It defines the relationship between collocations and colligations and demarcates collocations from idioms, compounds, proverbs, concords, formulas, clichés, and citations.

The study reviews different definitions for collocations considering a variety of definitional aspects, the most important of which are the predictability of occurrence and the restrictive selection (a word restrictedly selects to appear with another). It then proposes the following definition: "The frequent co-occurrence of lexical items that naturally share the characteristics of semantic and grammatical dependencies" (p.18).

While the predictability of occurrence can be related to both Dickins' (2020) notion of frequency and the idea of conventionality being a habitual association as mentioned by the author, semantic and grammatical dependencies can be considered to correspond to Dickins'(2020) semantic independence and syntactic relationships.

When it comes to colligation, Ibrahim (2003) confined the relationship of collocations and colligations to how well the idea of separability of grammar and lexis is received. Retaining the Fithian definition of it, colligation as a notion has been defined as the appearance of words in grammatical syntagmatic relationships. In more specific terms, colligations do not view the relationship of phrases or word combinations as lexical associations but rather as grammatical relationships among grammatical categories. The author indicates also that there is an overlap between collocations and colligations in that collocations not only account for semantic associations, but also syntactic/grammatical associations, as is clear in his definition of collocations.

As in Dickins' (2020) model, idioms are accepted to be part of collocations and are syntactically immutable as mentioned by Mitchell (1971, 57-59): "The idiom belongs to a different order of abstraction. It is a particular cumulate association, immutable in the sense that its parts are unproductive in relation to the whole in terms of the normal operational processes of substitution, transposition, expansion, etc.". As far as compounds are concerned, the author adopts Mitchell's definition (1971, p. 60): "Compounds … may occur within the scatter of a collocation or even, though more rarely, of an idiom". While Mitchell's account of compounds is similar to that of Dickins in that compounds can be subsumed under collocations, it is different in that compounds are not included in idioms.

On the one hand, unlike Dickins (2020) model, Ibrahim (2003) explanation does not involve statistical significance as a defining feature to formula. On the other, it focuses on the syntactic relationships between lexical constituents and the pragmatic function of serving social interactions. Proverbs are defined as lengthier syntactic combination than collocations and aim at evaluating human actions and giving a moral example

without clearly setting clear boundaries among them and collocation Unlike, idioms, compounds, formula, and proverbs, the notions of cliché, concord, and citation are absent from Dickins' (2020) analysis of categories.

Ibrahim (2020) draws another analogy between collocations and clichés. The account concludes that neither of the two concepts is inclusive of the other, but discusses the fact that generally clichés can be described as more fixed collocations or idioms that lost their precision or force due to redundant use. Another concept, not present in Dickins (2020) ontology which Ibrahim's (2003) thesis deals with, is concord. He defines this as a grammatical agreement between words and sentences such as when a singular noun appears with a singular verb.

Citation, which is not present in Dickins's model, is defined as a closely related notion to collocation in the sense that they both are forms of lexical relationship. The description however, claims that while citation characterises the relationship of a word with its wider linguistic context, collocations are viewed as more specific in that they describe the lexical relationship of one word with another.

The study also identifies the following translation strategies for collocations in dictionaries: substitutability, which entails replacing the source collocate by a more or less general semantic TL; expansion and contraction, which are opposites in that the former produces longer TL equivalents while the latter results in fewer words in the TL; transposability: altering the word order in the TL equivalents; predictability, including anticipating the most predictable collocates in the TL based on the strength of attraction between lexical items; and cohesion, touching upon the possibility of transferring the formal, i.e., grammatical associations of lexical items from the SL to the TL. The latter can be exemplified by the fact that dual form, for example, does not exist in English. This may require adding a cardinal number 'two' when we translate to English as TL, an Arabic SL collocation in which the dual form is expressed by adding a suffix. Each of these strategies is discussed separately with illustrative examples, highlighting of the different potential ways of handling collocations using each strategy. The examples included throughout the study have been systematically extracted from two different texts: first, English-Arabic bilingual dictionaries, and second, Modem Standard Arabic from the Arab Press. The researcher claims that the main contribution of this research is showing how each of the adopted strategies results in producing natural and acceptable Arabic equivalents for English collocations, particularly, when obvious, basic TL equivalents are not found due to untranslatability. This remark suggests that collocations can be reproduced as either collocations or non-collocations in the TL.

2.6.3 Shehata, 2009

Very much like Ahmed (2012), Shehata (2009) assessed the collocational competence of 97 Arabic-speaking English learners: 35 thirty-five of them were enrolled at a university in the United States and 62 were undergraduate students in an English Department at an Egyptian university. The study examined both the receptive and productive collocation proficiency in relation to two factors: the learning environment of English as a second language (ESL) or foreign language (EFL); and the time of exposure to English. The data was collected through two productive and one receptive test. The instruments used were: a self-reporting questionnaire, two tests of 16 items each, where participants were required to fill gaps with the right verb/adjective, and a 50-item acceptability judgment test. The questionnaire aimed at measuring the exposure of EFL learners to the English language, and the acceptability judgment test attempted to assess the participants competence in recognizing English collocations. The two productive tasks of 30 items in total, 16 adjectives and 16 verbs, each aimed at exploring students' proficiency in producing English collocations. The two types of collocational patterns used in the study were *verb+noun* and *adjective+noun*. Data analysis revealed that the three factors of the first language (Arabic) and status of English as SL or FL in addition to time exposure to English all had an influence on the participants' translation performance. The study demonstrated that L1 interference had a detrimental effect on the collocational competence of both EFL and ESL learners albeit mitigated in the case of SL environment as EFL learners depended more on Arabic as FL in both reception and production of English collocations. Moreover, the time of exposure to English moderately and positively correlated with the collocational proficiency of the learners. As noted in many other studies, *adjective+noun* was the most challenging lexical pattern in both reception and production tasks and students performed better in the reception than in the production task. Based upon these findings, the researcher, like many others, suggested explicitly teaching collocations and particularly the *adjective+noun* pattern.

2.6.4 Boussalia, 2010

The study explored the reasons behind the difficulties students encounter when rendering collocations from English into Arabic. The sample of the study was 30 third-year students drawn randomly from two groups at the Department of English at Mentouri University (Constantine). The data was collected through a 10-item translation task and 10 multiple-choice questions. The former task aimed to test both cultural and language-specific influences in translating from English to Arabic. In the latter, students were asked to fill in gaps with the suitable collocate from three given choices to test their familiarity with English collocations. The results of the study

revealed that third-year English students were unaware of the relevant English phenomena. The answers to the first task demonstrated that students were not familiar with English collocations. Their mother-tongue interference was a huge obstacle making them fail to recognize natural English collocations. The second task showed that the main strategy adopted in translating collocation was word-for-word translation, which created free compositional combinations rather than idiomatic collocations. This was largely due to replacing each source constituent in the collocation with a standard target equivalent, ignoring the fact that a word may have a different meaning than its usual one when combined with other words. The study concluded by recommending explicit teaching of collocations through implementing a contrastive analysis between Arabic and English and providing socio-cultural information to familiarize students with English collocations. It also recommended drawing students' attention to the need to learn as many collocations as possible.

2.6.5 Dweik and Abu Shakra, 2010

Dweik and Abu Shakra (2010) dealt with the issue of translating lexical and semantic Arabic collocations into English. According to the authors, lexical collocations are particular collocational patterns (*verb+noun, noun+noun, noun+adjective*) in which one of the constituents has a specific denotative meaning within that particular collocation. This recalls Dickins' non-fully free compositional collocations that are semantically less transparent than fully free compositional collocations in that they have at least one component which does not have an independent sense in unlimited contexts. A semantic collocation is defined by Dweik and Abu Shakra (2010, p.12) as a metaphorical collocation, where the presence of the figurative meaning causes it to be perceived as more non-free compositional than a lexical collocation.

Both types of collocations were taken from religious texts particularly the Quran, the Hadith, and the Bible. The study aimed at identifying the different strategies adopted by 35 Masters students majoring in translation, in translating religious semantic and lexical collocations from Arabic into English. To this end, the researchers devised a 45-item translation test, 15 items per text, for students in three different Jordanian universities to get some insights into the different possible strategies that these students used when rendering both semantic and lexical collocations in the previously mentioned religious texts. The findings demonstrated that students adopted different strategies of translation for different texts including generalization, synonymy, deletion, paraphrasing and literal translation. Generally, students used synonymy as the first resort in translating lexical collocations. However, the strategy that was most commonly used by students when translating collocation in the Hadith was deletion. As far as the Quran and Bible's semantic collocations are concerned, students resorted to literal translation.

2.6.6 Alsakran, 2011

This study evaluated 68 advanced Arabic-speaking students' receptive and productive collocational knowledge in two different settings: ESL (38 advanced English learners at the Institute of Public Administration in Riyadh, Saudi Arabia), and EFL (30 Advanced English learners in the Intensive English Program at Colorado State University). Receptive knowledge was measured via an appropriateness judgment test, while three gap-filling ('fill in the blank') tests were used to measure productive knowledge. This was done through measuring students' performance on three types of collocations: *verb+noun, adjective+noun* and *verb+preposition*. Although the results revealed that there was a significant difference in performance between EFL and ESL learners across both receptive and productive tasks, both groups exhibited poor collocational knowledge. Learners' performance also varied across the three types of collocations; *adjective+noun* and *verb+preposition* proved more difficult than *verb+noun* collocations. Alskaran recommended replicating his study with university students, as they are more advanced than English language intensive-programme students and using free production rather than cued production collocation tasks. He also suggested that more attention should be given to teaching collocations both explicitly and implicitly through a variety of exercises.

2.6.7 Ahmed, 2012

The study investigated the collocational knowledge of the 185 undergraduate EFL students in the Department of English at Al-Jabal Al-Gharbi University in Libya in relation to both receptive and productive tasks. The data was collected through a self-reporting questionnaire, a 60-item multiple choice test, and a 28-item translation task. The questionnaire aimed at measuring the exposure of EFL learners to the English language, the multiple-choice questions (intending to assess students' receptive competence in recognizing correct English collocations) and the receptive test attempted to examine Libyan EFL learners' competence in recognizing English collocations. The 28-item translation task aimed at exploring students' proficiency in producing English collocations. Six types of collocational patterns were used in the study (*verb+noun, noun+verb, noun+noun, adjective+noun, verb+adverb,* and *adverb+adjective*). The study showed that the ability of Libyan students in both recognizing and producing collocations of all lexical patterns was lower than expected. The year of study influenced the outcome of the test; third-year students performed better than second-year students, but gender had no effect on their scores. The study also showed that students performed better in the receptive task than the productive task (translation). It demonstrated that L1 interference had a negative effect on the collocational competence of Libyan EFL learners. By contrast, exposure to English had a strong effect on developing collocational competence. Those who

engaged more with English-related activities such as watching movies and shows, reading English material and surfing the internet in English achieved better scores than those who had lower or no exposure to English. The study revealed that *adverb+adjective* and *adjective+noun* were the most challenging lexical patterns in the receptive task, while *adjective+noun* was the most challenging in the production task.

2.6.8 Shammas, 2013

This study aimed at investigating both the comprehension and use of collocation of 96 Arabic-speaking Masters students at four universities in different Arab countries. Three questionnaires were devised. The first comprised twenty collocations in Arabic for the participants to translate into English. The second questionnaire consisted of twenty English collocations to be rendered into Arabic. The last and third questionnaire comprised nine English collocations, each of them accompanied by four possible Arabic translation equivalents; respondents were asked to choose among the four options the best Arabic collocate for each node. Participants in this study had no access to any references. The findings revealed that the errors in the first questionnaire totalled 1,478 out of 1,920 responses, which equates to 76.979%. The percentage of errors in the second questionnaire amounted to 63.437% which corresponds to 1,218 errors out of 1,920 attempts. The errors in the third questionnaire amounted to 2,712 out of 3,456 attempts, which corresponds to 78.472%. Across the different institutions and questionnaires, the researcher reported that the performance of the participants was far poorer than expected, albeit that it varied slightly. This was due to the interference of the mother tongue evidenced by using the literal translation strategy.

2.6.9 El-Dakhs, 2015

This study investigates the collocational competence of 90 Arab undergraduate students studying English at different levels at a private Saudi university. The study focuses on four main points: 1/ the learners' collocational competence; 2/ the effect of language exposure on their collocational competence; 3/ the effect of the type of collocations on the students' attainment; 4/ the types of errors produced. The results of this study demonstrated that the overall performance in the test was unsatisfactory. This was unexpected to the researcher given that English was the medium of instruction for these students and that most instructors were native speakers of English while a few were native-like speakers of English. The attainment of the students slowly increased when exposure was lengthened. It was also noted that across the three different levels, the *adjective+noun* collocation pattern was more difficult for students than *verb+noun* pattern. The predominant type of error was intralingual

rather than interlingual, and this resulted in producing unacceptable collocations. The researcher concluded by giving some pedagogical recommendations such as the explicit teaching of collocations to hasten the collocational competence of learners. El-Dakhs (2015) suggests that teaching vocabulary should not be confined to teaching words separately but rather teaching frequent word combinations in wholes as ready-made chunks in an attempt to enhance the collocational competence of English language learners. He highlights collocational restrictedness, which does not permit using synonymous equivalents for collocates.

2.6.10 Hammadi, 2015

This study used two questionnaires as tools to collect data. The first investigated EFL teachers' perception of collocational difficulties faced by third-year university students in the English department of Tlemcen University and their suggestions for overcoming these difficulties. The second questionnaire examined collocation awareness of 30 third-year students of English at the same university. The questionnaire was divided into three tasks. First, it asked the students about the difficulties they encountered when speaking and writing English and how lack of collocational knowledge affected their production. Second, it gave a translation task involving both lexical and grammatical collocations (*adjective+noun, noun+noun, noun+verb, verb-adverb verb, verb+preposition, noun+preposition, adjective+preposition*), where students were required to translate 10 English collocations into Arabic and vice versa. Third and last, it provided a collocational grid in which students needed to match each Arabic collocation with the best equivalent out of several options in English. The study demonstrated that EFL university learners struggled in producing both grammatical and lexical collocations. It also showed that students struggled more with *adjective+noun* and *verb+adverb* collocations than other types, for three main reasons: L1 transfer, synonymy, and overgeneralization. The study concluded by identifying a set of practical implications for teachers, learners, and curriculum designers. For teachers, it suggested raising students' awareness of collocations and highlighting the limitations of literal translation. It also recommended using a set of exercises where students use monolingual and collocation dictionaries such as the Oxford Dictionary of Collocations, the BBI Dictionary of English Word Combinations and the LTP Dictionary of Selected Collocations. The dictionaries could be used in class to generate lists of collocates for particular words, especially near-synonyms such as 'do' and 'make'. The study also advised students to read, and paraphrase English texts in order to become aware of the collocational patterns of the texts, and to create acceptable collocations in their own production. For learners, the study focused on exposure to the English language through reading and watching English videos and

channels, recording every new word and learning its collocates. For curriculum designers, the author suggested a through revision of old materials to fit in collocation activities and vocabulary exercises where possible to enhance collocational competence.

2.6.11 Jabak, Abdullah and Mustapha, 2016

The study attempted to investigate difficulties in translating Arabic collocations into English and the underlying reasons for these difficulties. It also provided practical ways to improve the learning and teaching of collocations at universities. The participants were Saudi undergraduate students, a sample of 50 undergraduate students at the English Department at the Teachers College of King Saud University being chosen. The instrument used was a translation quiz of 15 Arabic sentences containing collocations. Students were given 90 minutes and were able to use monolingual and bilingual dictionaries to look up the meanings of the words. The study demonstrated that the participants struggled in rendering collocations from Arabic to English mainly due to literal translation via the use of bilingual dictionaries, and lack of knowledge about collocations in both Arabic and English. The researchers recommended the following. First, students should be taught two courses of translation each per semester during the whole period of study, one for Arabic-English translation and other for English-Arabic translation. Second, the use of monolingual rather than bilingual dictionaries should be promoted, as the former help with creating a more linguistically and culturally immersive environment in the language. Teachers should provide a model translation for the given translation tasks so that students can highlight areas of weakness and seek improvements. They should also provide examples of how literal translation can distort or alter the meaning of the source text and always provide acceptable and valid translations for such instances.

2.6.12 Almaktary, 2017

This study examined the productive collocational knowledge, both spoken and written, of 84 Tunisian EFL learners over seven collocational types using a collocation test of 100 items (10 items for each type), a focus group discussion, and document analysis. It explored the familiarity of these learners with different English word combinations, mainly collocations. It then tested their awareness of English collocations and erroneous use of collocations, and inspected the correlation between the learners' collocational competence and their performance on two productive tasks (written and oral). The study concluded that general collocational competence was unsatisfactory for different reasons, the main ones being: 1/ the informants' limited knowledge of collocations (38%); 2/ application of the strategy of transferring L1 to L2 collocations (27%); 3/ the effect of the source text patterning (14%); and 4/ overgeneralization (18%). It also revealed a positive relationship between the

learners' general collocational competence and their written and oral skills. The results also showed a correlation between the collocation test scores of the participants and their writing and speaking skills measured by the interview and document analysis respectively. The study recommended enhancing the exposure of EFL Tunisian learners to authentic English language, both written and spoken, and raising learners' awareness of the importance of collocational competence in learning a foreign language and encouraging them to be more independent in their learning. According to the author, this should be done via extending and expanding learning outside the university setting and getting students to develop their own strategies for learning collocations.

2.6.13 Dukali, 2018

This study, aspects of which have already been discussed in section 2.5.3, was corpus-based and examined Libyan undergraduate English majors' challenges in using *verb+noun* and *adjective+noun* collocations. The sample of the study were fourth-year English major students at the Department of English at Tripoli University. Data was collected through a corpus which comprised 240-word academic essays freely produced by each student in 45 minutes. Twelve verbs and twelve adjectives were the focus of the study. The data was analysed using AntConc3.2.1w (Anthony 2007). The collocations extracted from the corpus were analysed in four different ways: 1.Oxford Collocational Dictionary (OCD) 2009; 2. British National Corpus; 3. Native speakers' intuitions; and 4. the acceptability survey of collocability. This last was used to triangulate the three previous methods. Erroneous *verb+noun* and *adjective+noun* collocations were classified into three main categories: grammatical, lexical, and usage-related errors. These errors were broken down into sixteen subtypes for *verb+noun* collocations and twelve subtypes for *adjective+noun* collocations. The results showed that grammatical errors were more frequent than lexical errors in the produced collocations. Accordingly, the author recommended introducing the teaching of the whole collocational pattern rather than just its components, i.e., teaching within collocational patterns pre-modifications and/or post-modifications of nouns such as: articles, intensifiers, prepositions, and possessive pronouns. The author's recommendation stems from Nesselhauf's (2003, p. 238) suggestion to adopt a comprehensive approach in teaching collocations through teaching the entire collocational combination, thus going beyond its constituent lexical items. Al Dukali's study also revealed that register can affect word choice in producing collocations. The focus of her study was academic written English, in which non-awareness of register may lead students to choose the wrong verb or adjective although thus may result in acceptable collocations in spoken English. Al Dukali identifies that L1 interference is one of the main reasons why students produce

unnatural collocations, these being based on collocations found in colloquial Libyan Arabic. In addition to L1 interference, overgeneralisation, and use of synonymy are other factors that resulted in erroneous combinations according to the author, who recommends that collocational activities should focus on these problems.

2.7 Relevance of the reviewed collocation studies

Previous research has investigated learners' collocational competence over productive and to a lesser extent receptive skills such as (Alsakran, 2011; Ahmed 2012; Zughoul and Abdul-Fattah 2001) in ESL and EFL settings using different methods, the main ones of which are corpora of students' writing and translation test instruments. These studies show that learners' productive collocational competence is always worse than receptive. Thus focusing on productive collocational competence is of paramount importance (Zughoul and Abdul-Fattah, 2001; Alsakran, 2011). In this light, more research is needed in particular to elucidate why producing collocations is always harder than recognising them. This study therefore attempts to do this by using a translation test instrument to investigate the issue. This explains why the above review of studies prioritised collocational studies using translation test instruments rather than compiled corpora of students' writings. Another common aspect found in the surveyed studies is the use of small size samples combined with few test items. Exceptions are Ahmed (2012) and Almaktary (2017). Similarly, the procedure for producing the sample from the population is rarely discussed in the reviewed works. Furthermore, most of these studies do not adopt a systematic approach to the choice of test items. Thus, it is not clear how the test items were selected and sometimes the whole process of item selection is barely described, if at all. With small sample sizes and a small number of ambiguously selected test items, it becomes impossible to draw conclusive and generalizable results. Besides, no definite association is drawn between general language proficiency and collocational knowledge. In the light of this, very few studies explored collocational knowledge within different levels (exceptions are Ahmed 2012, and Zughoul and Abdul-Fattah, 2001), to clearly determine whether collocational proficiency increases with level of study or not.

2.8 Conclusion

This chapter gave a brief overview about the common approaches used to define 'collocation' and argued in favour of the amalgamation of the statistical approach and the phraseological approach to identifying and producing collocations. This can be operationalised through the use of the statistical or frequency based-approach as a starting point to investigate the phenomenon of co-occurrence and then back this up with evidence from the phraseological or linguistic approach that these word

combinations are native-like. A distinction was drawn between collocations and other related phenomena using Dickins' (2020) model of semantic independence. Throughout this thesis, the term ‘collocation’ refers to what Dickins (2020) terms a 'statistically significant collocation' that can be analysed according to its semantic composition and grammatical rules. Taking into account both semantic compositionality and syntactic relations, this thesis adopts a definition that combines both phraseological and frequency-based approaches. The chapter also presented some related works that examined collocational knowledge of Arabic speaking students and briefly discussed their relevance to the current research.

3 Chapter three Collocational false friends and translation strategies

Introduction

This chapter comprises three main parts. The first part proposes an eight-category typology for translation strategies that are adopted for analysing students produced translation in this thesis based on the previous categorisations of three theorists and the students' decisions when dealing with the collocations. It begins with a discussion about how the terms 'strategy', 'procedure', and 'technique' are used in the literature and introduces a set of translation strategies that are adopted for the purposes of the analysis in this thesis. The second part of the chapter introduces the notion of language transfer particularly non-positive transfer and how it is to be investigated through collocational false friends in this thesis. This section starts with reviewing the relationship between false friends and collocations in the literature by examining previous studies that dealt with false friends. It then proposes a new sense for the term 'collocational false friends' (which has been used in a few other works previously, but in a different sense) for the purposes of this thesis. The third part of this chapter provides an overview of false friends and their typologies, and surveys studies that have investigated this phenomenon in the context of both second language (SL) interference on third language (L3) learning and translation.

3.1 Strategy vs. procedure vs. technique

According to Scott-Tennent et al. (2000, p. 108), translation strategies are "steps taken to solve a translation problem, which has been consciously detected and resulting in a consciously applied solution". Retaining the same concept of problem-solving procedure in translation, Lörscher (1991, p.76) and Séguinot (1991, p.82) provide a very similar definition except that the latter adds the possibility of unconsciousness to the process. Similarly, Baker (2018, p. 26) defines translation strategies as practical procedures taken by translators to overcome "various types of non-equivalence".

Many terms have been used more-or-less synonymously to refer to these problem-solving measures such as *procedure* and *technique*. While the latter goes back to "techniques of adjustment" introduced by Nida (1964, p. 23); the former was coined by Vinay and Darbelnet (1958), adopted by Malblanc (1968, p. 35), and further developed by Vázquez-Ayora, 1977; cf. Newmark, 1988). Conversely, a single term

like ‘method’, ‘strategy’, ‘procedure’, or ‘technique’ may be used with different meanings by different authors (see below).

Given this variety of terms referring to the same general concept of finding practical solutions to overcome difficulties while transferring meaning from the source language to the target language, as well as the use of the same term to refer to different concepts, it is unsurprisingly difficult to find a common consensus over a prototype definition (Gutiérrez, 2018, p. 23).

Among the theorists and researchers who use these three terms interchangeably are Wotjak (1981; cf. Mason, 1994, p. 63) along with *rule* and *method* respectively. In contrast Newmark (1988), Zabalbeascoa (2000) and Hurtado Albir (2001) distinguish between these terms. Newmark (1988, p. 45) draws a distinction between translation *procedures*, as specific decisions applied to smaller linguistic units, and translation *methods* as the overall approaches used with texts.

While Newmark differentiates between *method* and *procedure*, Zabalbeascoa (2000, pp.120-121) distinguishes between *strategy* and *technique*, where the former refers to the overall actions taken to facilitate the rendition of a text, while *technique* is more about the decision-making process and acquired skills. Seemingly, Hurtado Albir (2001, pp. 249-250) retains both Newmark’s definition of *method* and Zabalbeascoa’s definition of *strategy* and equates Newmark’s *procedure* with Zabalbeascoa’s *technique.* In this research, while the term *strategy* is used for the general translator’s ‘gameplan’ in dealing with an ST/TT pair, *technique* which is generally used synonymously with *procedure,* refers to the individual decisions made by the subjects of the study while translating collocations.

Translation strategies are of two major types: *literal* and *free*, which, in turn, have been given various synonymous names. Vinay and Darbelnet (1958) call them ‘direct’ and ‘oblique’. This dichotomy corresponds roughly – though not exactly – to Nida’s (1964) formal vs. dynamic translation, House’s (1986) overt and covert translation, Newmark’s (1988) semantic and communicative translation, Nord’s (1991) documentary and instrumental translation, Jääskeläinen’s (1993) global and local translation, and Chesterman’s (1997) comprehension and production strategies. Among all these dichotomous translation types, Vinay and Darbelnet (1958) provide the most widely acknowledged and develop a list of seven translation strategies (as shown in table 1), three of which – 1. calque, 2. borrowing, and 3 literal translation –

fall under the subheading of direct translation and, four – 4. transposition, 5. modulation, 6 equivalence, and 7 adaptation – under the subheading of indirect translation. Twelve supplementary strategies to the seven basic ones of Vinay and Darbelnet were added, ten of which are opposing pairs (see table 2) – concentration vs. dissolution; amplification vs. economy; reinforcement vs. condensation, explicitation vs. implicitation, and generalization vs. particularization – while the two remaining techniques – compensation and inversion – are non-paired strategies. While Vinay and Darbelnet's proposed translation strategies are used in translation in general, Baker suggests eight translation strategies (see table 3) to specifically tackle the issue of the non-equivalence at word level: 1. using a superordinate (more general word), 2. using a more neutral or less expressive word, 3. cultural substitution, 4. using a loan word with a subsequent explanation such as footnotes, 5. paraphrasing using related words, 6. paraphrasing using unrelated words, 7. omission, and 8. translation by illustration (pp. 26-38).

Translation strategy		Definition
Direct Translation strategies	Borrowing	Loan
	calque	Foreign word or phrase translated and incorporated int the TT
	Literal translation	Transferring the SL into the TL Word-for-word
Oblique Translation strategies	Transposition	Change in the grammatical class
	Modulation	Change in the point of the view of the SL message
	Equivalence	Functional replacement
	Adaptation	Modifying the cultural aspects

Table 1: Vinay and Darbelnet's main translation strategies

Translation strategy		Definition
Complementary translation strategies	Concentration	Shortening the ST expression when translated into TL
	Dissolution	Lengthening the ST expression when translated into TL
	Amplification	Using more words to make up for lexical or syntactic gaps
	Economy	Reducing a shorter version of the ST in the TL
	Reinforcement	Reinforcing grammatical elements of ST in the produced TL
	Condensation	Condensing the ST grammatical elements in the TL
	Explicitation	Making explicit in the TL the implicit information of the ST
	Implicitation	Making obvious information of ST implicit in the TL
	Generalisation	ST word/phrase is translated into a more general term/ phrase in the TL
	Particularisation	Producing a more specific term/phrase in the TL
	Compensation	An ST element is rendered somewhere else in the TL
	Inversion	Change in word/ phrase order

Table 2 Vinay and Darbelnet's complementary translation strategies

Translation strategy		Definition
Baker's Translation Strategies for non-equivalence	Using a subordinate	Translating the ST word/phrase by a more general term
	Using a more neutral term	Producing a less expressive term
	Cultural substitution	Replacing a ST cultural reference with its correspondent in the TL
	Loan word	Using a loan word or a loan word with explanation
	Paraphrasing with related words	Using related words to explain a non-lexicalised item or a less frequent form in the target language
	Paraphrasing with unrelated words	Using unrelated words to simplify a non-lexicalised item in the TL that is semantically complex
	Omission	Omitting a word /phrase if its meaning is not of paramount importance to the text
	Illustration	Using illustrations rather than lengthy explanations for items that have no equivalents in the TL

Table 3: Baker's translation strategies for non-equivalence

For the purpose of this analysis, Vinay and Darbelnet's equivalence and literal translation, and Baker's paraphrasing, omission and substitution, are used. Borrowing and calque are included under literal translation, and the dichotomy of generalization vs. particularization along with Baker's second technique (using a more neutral or less expressive word) are grouped under *approximation*. Molina and Albir's dissolution, economy, condensation and implicitation are all grouped under *reduction*. The main strategy of modulation and the four complementary strategies of concentration, amplification, reinforcement, and explicitation are collapsed into *paraphrase*. The other two complementary strategies, inversion and compensation, the two main strategies of transposition and adaptation, and Baker's illustration, are not adopted (for reasons, see below). Finally, synonymy is added. The following diagram shows the relationship between the strategies recognised for the analyses in this thesis and those adopted from Vinay and Darbelnet (1958). The strategies found in Vinay and Darbelnet are represented with a blue background, those added by Baker are represented with a red background, those added by Albir are represented with a brown

background, and those added by me are represented with an orange background (cf. Vinay and Darbelnet 1958, 1995, 30-42; Molina and Hurtado Albir, 2002; Baker 25-42).

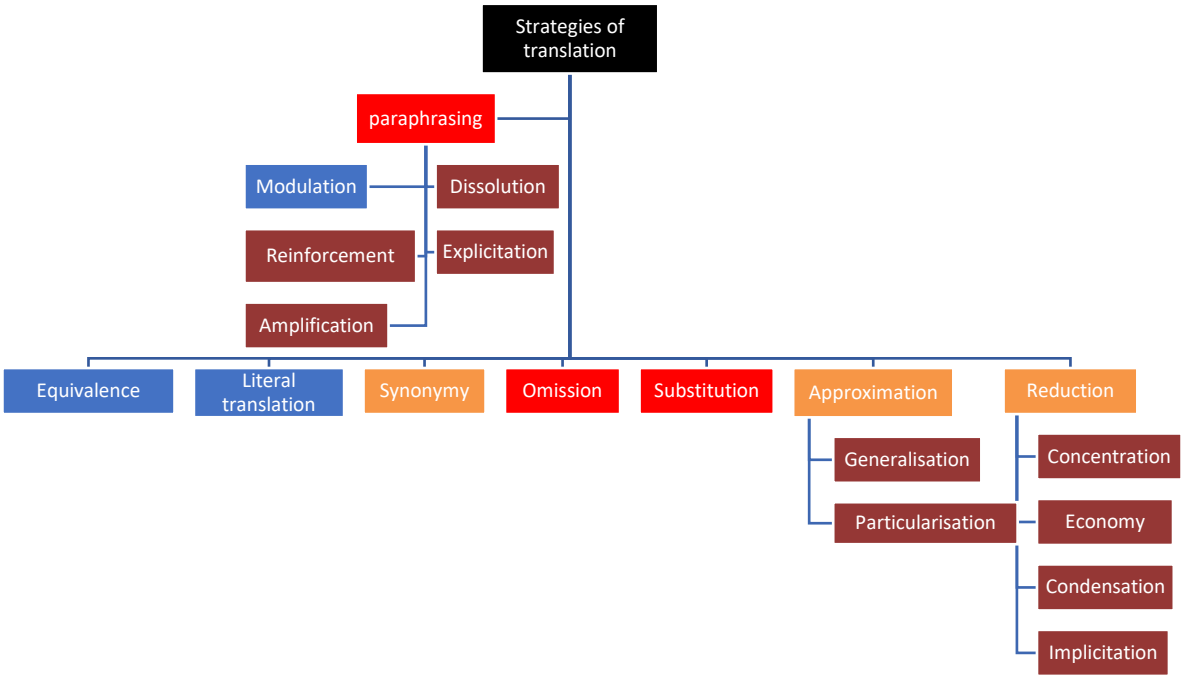

Figure 4: Translation strategies recognised in this thesis, in relation to those of Vinay and Darbelnet

Before discussing each of the adopted strategies, I shall explain the reasons for omitting transposition, inversion, compensation and adaptation and Baker's illustration strategy.

3.2 Non-included techniques

3.2.1 Transposition

Transposition entails altering either the grammatical class or the order of grammatical elements from the source language to the target language. This is a very common strategy that occurs between Romance languages, and English given that many of their syntactic structures are different. It also applies to English and Arabic. While sentences in English normally have Subject, Verb, Object (SVO) word order, Arabic sentences often have VSO word order (though SVO is also possible). Similarly, adjectives in Arabic and English are positioned differently: while in the former the adjective follows the noun, in the latter it precedes it. Since this study deals with the rendition of *adjective+noun* collocations, it is self-evident that all answers are likely

to involve this strategy while still fulfilling the criteria of other strategies. Accordingly, transposition here is regarded as a complementary strategy. In order to avoid having an overlap between categories, transposition is not considered as a separate category.

Although Vinay and Darbelnet (1958) categorized transposition as an indirect translation strategy, this research does not involve itself with this categorisation, and regards the particular translation as either direct or indirect depending on the other strategy with which it occurs. To put it differently, when affecting word category, transposition becomes more of a paraphrasing strategy as it results in reformulating language, which is the essence of the paraphrasing strategy. However, when involving the order of words, transposition instances can still also fit other criteria for other strategies. For example, some instances of literal translation can as well be considered instances of transposition.

3.2.2 Compensation

This involves a source language aspect that cannot be translated into the target language and needs to be compensated elsewhere in the text. It can be 'an item of information or a stylistic effect from the source language that cannot be reproduced in the same place in the target language' (Molina and Hurtado Albir, 2002, p. 500). To make up for the loss occurring in the translation, similar semantic or other features are added in another place in the text. An example where compensation may be needed is with nuances of formality of the French pronouns *tu* and *vous*, which are both standardly rendered as *you* in English. Vinay and Darbelnet (1958) were the first translation theorists to provide a definition for compensation. This translates into English as follows: "The stylistic translation technique by which a nuance that cannot be put in the same place as in the original is put at another point of the phrase, thereby keeping the overall **tone**" (1995, p.341). This definition was further developed by Hervey and Higgins (1992, p. 248): "the technique of making up for the translation loss of important ST features by approximating their effects in the TT through means other than those used in the source ST". As the translation task in this thesis is just to translate bigram collocations, compensation is irrelevant to this analysis.

3.2.3 Inversion

This is also referred to as *dislocation* Vinay and Darbelnet (1958, p. 222). It is a structural change that affects the syntax of sentences or paragraphs, and entails

inverting the word order of a whole sentence or just a phrase (Gutiérrez, 2018, p. 55) in an attempt to conform to particular stylistic features of the target language and make the translation more intelligible to the target reader. As this strategy is therefore only apparent in translating whole clauses or longer chunks of language, it is irrelevant to our data which is restricted to bigram collocations. The following sentence is an example of using inversion:

Did you send this letter, or didn´t you? > Cette lettre, tu l´as envoyé oui ou non ?

3.2.4 Illustration

This is often relevant when an equivalent word in the target language referring to a physical entity cannot be found and therefore illustrated using pictures. An example of this is using an illustration for a tagged teabag instead of going into a lengthy explanation of what the word 'tagged' means in Arabic. As the main focus of this strategy is the use of illustrations, it is completely irrelevant to this study.

3.3 Adopted strategies

3.3.1 Literal translation

This is a direct word-for-word strategy where both source and target languages share the same concept, and structural and lexical equivalent (Molina and Hurtado Albir 2002). This usually occurs when two languages are very similar to each other (Vinay and Darbelnet, 1958). Literal translation is a source language-based strategy (Dweik and Abu Shakra, 2010, p. 17) which entails transferring the elements of the collocations word for word into the target language.

Literal translation is, in fact, a superordinate (hyperonym/hyerpnym) of the other two direct translation techniques, calque and borrowing, since all instances of these two categories are also instances of literal translation. The former is a word-for-word translation at the phrase level of a specific expression like: *Adam's apple* which is taken from *pomum adami* and the English expression *black market*, which apparently originated in English in 1931 as a translation of the German word *Schwarzmarkt* and is rendered literally as *mercado negro* in Spanish and *mercato nero* in Italian. Borrowing, by contrast, entails keeping source-language terms in the target language due to the unavailability of equivalents in the target language and to preserve their cultural aspect of the source language. Borrowing can result in either unchanged words like the Italian *pizza*, which remains the same in most languages or incurs slight

changes to fit target language rules like *meeting* in English becoming *mitin/mitín* in Spanish. All instances of borrowing and calque in this study are to be considered as examples of literal translation.

3.3.2 Omission

This is also termed *avoidance* by Abdul-Fattah & Zughoul (2001). It involves avoiding the rendition of the collocations and one cannot be clear whether it is adopted due to lack of determination or to the participants finding the test items difficult (Abdul-Fattah & Zughoul, 2003; Dweik & Abu Shakra, 2010). Abdul-Fattah & Zughoul (2001) found that omission, which they referred to as 'avoidance' was the second most adopted technique with 28.24 % after paraphrasing. They ascribed the use of this technique to either lack of knowledge, or unwillingness to undertake a voluntary task of half an hour. It also partially accords with Dweik and Abu Shakra's (2010) deletion strategy, involving eliminating either a part or whole answer, which they found to be the commonest technique used in rendering Biblical and Hadith collocations and the third most commonly adopted technique for rendering Quranic collocations.

3.3.3 Synonymy

This strategy entails the use of synonymous words of either one or both constituents of the collocation of the expression in an attempt to provide an equivalent target collocation. What might cause erroneous or less accurate translations when adopting this strategy is that full synonymity is not always attainable especially if the collocation in question is clearly conventionalised. The consensus among linguists and translators seems to point towards the non-possibility of full-synonymy Cruse, 2000, pp. 157-158). However, some argue that full-synonymy is possible if the lexical times are carefully chosen within particular contexts. Al-Barakati claims that:

> In translation, we are normally dealing with a defined context; therefore, synonyms can easily act like full equivalents if carefully picked. Therefore, approaching equivalence by taking such a narrow view will only lead to drawing general and inaccurate assumptions. For example, to refute Jakobson's notion of equivalence … the verbs 'begin', 'start', and 'commence' may also be used interchangeably in various contexts without any noticeable loss of meaning as long as they collocate with the occurring lexical items in the sentence (2013, p. 93).

3.3.4 Equivalence

The concept of equivalence had always been "a vague, hard-to-define concept" as stated by Wilss (1996, p. 3) and thereby multifariously defined. In this thesis, Vinay and Darbelnet's view on equivalence as a strategy which "replicates the same situation as in the original, whilst using completely different wording" (ibid. p. 342) maintaining the stylistic features of SL in the TL text, is adopted. By stressing the role of the situation in determining equivalence, this definition, although it imposes a burden on translators, equally gives them more flexibility to choose – among different cultural references or images in the TL – the ones which not only are they listed semantic equivalents in a dictionary or a glossary, but which best mirror the situation expressed in the SL. An example of this is the fixed expression *Take one* which would have as an equivalent French translation *Prenez-en un*. However, if the expression appeared as a notice next to a basket of free samples in a large store, the translator would have to look for an equivalent term in a similar situation and use the expression *Échantillon gratuit* (ibid. p. 256). Having said that, the idea that equivalence is the best method to deal with proverbs, idioms, clichés, nominal or adjectival phrases seems sound and plausible.

3.3.5 Approximation

The term 'approximation' is suggested in this thesis to denote a strategy whereby the translator or student tries to approximate the meaning of the source language expression in the target language. It encompasses Vinay and Darbelnet's generalisation vs. particularisation dichotomy and Baker's strategies of using a superordinate (a more general word) or a more neutral or less expressive word. Approximation, therefore, involves either using a term that is semantically broader than the one of the source language (subordinate) (Baker, 2008, p. 42), a more specific term (hyponym), or a more neutral or less expressive word in the target language as a translation equivalent for the source language term or expression.

3.3.6 Substitution

This is defined in this context as a strategy whereby subjects replace the source collocation with a target collocation that is semantically distant from the source language collocation or completely unconnected and the choice of which cannot be ascribed to any clear reason. Substitution results in good collocations but unacceptable translations. In the translation task, Arabic collocations that are barely connected to

the source collocations were produced as equivalent target collocations, such as حقيقة مرة / بشعة *ḥaqīqa murra / bašiʕa*, which means *bitter truth* as an equivalent for *rude awakening*.

3.3.7 Reduction

This covers the four complementary strategies of Mona and Albir (2002) - concentration, economy, condensation, and implicitation, all of which are shortening techniques. Implicitation, for instance, reduces the details found in the source language when they are clear from the context in the target language by omitting. This is similar to concentration, whereby the translator shortens or reduces the number of words in rendering the source expression into the target language. This can be noticed in translating some phrasal verbs and prepositional phrases from English to French or Arabic like *to make up for*, which translates as *compenser* in French and عوض *ʕawwaḍa* in Arabic.

Seemingly, by compressing the original expression and reducing its size, concentration and implicitation are no different from economy and condensation. All of these involve suppressing information of the source language in transferring it into the target language. This is reflected in Molina and Albir's (2002) definition of reduction: "to suppress a ST information item in the TT" (p. 510). Given this, these four complementary strategies are referred to 'reduction' in this study, and include all instances of collocations being contracted into one word as in 'trends', 'spectrum', 'affiliations', 'indication', and 'conflicts' when attempting to *render* توجهات سياسية *tawajjuhāt siyāsiyya*.

3.3.8 Paraphrasing

This entails producing lengthier equivalents by means of description or definition. It corresponds to each of the following: amplification, dissolution, reinforcement, explicitation, and to a lesser extent modulation. Modulation entails reformulating the source language expression from a different point of view in the target language without distorting the original meaning, with the aim of making the target expression sounds more acceptable for the target reader. An example of this is when instead of literally rendering the English expression *Maybe you're right* as *Peut-être tu as raison* in French, it is translated through reversing *right* as 'not wrong', as in *Tu n'as peut-être pas tort*, which back-translates as *Probably you are not wrong.* Apart from the fact that modulation is highly likely to occur in long phrases and sentences, where the

change of point of view can be really noticed, it might also occur with small chunks of the language as in the collocation *uninhabited islands* when translated as جزر غير مقطونة *juzur ġayr maqtūna* and جزر غير مأهولة *juzur ġayr ma'hūla*, rendering the negative prefix 'un' as غير, which has the same function, instead of rendering the collocation as جزر نائية *juzur nā'iya*, where the one-word adjective conveys the entire meaning of the negative phrase.

Modulation entails rephrasing/reformulation, which is the core idea in amplification, dissolution, reinforcement, and explicitation. Amplification, for example, has been described by Molina and Albir (2002) as 'explicative paraphrasing' and 'includes explicitation'. Similarly, explicitation is meant to make the implicit meaning of the source text explicit in the target text, which will ultimately result in adding more words to be more accurate. Dissolution and reinforcement are no different from amplification and explicitation; all are 'lengthening' strategies for target language expressions resulting in an extended version of the short source language expression.

3.4 Collocational False Friends

The relationship between collocations and false friends has multifariously, if at all, been drawn in literature. Memišević and Margić (2011, pp. 73-74), for example, consider amongst types of false friends those false friends which involve collocations. Subsequently, they claim that collocations (involving false friends), partial false friends and full false friends are the most common false friends, often used by language learners in both second language acquisition and translation. In their study, the authors explored how the level of language proficiency and the formal exposure to false friends, as well as explicit teaching of these pairs, affect Croatian English learners' linguistic output. According to their findings, the rate of correct answers in the first two types of false friends (full and partial) increased across the four years of study which they looked at. Nevertheless, when it came to the third type of false friends (collocations), fourth-year students had the lowest success rate. The authors ascribed the disparity in performance with collocations to the fact that students in the first three years were taught according to a new curriculum that focused on collocations. Conversely, current fourth-year students were taught using a programme that did not entail passing previous exams to attend further advanced level courses, which might have detrimentally affected their motivation towards studying.

Contrasting with the above approach to the connection between collocation and false friends, other researchers view this relationship through the notion of 'collocational false friends', which is a complex idea that has been dealt with in a few works (e.g. Klégr and Saldová, 2006; Nuccorini, 2006; Heid and Prinsloo, 2008). In order to understand this concept Klégr and Saldová (2006) raised two crucial questions regarding what lexical false friends have in common with collocations and what makes collocations false friends. They explain then that the common element between false friends and collocations lies in the assumptions underlying them that "both deceptively suggest equivalence of meaning and/or grammatical properties in their TL counterparts" (p. 172). While a single false friend in the source language suggests 'deception' in relation to its target language counterpart due to their resemblance in form, translating a source collocation might suggest rendering it by its constituents' standard equivalents in the target language. Accordingly, instances of collocational false friends emerge due to absence of equivalence when rendering the constituents of a source collocation by their immediate target counterparts. Collocational false friends are thus "an interlingual concept for translation and ELT" (p. 170).

This suggests that collocational false friends occur at the level of collocations between source single word constituents and their target counterparts even if each of the constituent elements are true friends. This is also supported by Heid and Prinsloo (2008, p. 1357) claiming that "true friends in their use as individual lexical items often become false friends in collocations". The fact that collocations have combinatory idiosyncrasies in each language complicates the search for equivalent target collocations in another language even when a true friend single-word equivalent exists in the two languages. This extends the idea of false friends and lexical preference from the single word level to the collocational level. Heid and Prinsloo (2008 p. 1360) state that "As collocational combinatorics is often more a matter of preference than of categorical decisions, lexical false friendship is sometimes also a matter of preference". Collocational false friends exist even between closely related languages or varieties within the same language through the occurrence of "a network of true/false relations exists in terms of the collocative function of the translation equivalents" (ibid. 2008, p. 1361). To illustrate this, the authors give the example of German *Angst bekommen* meaning 'become anxious' with its Afrikaans and Dutch equivalents. While the German collocation *Angst bekommen* is fully equivalent to its Dutch counterpart *angst bekomen* as evidenced by the frequent co-occurrences of

these collocations in German and Dutch corpora and Google searches, the Afrikaans *angs bekom* is not found. This does not necessarily mean that the Dutch word *bekomen* makes a collocational false friend with its Afrikaans counterpart *bekom* in other collocational network of relationships (p. 1361).

Similarly, Nuccorini's study (2006) investigates how some pairs of true friends in English and Italian behave in collocational relationships and often prove to be translational non-equivalents. The study uses comparable corpora to search for true cognates and explore their use in the two languages. The author gives the example of the pair of adjectives of terrorist and *terroristico* in 'terrorist attack' and in *attacco terroristico* respectively. While the perpetrators of terrorists attack are often referred to as 'suicide bombers' (p. 34) in English, they are referred to by two words in Italian. The first is as *terroristi*, which has a broader sense than its English equivalent 'terrorists', in that not all *terroristi* are suicide bombers. The second is by using the Japanese loan word *kamikaze*, which has a restricted meaning in Italian, in that it refers only to suicide bombers. The author concludes that although the pair 'terrorists'/*terroristi* are reported as lexical, semantic and translational equivalents in dictionaries such as the Oxford-Paravia, they are only partial translational equivalents in the sense that *terroristi* are often but not necessarily 'terrorists' (p. 34). This difference in use in the English-Italian pair is clearly noticed through the different network of relations developed with other words surrounding them in both languages.

In the studies reviewed above, the term 'collocational false friends' is used in the sense of true cognates in different languages becoming either not translation equivalents or only partial translational equivalents when these cognates appear as collocational nodes and attract different collocates. Contrary to how the term 'collocational false friends' is used in the literature, this thesis expands the scope of collocational false friends to also refer to false friends being used as nodes within semantic and syntactic associations connecting them to their collocates.

To understand the idea underlying the notion of false friends, we need first to trace back the term itself, finding out its equivalents, converses and related linguistic phenomena. 'False friends' as a term is the opposite of 'true friends', which are simply cognates (for the sense of 'cognate' as used in this thesis, see section 1.5.2 below). False friends and cognates have been considered different concepts. One way of defining false friends, then, is to highlight the difference between these two concepts.

3.5 Language transfer

Languages are idiosyncratic. They however share many similarities that may help as much as hinder in the process of learning a foreign languages. When learning an L2, students habitually refer back to their L1 repertoire to associate what they are learning with their previous knowledge, because knowledge is accumulative, so that they can find a common ground for L2 learning (Negadi 2015). Similarities can be useful when they help students in learning another language through what is known as "positive transfer". This can be depicted in the occurrence of what is called "cognates". In some other cases, similarities can pose a significant hurdle rather than an advantage. This can be clearly manifested in the occurrence of false friends as they prove to be a complex area for both foreign language learners and translators (Brenders, van Hell and Dijkstra 2011).

Language transfer, also termed 'cross linguistic influence', is the result of transferring/applying previous knowledge be it lexical items, rules, or structures from L1 (source language) to L2 (target language). Accordingly, it is a 'you-know-it-when-you-see-it phenomenon' (Jarvis, 2000, p. 246). A learner's previous knowledge of his L1 entails all aspects of the language including phonology, morphology, syntax, semantics and pragmatics. Research on transfer mainly focused on phonology and syntax until the late 1980s, when interest in researching lexis in second language acquisition began to develop (Trimasse, 2018). Jarvis (2009, p. 99) defines lexical transfer in second language acquisition as: "The influence that a person's knowledge of one language has on that person's recognition, interpretation, processing, storage and production of words in another language". This definition points towards the influence of knowing all linguistic features of words in the source language on producing words in the target language (TL). This necessitates specifying what it means to know a word before investigating crosslinguistic lexical influence. In terms of vocabulary knowledge, there are several theories and models that are based on various conceptualisations of vocabulary learning and vocabulary use. These models attempt to account for the role of collocations specifically when it comes to vocabulary acquisition and usage. Jarvis' model, for example, (2009, p. 100) argues that knowing a word entails knowing its: 1. spelling and pronunciation; 2. meaning; 3. grammatical class; 4. syntagmatic associations; 5. lexical and conceptual associations'; 6. frequency of occurrence in the language; 7. degree of formality and

register in which it fits; and 8. where can be used appropriately and conventionally. Jarvis' model can be related to Nation's theory comprising nine aspects of what is involved in learning a word. These nine aspects are grouped under three main categories which are form, meaning and use. In this light, knowing the form of words entails knowing how the word is spoken and written, and its recognisable parts. Knowing the meaning of a word necessitates form-meaning mapping i.e., connecting a particular form to what it signals or refer to in terms of concepts in addition to its associations. Additionally, knowledge of word use needs a thorough understanding of a word's grammatical function, collocations and various restrictions related to its frequency and register. Even though Nations' model presents a convenient way to cover a wide range of word knowledge aspects, it can be viewed as a multidimensional conceptualisation in the sense that these aspects of knowledge are treated as separate dimensions. In this regard, González-Fernández (2022) suggested reconceptualising Nation's framework after finding out that various dimensions of word knowledge are shared across the two learner groups used in her study, indicating that L2 vocabulary knowledge is one-dimensional. Despite the fact that Nation's model does not explain how the proposed word-knowledge dimensions behave in actual vocabulary use, relate to each other and affect vocabulary knowledge in general, it remains the most detailed word-knowledge framework presenting a convenient description of the possible range of knowledge dimensions learners can possess about words.

Using Nation's model of vocabulary knowledge as an initial point and accepting González-Fernández's (2022) claim that the fundamental principle underpinning the structure of vocabulary learning is unidimensional, this study employs false friends and collocations as interconnected aspects.

On another note, this study attempts to explore how a range of factors such as gender, length of exposure to the English language through different activities, and perceived proficiency may affect collocational proficiency (González-Fernández and Schmitt, 2020). In this regard, these aspects are perceived as parts of the conceptual framework adopted in this thesis.

In the context of different languages, the existence of similarities between words at different levels constitutes lexical resemblance at the levels of both meaning and form, resulting in true cognates/friends. Lexical resemblances between languages can be a double-edged-sword for learners. In the case of true cognates, it causes positive transfer, while differences in either form or meaning may cause negative transfer. In this regard, Ringbom (1987, 2001) classifies negative lexical transfer into two types: formal and semantic. The former refers to a situation where the form (spelling) of the words gets transferred, as is the case in false friends, while the latter refers to transmitting the semantics and lexical associations of the SL word to the TL form and is apparent in calques and collocations. In addition to being a feature of interlingual lexical differences between languages (Harouni, 1998), false friends and collocations are also two facets of negative lexical transfer (see Figure1). Therefore, combining these aspects will result in collocational false friends, which, in turn, can uncover a potential network of false collocative translation equivalents to which students are susceptible. This study, thus, claims that collocational false friends are a more comprehensive means to capture *negative lexical transfer* in translation than each of the two aspects, false friends and collocations, separately. The word 'negative' is not universally accepted in describing non-positive transfer, and other terms are sometimes preferred, the most precise (and lengthy) of which is 'absence of relevant concrete positive transfer'.

Linguistic transfer, be it positive or negative, is often studied within the context of L1 influence on L2 learning. However, this can be extended to include L2 influence on L3, as is the case in the study at hand. Only a few studies have tackled the influence of L2 on L3 learning either in terms of positive or negative transfer using different approaches/ methods. Unlike positive transfer which is easy to define but challenging to identify in practice (Woll, 2018), negative transfer is more difficult to define but more evident in practice in EFL and translation learners' production. This study identifies the negative interference of French as L2 in English as L3 triggered by false friends through the lens of collocations.

The following study thus highlights the correlation between a non-native language (French) and third language L3 (English) by Arabic-speaking Algerian students of translation where instances of L2 negative lexical transfer are operationalized by wrong translations that can be traced to an association with a French cognate/false friend. Before explaining and reviewing false friends, along with other related terms,

as focal constituents to collocations in this thesis, the central and holistic term 'collocational false friends' is first explored in the next section of this chapter.

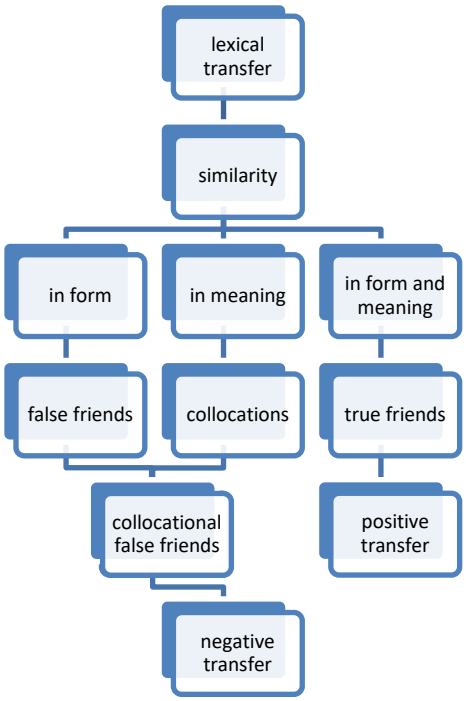

Figure 5: Relationship between lexical transfer, false friends, and collocations

3.5.1 Cognates

Cognates, as defined here, or true friends, are pairs of words that have similar, if not identical in some cases, form manifested through spelling with exactly or roughly the same sense in two or more languages, and which have a common linguistic origin. (There are some other definitions of 'cognate', but these are not of interest to us here.) According to Mitkov *et al.* (2007), the knowledge of cognates is of great importance for learners in terms of expanding their vocabulary repertoire as well as their reading comprehension.

False friends, by contrast, have similar, if not identical in some cases, spelling with clearly different meanings in two or more languages. False friends confuse the reader and create an obstacle for foreign language learners and translation students. In an experiment carried out by Brenders, Van Hell and Dijkstra (2011) on Dutch learners of English as a Second Language false friends proved to complicate the task of recognising cognate forms in both languages English and Dutch. The learners needed more time to react relative to another experiment carried out without implementing

false friends in the task. This is due to the similar way in which false friends are written combined with the dissimilar meaning conveyed by these terms in different languages. Cognates are found in languages mainly because they are the direct or indirect result of the shared proto- or ancestor language from which these terms are originally derived (Beltrán, 2006). They can also be the result of borrowing from one language to another.

Despite the fact that English and French belong to different branches of the Indo-European language family, they share a huge number of cognates (Inkpen, Frunza, and Kondrak 2005; Frunza and Inkpen 2010) mainly because English took in a huge amount of French vocabulary following the Norman conquest of England in 1066. Examples of such cognates are "nature" in English, and "nature" in French; "couleur" and "colour" (see Inkpen, Frunza, and Kondrak 2005).

3.5.2 Types of Cognates

Cognates are of two types: genetic cognates and partial cognates.

3.5.2.1 Genetic cognates

These are pairs of words that are derived from the same proto- or ancestor language but have undergone some phonetic and graphic changes. Genetic cognates exhibit orthographic or/and semantic dissimilarity like "chef" and "head" (Inkpen, Frunza and Kondrak 2005). Borrowed terms are not recognised as genetic cognates.

3.5.2.2 Partial Cognates

Another notion that needs to be explored in the plethora of terms in this study is 'partial cognates'. These kind of cognates are pair of words that share similar meanings in some contexts, where they can be the mutual translations of each other on the one hand, and differ in some others, on the other hand. Inkpen, Frunza and Kondrak's notion of partial cognates is the same as Dominguez and Nerlich's idea of 'partial false friends' and what I will refer to as 'weak false friends' in this thesis.

3.5.3 False Friends

For the purposes of this study, I shall define false friends as follows:

> False friends are pairs of words in two (or more) different languages that exhibit similarities in terms of graphics and/or phonetics, but have some clearly different senses (even if some of their senses may be the same, or virtually the same). The similarities between these words can be measured automatically applying

> various measures of similarity. False friends may occur either by chance (incidental false friends) or due to semantic and/or phonetic changes.

This definition draws on many existing definitions in the literature such as that of Gouws, Prinsloo and De Schryver, (2004 p. 798) who define false friends as follows: "two lexical items from different Languages with the same form but different meanings". With this definition I shall include incidental false friends and weak false friends within the scope of false friends.

3.5.4 Types of False Friends

3.5.4.1 Incidental False Friends

These are false friends by chance, i.e., they share no etymological origin which would ultimately explain the overlap and the similarity in terms of graphics and/or phonetics. They may be regarded as homonyms but in different languages. Homonyms are words that are similar graphically and/or phonetically in the same language (Domínguez and Nerlich 2002).

Some linguists exclude instances of accidental homonymy like "pain" in English and "pain" in French from the scope of false friends and confine false friends to semantic false friends having the same etymology like Roey, Granger and Swallow (1988). Some others, however, consider to be cases of incidental false friends lexical similarities between etymologically distant languages that either occur by chance or by means of lexical borrowing due to cultural and language contact as is the case with English and Arabic (Al-Athwary, 2021).

3.5.4.2 Semantic False Friends

This type of pair, by contrast with incidental false friends, involves words which share the same etymological origin but have acquired different meanings that are linked through figurative relationships, the best known of which are: metonymy, metaphor generalisation, and specification (Domínguez and Nerlich 2002). Semantic false friends can be further divided into two subtypes: strong false friends (also called 'full false friends') and we false friends (also called 'partial false friends').

English and French have some vocabulary used in both languages with slightly or completely different denotations or having different connotations. Since this study focuses on particularly French and English false friends, a list of 365 such pairs has been compiled from *Faux Amis and Key Words* and *Dictionnaire des Faux Amis /*

Dictionary of Faux Amis Français-Anglais English-French (Thody, Evans, and Rees, 1985; Roey, Granger, and Swallow,1998).

Some false friends in different languages have many senses. When the relevant sense of the word in the two languages is clearly different then it is categorised as strong false friend. For the example, the English word 'sensible' is polysemous, as is French *sensible*; but when we compare them as false friends, we tend to take the most common/basic sense in the two languages for comparison, which is very different. By contrast, when the relevant basic sense is similar in both languages despite having at least one sense not in common among other shared senses. The false friend is grouped as a weak false friend.

On the basis of whether the relevant basic sense (meaning) is slightly or completely different in English and French, all the false friends compiled in the Excel file list are divided into two categories: weak false friends and strong false friends. This applies to polysemous words having all the senses in common apart from one as in 'sensible' in English and *sensible* in French, for which the most common basic sense is considered for comparison.

3.5.4.3 Strong False Friends

These are also referred to as 'full false friends' and are considered homonyms in the context of two or more languages. They are pairs of words that are orthographically and sometimes also phonetically similar, but semantically clearly different in all of their senses. Examples of strong false friends are *éventuel, actuel* and 'eventual', 'actual' in French and English respectively. In French, the meaning of the adjective *éventuel* is 'possible' or 'probable' but not 'final', which is the sense of the word in English. Also, 'actual' in English means the real, non-virtual, and existing in fact, rather than *current* which is the sense of the French word *actuel*.

3.5.4.4 Weak False Friends

These are partial false friends and I will refer to them as 'weak false friends' in this work. They have the same or similar spellings but their basic sense in the two different languages under consideration is different. Some of their senses in these two languages may, however, coincide. The different senses reflect the fact that the words have undergone semantic changes over time although they are etymologically related.

The difference can also be noted when two words have related senses but are used in different contexts, i.e. when words have similar senses but different contexts of use like ‘interference’ in English and *interférence* in French. The difference between these words is that in French *interférence* is used only in physics and radio or television waves but not in a political context, unlike ‘interference’ in English. If we want to express the sense of ‘interference’ in English when used in a political context, as in ‘the interference of the government’ we have to use the word *ingérence* in French, thus “l’ingérence de la part du gouvernement” rather than “l’interférence de la part du gouvernement”*. The English word ‘interference’ accordingly shares one sense with French *interférence* but has another sense which is exclusive to English.

This situation, therefore, seems to be trickier with weak false friends not only because they share on or more senses between the two languages, but because language learners may to be exposed to these shared senses before they are exposed to the non-shared senses in the two languages. Weak false friends generally outnumber strong false friends in the different languages and this is what the compiled list of false friends in this research demonstrates.

The following diagram explains the relationship of false friends and the related terms discussed above:

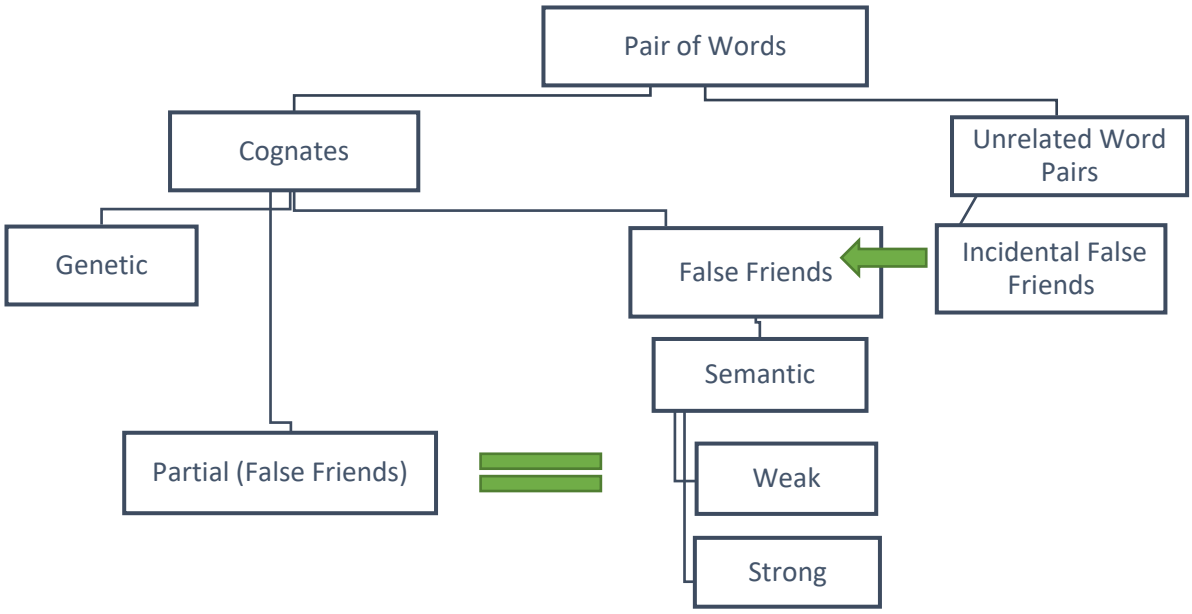

Figure 6: The basis for categorisation of weak and strong false friends

As a thorny linguistic phenomenon that appears in languages with significant overlap in their lexicon like French and English, false friends are confusing for language learners due to their orthographic similarities and semantic discrepancies. They are also difficult for different areas of research such as Natural Language Processing (NLP), Education and translation language studies. As a result, most of the research works that deal with false friends are focused on the identification of false friends and classifying them into true cognates and false friends or differentiating between 'total' false friends that do not share any senses in the two languages and partial false friends, which share one or more senses in the two languages.

False friends research works that are computationally based (NLP) mostly use a two-stage analysis where a combination of orthographic and semantic similarity measures are used to automatically extract language pairs and classify them into true cognates, false friends, or unrelated terms. These results are compared against a manually compiled list of previously judged true cognates and false friends as noted by Fišer and Ljubešić (2013).

3.6 English and French Learners and False Friends

Linguistic transfer, be it positive or negative, is often studied within the context of L1 influence on L2 learning in the literature. Transfer from L2 to L3 has hardly been investigated. Sikogukira (1993) investigated the influence of French on English cognates with an experiment on Kirundi natives. The learners were confused by many false friends (especially weak ones) like *esteem* and *estimer*; *veterinary* and *veterinaire*. Sikogukira (1993) argues that weak false friends are a potential source of difficulty for these learners because they need to expand their knowledge about the other meanings that these words have in the target language although they know their meanings in French.

The false friends in the compiled list of this study may trouble a French learner whose first language is English or who has been widely exposed to English before he or she learns French. These false friends, therefore, may pose problems for the following learners:

- Native speakers of English who are learning French.
- Native speakers of French who are learning English.

- Learners whose L2 is English and who are learning French.
- Learners whose L2 is French and who are learning English.

Algerian students belong to the last category – one of the main aims of this thesis being to examine the effect of French as SL on English as L3 by Algerian Master's students of translation. Both kinds of false friends compiled in this list are highly likely to be a source of confusion for English learners who have a considerable knowledge of French or have been exposed to French before they learn English.

The compiled list of false friends could serve as a seed list to automatically build comparable corpora which could, in turn, be used for automatic training models. For the purpose of this thesis, this list is used as a linguistic resource that can be helpful for educational purposes particularly for Algerian learners of English as a third Language (L3). Since false friends are the result of pairwise comparisons across languages, the compilation in this list can be used to construct a number of bigram[5] collocations. Accordingly, some of the adjectives and nouns on this compiled list will appear in *adjective+noun* collocations. The reason for choosing *adjective+ noun* combinations in this study is that this collocational pattern has proved to be the most challenging type for native speakers of Arabic (Ahmed, 2012; Al-kattan, 2007; Al-Sakran 2011). Moreover, adjectives and nouns exist in all of Arabic, French, and English and this makes the comparison easier. Adjectives in Arabic and English share many common points (function, types, forms, formation) (Messaoudi, 2016), although English and Arabic belong to different language families. In addition to the fact that adjectives and nouns are important components of the language[6], it is in these two grammatical classes where false friends between French and English mostly occur.

3.7 False friends in Research investigating L2 interference in L3 learning

False friends have proved a challenging obstacle as a translation phenomenon as noted by Granger and Swallow (1998) and Chamizo Domínguez, who describes them as "extremely insidious traps" (2006, p. 426). Similarly, they are seen as a hurdle for

---

[5] A bigram is a combination of two words or two adjacent words.

[6] In English web 2015, which has 13,190,556,334 words, there are 678,848 adjectives with a total frequency of 1,122,332,104 and there are 4,491,021 nouns with 4,578,463,504 total frequency

language learning and teaching. Wagner, (2004) for instance views them as "word associations between the language to be learned and another language (often the native language) that are misleading, cause errors and thus not desired" (p. 1). Although the problem of false friends has been recognized since the 19th century (Makayev et al., 2020), practical solutions are rare in the literature. The subsequent section surveys the few accessible studies examining the issue of false friends in the context of L3 learning and translation studies

3.7.1 Martins, and Nunes, 2013

This research aims at identifying cases of transfer of English as a Second Language (L2) in Portuguese learning as a third language (L3) by Chinese learners using an oral sub-corpus of a larger corpus of their oral and written production. The corpus is a recording of narration task to a sketch from Charlie Chaplin's film Modern Times for native speakers of Mandarin with an intermediate level in Portuguese language in their third year bachelor's in Portuguese Language and Culture. These learners are classified as advanced English learners according to Chinese secondary school educational levels. Particularly, the study investigates whether proficiency in English causes transfer at all the linguistic levels (morphology, syntax, and semantics) in Portuguese. It then, looks at whether the cases of transfer at the level of form outnumber those of function; and what kind of transfer is being found positive or negative. The findings show that linguistic transfer occurred at three levels: morphological, semantic and lexical. Most cases of transfer were of the lexical type, which in turn included both content and functional loanwords, cognates, false friends and hybrid forms. The researchers explained the use of L2 lexical knowledge by the students as a resort to compensate for their lack of L3 proficiency "to fill in a gap in their L3 lexical repertoire". They added that this was a conscious move by the learners as hesitation always preceded any production of loanwords "… and that this strategy was a conscious act within their narration, as the use of such loanwords is always preceded by hesitation expressed via the utterance 'ah' and a pause in their chain of speech before inserting the L2 English loanword in the L3 Portuguese sentence" (p. 10).

Instances of use of L2 semantic and morphological rules were also found. Researchers explain the reason for English as L2 being more influential than Mandarin as L1 on the oral Portuguese production by Mandarin speakers, as being the etymological

closeness of English and Portuguese, which partly sharing a common Latin origin. This proximity can be clearly noticed at the phonological level, which led to the production of cognates and false friends.

3.7.2 Boumali, 2010

Boumali's study is an error analysis investigation of the effect of false friends on translating from French to English by a sample of 25 third-year students of translation at Mentouri University, Constantine. Both a questionnaire and a translation test were used. The aim of the former was to test students' knowledge of false friends as a linguistic phenomenon and their proficiency in rendering five strong false friends from French into English through choosing the correct answer. The translation test explored their awareness of the different senses of four weak/partial false friends by using the same false friend in two different sentences with two different senses, thus resulting in eight sentences in all. The results of the questionnaire showed that 64% of the participants did not know what false friends are. Both the questionnaire and test demonstrated that false friends are a critical problem facing students of translation, as most of them could not identify the relevant English equivalent for each French entry (false friend). Although this study was run on a very small sample with few test items, it suggests the importance of false friends in language learning and translation.

3.7.3 Sadouni, 2016

Sadouni wrote a paper that focuses on the difficulties which Algerian students of translation at Algiers 2 University face when translating from Arabic to English, and the influence of French as a second language on the process. Sadouni claimed that there is no unified curriculum in place for lecturers at the Institute of Interpreting and Translation. She categorised mistakes into three main groups: spelling, grammar, and structure. It is in the former that French interference comes into play. Algerian students writing/spelling in English showed a direct influence from French. Grammar mistakes were found in the misuse of prepositions, pronouns, verbs, nouns, and tenses. The latter category included mistakes in which students borrowed an Arabic or/and French structure and applied it to English, such as starting a sentence with a verb or placing the adjective after the noun. The author attributed the failure to teach English to translation students to many factors. These included: 1. the common use of colloquial Arabic and French rather than English in the classroom, resulting in minimizing students' exposure to English; and 2. focusing on teaching English

equivalent words to those of Arabic irrespective of the structure in which these words occur. The author recommended teaching English in primary school along with French and Arabic and adding more hours of English in schools.

3.8 False Friends in research related to Translation Studies

3.8.1 Hilu, 2018

This study examines the issue of false friends (or false cognates) when translating from English to Arabic. Final examination sheets done by 200 first year translation students at Al-Mustansiriya University in Iraq served as data for the analysis. The findings demonstrate that 76.5% of students made mistakes due to false friends. Hilu (2018) categorises the errors committed in relation to false friends into four groups. The first groups includes all instances of English source expressions transliterated into Arabic as target language forms, on the mistaken understanding that they are proper nouns. This type of error was made by 31% of the sample of this study. The second group of errors is associated with the students' confusion of one English word with another English one based on orthographical similarity because of the poor linguistic background of the students. 61.25% of the students in this study produced errors of this second type. The third group of false friends' errors occurring in this study are due to the misunderstanding of the source expressions. This goes back to the students' unawareness of the functions of affixes and how they affect the meaning of words. This leads them to transfer the meaning of the word stem/root into Arabic resulting in erroneous translations. 39.75% of the participants in this study made errors of this sort. The fourth and last type of false friends error emerged from borrowed English words into colloquial Arabic related to social media and TV commercials. Only 8.65% of the students in this study produced errors of the fourth category. The study suggests raising students' awareness of false friends and insights to improve their lexical, semantic, and etymological background knowledge to minimize or eradicate the potential confusion caused by false friends.

3.8.2 Cela Gutiérrez, 2018

This research investigates how evidentiality[7] is manifested in specialized English texts and particularly how evidential adverbs[8] i.e., 'actually', 'evidently', 'in fact', etc. are translated into Spanish and French. The study uses the multilingual UNESCO corpus which is one among twenty-one English-Galician-French-Spanish scientific-technical divulgation corpora of the University of Vigo (CLUVI) to analyse the function and frequency of evidential adverbs and the translation strategies used in rendering them into Spanish and French. There are 148 entries for the adverb 'actually' in the UNESCO Corpus in the English version for which omission, literal translation, transposition and modulation were the most used translation strategies when rendered into Spanish and French. The study also found that there were translation errors in which the evidential marker 'actually' was translated into a circumstantial adverb of time in the target languages French and Spanish as in *en la actualidad and actuellement,* which are false friends. In these cases, the lexical form was not transferred into its corresponding equivalents in the target languages because of the similarity of 'actually' and its corresponding false friends at morphological and phonological levels. This similarity resulted in switching the evidence expression from the source language (English) into a time expression in French and Spanish as target languages.

3.8.3 Anwar A. H. Al-Athwary 2021

Both a theoretical and exploratory study, Al-Athwary (2021), like Hilu (2018), tackles the existence of lexical false friends between English and Arabic as genetically unrelated languages. In addition to a few YouTube videos, the author predominantly depended on his own linguistic repertoire to identify and collect a list of fifty pairs between Arabic and English. Using the taxonomy proposed by Chamizo-Domínguez and Nerlich (2002) of semantic and chance false friends, which in turn correspond to Veisberg's (1996) proper and accidental false friends, the study categorised false friends into semantic and chance categories. The study showed that chance false friends outnumbered semantic false friends, which is reasonable for two

---

7 "Evidentiality is a stance marker of to know how we know, i.e. to indicate the way the speaker has obtained the information or acquired knowledge" (Cela Gutiérrez, 2018, p. 29).

8 "Evidential adverbs are powerful devices to provide evidence supporting the proposition, to reinforce the speaker's attitude towards the information and to keep the intention and effect of the original message unchanged" (Cela Gutiérrez, 2018,p.148)

etymologically unrelated languages. Pronunciation, form and meaning were examined to determine which category these pairs of false friends belong to. Although the author claims that these false friends can cause misunderstanding and confusion to speakers of English learning Arabic or speakers of Arabic speaking English, the study lacks a practical experiment to demonstrate how these pairs confuse learners and at what level specifically. This is due to the fact that confusing the word in Arabic دين /di:n/ with 'dean' in English seem to be particularly relevant only to absolute beginners. This can be explained by learners' awareness of the general distance between English and Arabic (reflecting the fact that Arabic is a Hamito-Semitic language and English an Indo-European one). This reduces the likelihood of any potential hypotheses about resemblances based on the notion of linguistic closeness. Another notable thing is that some semantic false friends according to the author's classification seem to occur more in colloquial Arabic, particularly Middle Eastern varieties, such as the word *receiver* ريسيفر /risi:far/, and are not used in North African Arabic varieties for example.

3.8.4 Roca-Varela 2015

Applying a corpus-based methodology, this study comprehensively examines the 100 most frequent English false friends produced by Spanish students learning English. This is operationalised through using three corpora covering both spoken and written English production by Spanish learners to compare spoken and written English productions of lexical false friends by Spanish learners. The Santiago University Learner of English Corpus (SULEC) containing both written and spoken samples of intermediate and advanced learners was used. The International Corpus of Learner English (ICLE) containing written samples and the Louvain International Database of Spoken English Interlanguage (LINDSEI) containing spoken samples were also used. ICLE and LINDSEI include a collection of texts produced by advanced Spanish learners of English. A total of 1,403 sentences were carefully and thoroughly examined. The findings revealed that false friends were more confusing in the written mode with 23% inaccurate use than in the spoken one with only 16% inaccurate use. It was also concluded that some false friends are more confusing than others. Thus, while *eventually* and s*ympathetic* were shown to be of low-level difficulty, *sensible, actual, argument, inhabited, and comprehensive* were classified as high-level difficulty false friends based on the level of accuracy of use by learners. Mother tongue interference, the semantic content of false friends, the context of occurrence,

and the communicative purpose of the message were all elements that contributed to the misuse of false friends.

The study is part of the book, False Friends in Learner Corpora, in which Roca-Varela (2015) took a novel approach that thoroughly and extensively covered false friends in English as a serious lexical phenomenon facing lexicographers, English language learners and translators. The book has four major parts. The first part reviews the relevance of false friends to different fields of language research such as Translation Studies, Lexciography and Language Teaching. The second part provides an extensive account of definitions and categorisations of false friends. In the third part, a systematic corpus-based method to examine produced English false friends by Spanish learners is reported. The author then discusses the difficulty of these lexical items by designing a chart classifying the 100 most frequent false friends into three groups based on percentages of accurate and inaccurate use. The index difficulty of false friends included in the first group ranges from 0% to 19%. The second set of false friends has an average of difficulty ranging from 20% to 50% based on percentage of errors in use. The difficulty of the highly problematic lexical false friends belonging to the third group ranges from 50% to 100%. Based on the findings of the practical study, the fourth part of the book is confined to presenting a series of practical tasks suggested to teach false friends in ELT classrooms like practice-production tasks, presentation tasks, and knowledge-expansion tasks.

3.8.5 Soglasnova, 2018

This study is based on observations made over several years on materials in Slavic languages by the author who is a practicing cataloguing librarian in a large academic library. While the empirical basis is similar to other studies that dealt with errors in cataloguing language materials in libraries, this study is slightly different in that it focuses exclusively on false friends errors in etymologically related languages. The study examines the linguistic reasons for the existence of false friends. Second, it makes practical suggestions that can help cataloguers and reference librarians avoid the traps of Slavic false friends. The author then examines titles of books to extract the false friends' errors and suggests checking some resources to avoid them such as online translation tools, in particular online lists for Slavic false friends.

While Soglasnova (2018) admits that the chosen examples used are not statistically significant in any way, she stresses that one can spot several obstacles in Slavic

cataloguing due to the presence of Slavic false friends. The author mentioned that false friends in Slavic languages such as Bosnian, Croatian, Polish, and Serbian cover a wide range of phenomena; researchers' criteria for phonetic, graphic, and semantic similarity vary, as do estimates of their number. The author proposes greater flexibility within libraries' organisational systems, which should allow for changes in foreign language materials cataloguing. This can be operationalised through considering the linguistic reality of false friends, particularly in Slavic cataloguing. The strategies suggested by the author, when combined with a continuing commitment to professional development, can yield positive outcomes in diminishing obstacles presented by false friends for subject cataloguing.

3.9 Relevance and gaps in the reviewed studies

Language transfer, particularly non-positive transfer, has attracted the attention of many researchers. However, most of the studies that have dealt with this phenomenon investigated the issue in the light of first language interference on second language learning. Only a few studies have investigated language transfer in the context of L3 learning and this study will constitute an addition to the scarce literature in this regard. Apart from Roca-Varela' study (2015), which examined false friends within their wider context of occurrence and included collocational meaning in its semantic analysis, none of the above studies on false friends considered these pairs in their wider context of occurrence. Thus, they were just treated as separate words for which equivalents were sought. By studying false friends within their larger linguistic environment, including their collocational relationship, this thesis attempts to shed light on one of the aspects that has been ignored in false friends' studies. Moreover, this thesis uses the notion of collocational false friends, a notion that is scarcely covered in the literature, in a slightly different way to how it is used elsewhere.

Rather than being used to refer to true cognates as non-translational equivalents due to their collocational behaviour, the term 'collocational false friend' is used in this thesis to refer to pairs of false friends which used as nodes for collocations. Furthermore, in the current study the investigation of language transfer using collocational false friends is more rigorous and comprehensive than it was in previous studies. Likewise, no previous study, to my knowledge, has investigated the impact of language transfer of the L2 on L3 using collocational false friends in general and specifically in the Algerian context.

3.10 Conclusion

The chapter was divided into three sections. In the first section, a proposed typology of eight translation strategies was introduced. The section also considered how the terms ‘strategy’, ‘procedure’, and ‘technique’ are used, confusingly, in the literature by different theorists. A proposed ontology (set of categories) for translation strategies was developed through a comparison between the different reviewed typologies coupled with observed emerging patterns of solutions adopted by the participants of this study.

The second section of the chapter explained the concept of language transfer and how non-positive transfer is to be examined in this thesis using collocational false friends. Different aspects of language transfer, particularly in relation to false friends, were discussed and an overview of false friends and their classifications was provided. The compiled list of false friends was also described and its potential uses were discussed. The third section of this chapter surveyed studies that have examined false friends as a phenomenon in the context of translation and second language interference in third language learning.

4 Chapter four Research methodology and procedures

Introduction

This chapter discusses the research design used for data collection in this research. First, it presents the aim of the study and the research questions. It describes thoroughly the research instruments and how they were devised, the setting of the whole study and the subjects who participated in the pilot and main study. It highlights the different paradigms of research in the social sciences and explores the mixed-methods approach and the reasons why this is most suitable in this study. The subsequent sub-section covers the obtaining of ethical approval and its role in the process of data collection. The last section discusses the process of conducting the pilot study and outlines the scoring and the coding of data.

As noted before, in chapter one, one of the aims of this research is to highlight the influence of adjectival false friends between French and English on the translation of English collocations from and into Arabic by Algerian students, given that they were exposed to the French language before English. To this end a test involving translating English collocations into and from Arabic is devised. The collocations chosen to be translated in this test are *adjective+noun* collocations, the adjectives of which are false friends with corresponding French adjectives. The reason behind choosing *adjective+noun* combinations in this study is that this type of collocational pattern proves to be the most challenging type of collocation for native speakers of Arabic according to previous studies (Ahmed, 2012; Al-kattan2007; Al-Sakran, 2011).

Before discussing how the collocations used in the translation test have been constructed, the following section will first explain how a list of false friends between English and French languages has been generated. It then demonstrates the different categorisations of the false friends compiled in the list.

4.1 Categorisation of False Friends

A list of thirty adjectival false friends used in this thesis came from a compiled list of false friends comprising approximately 365pairs , taken from *Faux Amis and Key Words* and *Dictionnaire des Faux Amis/Dictionary of Faux Amis Français-Anglais English-French* (Thody, Evans, and Rees, 1985; Roey, Granger, and Swallow,1998).

The words are classified into different grammatical classes (verbs, nouns, adjectives, adverbs), and the following broad semantic categories:

- Human Character, Emotions and Feeling.
- Politics, Business and Economics.
- Education, Media and Art.
- Justice Law and Administration.
- Aspects of Life, Society, Health, and Nature.

To distinguish these semantic categories, five colours are used on an Excel sheet which I have produced: blue, red, purple, green, and yellow respectively. The last semantic category covers all words that have a relationship with food, religion, health, housing, transport, clothes and nature. On the basis of whether the meaning is slightly or completely different in English and French, all the false friends on the Excel file list are divided into two categories: strong false friends and weak false friends.

Strong false friends are words that are completely different in meaning, not sharing any similar meaning(s). The difference between the words' meanings in this case does not pose a problem because the boundaries between the meanings of the word in the target language and the source language are obvious. An example of a strong false friend is "deceiving" and "deçevant". In French the word simply means 'disappointing'. However, "deceiving" in English can mean betraying, misleading, or fooling.

For the first part of the test consisting of English collocations, twenty strong adjectival false friends were randomly chosen to serve as nodes of the twenty English collocations.

	Type 1	Source	Type 2	English	French	Meaning of French	Semantic Category
1	Strong	Thody& Evans	ADJ	Actual	actuel	Current	Politics, Business, and Economics
2	Strong	Thody& Evans	ADJ	Candid	candide	naive or ingenuous but not honest or frank	Human Character, Emotions and Feeling
3	Strong	Thody& Evans	ADJ	comprehensive	comprehensif	understanding et sympathetic	Education, Media, and Art
4	Strong	Thody& Evans	ADJ	concurent	concurent	simultaneous in english but competing or rival In French	Education, Media, and Art
5	Strong	Thody& Evans	ADJ	consistent	consistant	solid, thick,substantial but not coherent	Education, Media, and Art
6	Strong	Thody& Evans	ADJ	eventual	éventuel	possible	Education, Media, and Art
7	Strong	Thody& Evans	ADJ	excited	excité	aroused	Human Character, Emotions and Feeling
8	Strong	Thody& Evans	ADJ	fastidious	fastidieux	boring or tedious	Human Character, Emotions and Feeling
9	Strong	Thody& Evans	ADJ	formidable	formidable	wonderful but not powerful, enormous or alarming	Human Character, Emotions and Feeling
10	Strong	Thody& Evans	ADJ	grand	grand	large	Aspects of Daily Life
11	Strong	Thody& Evans	ADJ	rude	rude	harsh, hard or intimidating	Human Character, Emotions and Feeling
12	Strong	Thody& Evans	ADJ	heinous	haineux	full of hate	Justcie, Law and Administration
13	Strong	Thody& Evans	ADJ	inhabited	inhabité	not lived in	Aspects of Daily Life
14	Strong	Thody& Evans	ADJ	jolly	jolie	beautiful	Human Character, Emotions and Feeling
15	Strong	Thody& Evans	ADJ	rentable	rentable	profitable	Aspects of Daily Life
16	Strong	Thody& Evans	ADJ	sensible	sensible	sensitive	Human Character, Emotions and Feeling
17	Strong	Thody& Evans	ADJ	secular	séculaire	happening once every hundered years	Politics, Business, and Economics
18	Strong	Thody& Evans	ADJ	sympathetic	sympathique	nice	Human Character, Emotions and Feeling
19	Strong	Thody& Evans	ADJ	ulterior	ultérieur	not secret but later	Human Character, Emotions and Feeling
20	Strong	Thody& Evans	ADJ	petulent	pétulant	sulky and gloomy or full of energy	Human Character, Emotions and Feeling

Figure 7: The list of the twenty false friends and their semantic categorisation

The meanings of the 20 English false friends used in the study are completely semantically divergent having no significant overlap with French counterparts. In addition, these 20 chosen false friends provide good potential candidates for semantically independent *adjective+ collocations.*

Weak false friends involve similar words that most often belong to the same grammatical class but have slight differences in meaning. This may occur through there being additional meaning(s) of the word in one of the languages besides the shared meaning(s) in the two languages. The verbs "to realise" and *réaliser* are an example of a weak false friend. "To realise" means to become aware of something, and this can be rendered to French as *rendre compte de*. However, *réaliser* in French means often to make, to create, or to produce in addition to its meaning in English which is "to realise".

The difference can also be noted when two words have nearly the same meaning but are used in different contexts, i.e. when words have similar meanings but different connotations like "interference" in English and *interférence* in French, previously discussed in section 3.5.4.4. As noted there, the difference between these words is that in French "interférence" is used only in physics and radio or television waves but not in a political context like "interference" in English. If we want to express the meaning of interference in English in a political context, as in 'the interference of the government' we would say in French *l'ingérence de la part du gouvernement* instead of *l'interférence de la part du gouvernement**. Therefore, the word 'interference' in English shares same meaning with *interférence* in French but has another exclusive meaning used only in English.

This situation seems to be trickier with weak false friends not only because they share many meanings but also because language learners are highly likely to be more exposed to these shared meanings before the different ones in both languages. Weak false friends outnumber strong false friends in the language as demonstrated by the list of false friends compiled for the purpose of this research.

For the second part of the translation test comprising English collocations, a list of ten Arabic nouns synonymous with ten false friends between English and French was chosen to serve as nodes of the Arabic collocations as shown below.

	Type 1	Source	Type 2	English	French	Meaning in French	Translations in Arabic	Semantic Category
1	Strong	Thody& Evans	N	adept/ness	adept	skilfulness	مهارة	Human Character, Emotions and Feeling
2	Strong	Thody& Evans	N	apology	apologie	excuses	أعذار	Human Character, Emotions and Feeling
4	Weak	Thody& Evans	N	argument	argument	evidence	حجة	Human Character, Emotions and Feeling
3	Strong	Thody& Evans	N	deception	déception	Disappointment in French	خداع	Human Character, Emotions and Feeling
5	Strong	Thody& Evans	ADJ	formidable	formidable	wonderful but not powerful, enormous or alarming	كؤود	Human Character, Emotions and Feeling
6	Strong	Thody& Evans	N	impotence	impotence	crippled	عجز	Human Character, Emotions and Feeling
7	Weak	Thody& Evans	N	humour	humour	temper in english	مزاج	Human Character, Emotions and Feeling
8	Strong	Thody& Evans	N	injury	injure	insult not injury	جرح	Aspects of Life, Society, Health and Nature
9	Weak	Thody& Evans	N	professionalism	Professionalisme	also, competence in French	احترافية	Education, Media and Art
10	Strong	Thody& Evans	N	sensibilties	sensibilités	tendencies	توجهات	Human Character, Emotions and Feeling

Figure 8: The list of the ten false friends, their semantic categorisation, and some of their Arabic semantic equivalents

Both strong and weak false friends in this list are highly likely to be a source of confusion for the following learners:

- Native speakers of English who are learning French.
- Native speakers of French who are learning English.
- Learners whose L2 is English and are learning French.
- Learners whose L2 is French and are learning English.

4.2 Collocations: Automatic Extraction

The collocation extraction is two-stage process. The first and main part entails extracting collocations by means of consensus of both Log Likelihood Ratio (LLR) (Dunning, 1993) and Log Dice (LD) (Rychlý, 2008), which are association measures which will be explained and detailed in the subsequent section. Opting for the latter is a natural extension of the theoretical framework of this thesis which adopts the statistical approach as a starting point. The second stage analysis is used when agreement between the association measures particularly Log Likelihood and Log Dice cannot be reached. In this case the bilingual lists are used to highlight collocates belonging to any shared semantic category between collocates in English language and French. This step extends the idea of the statistical approach used in the first stage analysis to two languages. The bilingual sketch function allows for analysing the mutual relation between the words in different languages through choosing a target term in a target corpus for the source term in the source corpus (Kovář, Baisa, & Jakubíček, 2016). The merged sketch list shows a bilingual list of candidate collocates

next to each other sorted by Log Dice (Rychlý, 2008), which is the default measure using in scoring collocational strength in word sketches.

The use of word sketches is an operationalization of using a restricted bilingual context to show how the two node words behave collocationally in the two different languages, English and French. This is on the premise that collocations and pairs of translation in an aligned bilingual corpus are often used to set specific constraints on word occurrence and meaning as in sense disambiguation, which is based on the idea that words surrounding an ambiguous word, help disambiguate its sense (McKeown, and Radev, 2000). Therefore, using bilingual word sketches would highlight any potential collocates that belong to the same semantic category which might trigger the occurrence of any lexical transfer.

For the purposes of this study, Sketch Engine has been used for analysing the data set using two large representative corpora for English and French respectively: English Web 2015 (enTenTen15) downloaded by SpiderLing in Nov & Dec 2015 and French Web 2012 (frTenTen12) crawled by SpiderLing in Feb 2012. These are both encoded in UTF-8, cleaned, deduplicated, tagged by TreeTagger. I also used Arabic web corpus 2012 (arTenTen12), which was crawled by SpiderLing in January 2012, encoded in UTF-8, and cleaned, deduplicated and tagged by Stanford Arabic Parser in August 2015.

Before proceeding to the first step of collocation extraction, the notion of association measures is to be explained. In the subsequent section of this chapter, I shall explain what the association measures are and what their practical effects are. However, before discussing the association measures, I will highlight some dimensions of collocability in corpus linguistics in order to pave the way to a better understanding of how these association measures work.

4.3 Collocations dimensions

There are four collocation dimensions which I shall introduce below. It is very important for a researcher to to consider which of them are most pertinent to the particular research task before selecting the association measure.

4.3.1 Raw Frequency

Frequency is one of the criteria that is crucial for the identifying collocations and defining collocability. The frequency dimension is based on the idea that the more frequent the collocation is in a language the more predictable it becomes given that corpora are representative of the language.

The first dimension of collocability is raw frequency, which highlights the repetitive units generally in the corpus but may not be the best predictor of regularity of occurrence and predictability in use (Gablasova, Brezina and McEnery, 2017). This can be shown through the fact that most occurrences of a particular collocation may occur in certain units in a particular text in a given corpus. Frequency, in such cases, cannot be a good indicator of collocability because the collocation is not equally dispersed across the whole corpus and is used only in one particular context. Although, frequency is one of the conditions in corpus linguistics to account for collocability, it is still affected by corpus size and how equally collocations are dispersed across the corpus.

4.3.2 Exclusivity

The second dimension of collocability is exclusivity. Exclusivity stresses the positive relationship between the occurrence of two units in each other's company and the extent to which words co-occur together most of the time. This is referred to as "degree of exclusivity". It is also referred to as "mutual information". Exclusivity is typically measured by the number of times each unit in a bigram appears on its own in the corpus compared with its overall number of co-occurrences with another word (Gablasova, Brezina and McEnery, 2017). Therefore, the more two units co-occur, the stronger their exclusivity is.

4.3.3 Directionality

The third dimension is directionality. Directionality stems from the idea that components of collocations do not attract each other with equal strength; the degree of predictability within one collocation varies from one component to another. This can be explained through the strong predictability of one component in a collocation co-occurring with a specific word, while this specific word does not occur with very high frequency with the first component. An example that clarifies directionality is the word 'affair' is much more likely to occur with 'love' more than the word 'love' does with 'affair' (Brezina, McEnery and Wattam (2015 p. 141). Thus, 'affair' attracts 'love' more than 'love' does 'affair'.

4.3.4 Compactness or proximity

This the fourth dimension of collocability. There are two approaches for determining the proximity of collocations: the n-gram approach and the window approach (Gablasova, Brezina and McEnery, 2017). These are discussed in the following sections.

4.3.4.1 The n-gram approach

The first approach is called the n-gram approach. This highlights all the adjacent words as well as bi-grams as a special case. An n-gram is a string of adjacent words, in which (n) stands for the number of words starting from one word up to n (any) number of words.

4.3.4.2 The window approach

The second approach, is the window approach. This selects adjacent words in a given span or window from left and right of the node or the target word. This approach covers a wider range of pattern possibilities and allows more flexibility for different grammatical patterns to appear than the n-gram approach (Manning and Schütze, 1999). Since the aim of this study is to look for adjacent attributive adjectives that often precede the noun, a window approach of one or two words following and two words preceding the node, has been applied.

4.4 Association measures

The extraction of collocations relies on considering the above dimensions. These dimensions can be represented in a range of association measures. An association measure is a mathematical formula which identifies among different candidates in a corpus the ones forming collocations through calculating some scores expressing the likelihood of candidate phrases to be reliable collocations (Pecina, 2010; Gries, 2013). This can be achieved through comparing the co-occurrence of the components of the candidate collocation with the co-occurrence of the two components together. Association measures can also be used for ranking data and classifying them. According to Pecina (2010), the scores obtained by association measures decide which collocations are at the top of the list. Association scores help as well with setting a threshold to discard all the combinations below them.

In theory, creating an association measure accounting for all the dimensions of collocability measures mentioned above would result in a perfect association measure

in extracting collocations. In practice, however, there is no association measure that covers all the dimensions discussed above due to the very different nature of some dimensions. For the purposes of this study four association measure scores will be considered: T- score, MI, Log Likelihood, and Log Dice, with most emphasis on the Log Dice and Log Likelihood scores.

4.4.1 T- Score

When two constituent words frequently occur separately in a corpus we can expect that their co-occurrence is random and just by chance. Assessing whether or not something is a chance event is one of the classical problems of statistics. For this hypotheses testing is used. This is done through formulating a hypothesis which says that there is no association between the node and its collocate beyond chance and compute the probability *p* that this hypothesis is true and then reject it if the *p* value is too low (typically if it is less than a sig level of 0.05). Assuming the random distribution model for language, the t-test, in its equation looks at the difference between the observed and the expected frequency scaled by the observed value.

T-Score expresses the degree of certainty with which we can argue that there is an association between the words, i.e. their co-occurrence is not random. The normal distribution assumption of the t-test may not be suitable for linguistic data as words may not be equally dispersed across the whole corpus. Therefore, T-Scores are affected in one way or another by raw frequency scores. This is why very frequent word combinations tend to have a high T-Score value even if they are not true collocations. It is also worth noting that the score is influenced by the corpus size and the results cannot be compared across other corpora (Manning and Schütze, 1999).

4.4.2 Mutual Information

This measure is also known as ‘association of strength’ as it measures the dependence of the collocation units through finding out how much information each of the unit provides about the other. So, if one of them is entirely independent of the other one, this will ultimately demonstrate no shared information between the two. Mutual information (Church & al. 1991), therefore, is not a frequency-based measure like T-Score and for this reason, it favours low-frequency collocations. These result in assigning high scores to rare exclusive combinations like technical terms or names. Although mutual information is standardised and comparable across corpora, the scale with which it works does not set a theoretical maximum or minimum value. This

requires significant attention when interpreting high scores as these do not mean necessarily that the collocation is stronger (Gablasova, Brezina and McEnery, 2017).

4.4.3 Log Likelihood

Like the t-test, Log Likelihood is a test of significance. Unlike the t-test, which assumes normal distribution of the language, LLR (Log Likelihood Ratio) assumes a binominal distribution. This association measure is based on "a ratio of two likelihood functions which correspond to two opposite hypotheses. The first hypothesis formalises the independence of one component`s (word's) occurrences against the other component's (word's) occurrences. W1 (word 1) does not necessarily co-occur with W2 (word 2). The second hypothesis stipulates that the co-occurrences of both words are dependent on each other. Log Likelihood score tells us how much more likely a collocation is to occur under one of the hypotheses than the other (Manning and Schütze, 1999, pp.172-4).

4.4.4 Log Dice

Log Dice is a measure of exclusivity. This explains why the mathematical expression of this measure looks similar to that for MI. However, Log Dice was developed to compensate for the bias in favour of rare exclusivity, which is a side-effect of Mutual Information (MI) (Gablasova, Brezina and McEnery, 2017). Log Dice is a modified version of Dice Equation, in which the ratio of co-occurrence to the sum of each word occurrence separately is always very small. When compared with T-Score, Log Likelihood and Mutual Information, Log Dice assumes neither a random distribution model of the language of the T-Score nor the binominal distribution that LLR measures invoke. Also, Log Dice does not penalize frequent word combinations which MI does. This can be shown through the non-inclusion of the expected frequency in its equation. One of the other strengths of Log Dice is standardization and fixedness of its maximum value (14) (Gablasova, Brezina and McEnery, 2017). Being standardised means that the results can be directly compared across corpora. The fixedness of the maximum value means that it operates on a specified scale, the highest score of which is 14. Log Dice has also been useful for translation pairs in machine translation (Petrović, Šnajder and Bašić, 2010).

Association measures are statistical procedures/equations that calculate the strength among nodes and their collocates based on different aspects of co-occurrence

relationship in other words collocational dimensions (Brezina et al. 2015, p. 144). There are dozens of statistical measures. Each of them generates a slightly different list of collocates. In practice there is no association measure that covers all the dimensions discussed and fits all research questions due to the very different nature of some dimensions. This is why an understanding of the reasoning underpinning each of these associations is important to select the one/ones that best highlight/s the collocational dimension we are interested in.

Frequency and exclusivity are two opposing notions at the heart of the proposed definition of collocation in this thesis (as discussed in 2.5) and therefore associations highlighting these two dimensions are of interest in this study. These two represent very different approaches to collocation extraction. While T-Score and Log Likelihood are known as frequency association measures, Mutual Information and Log Dice are typically known as exclusivity association measures. Therefore, positing the four associations along a continuum with exclusivity and frequency at the opposing poles, we would have MI at the far end of the 'exclusivity' continuum resulting in strong and rare collocations and T- Score at the end of the 'frequency' continuum yielding frequent combinations. Log Likelihood seems to be the best among the measures of frequency and Log Dice seems to be the best measure among those for exclusivity. Therefore, the two lists of the top ten collocates in this analysis will be sorted according to these two measures.

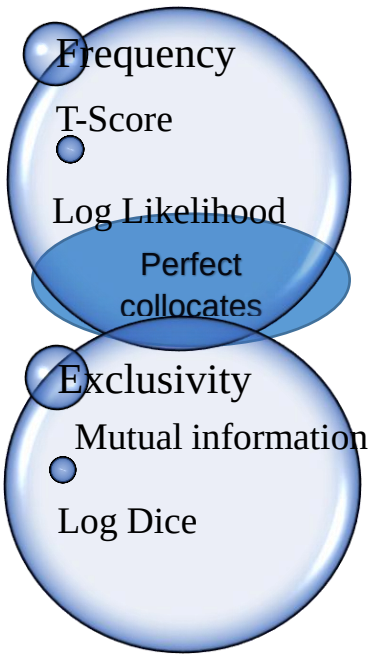

Figure 9 :Relationship between the collocational dimensions and the association measures

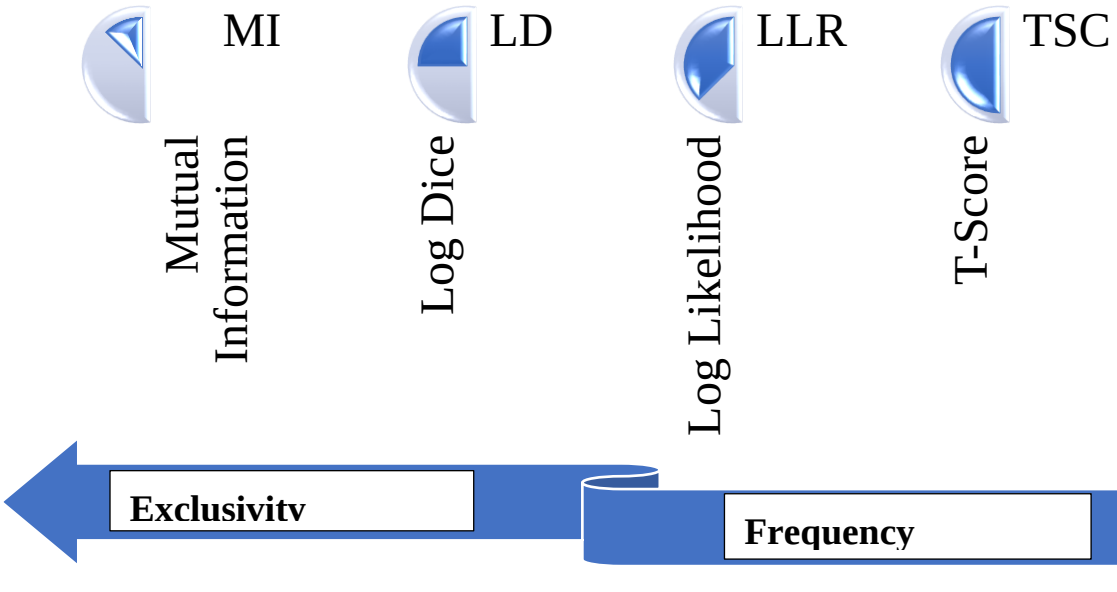

Figure 10: Association measures along the continuum of frequency/exclusivity

The first stage analysis entailed automatically extracting collocations using the T-Score, Log Likelihood, MI, and Log Dice association measures available on the Sketch Engine toolkit to identify both frequent and exclusive collocations. To do this, two chosen lists of focal words obtained from two previously compiled list of friends between English and French (Thody and Evans, 1985; Roey, Granger, and Swallow, 1998**)** was explored using Sketch Engine. The first list comprises 20 English adjectives which are false friends with French. The second list comprises ten Arabic nouns synonymous with English nouns that are false friends with French nouns. For each of the words in the list, the best adjectival collocate was found. Target collocates, which included technical terms, function words and punctuation in this analysis, were considered only if they occurred in the top ten collocates of the Log Likelihood and Log Dice score lists respectively. Moreover, an adjectival collocate was to be considered a perfect collocate if it occurred in both lists with the same rank. The concordance was manipulated using the 'expert' option on Sketch Engine to set specific parameters. This option allowed for imposing various specifications on the lemmas and part-of-speech filter. In this study, the lemma filter was set on both sides

of the node with a window of three tokens. Therefore, after inserting the adjective (node) in the query box and setting specific parameters (lemma filter and part-of-speech filter), the concordance demonstrated many concordance lines in which the node co-occurred. To obtain the most significant and exclusive candidates for the focal adjective, the collocation function, which allows for a choice of specific mathematical metrics of collocability, was used. The noun candidate to be chosen needed to occur with a threshold of top ten collocates in two lists in which the results were sorted by Log Likelihood and Log Dice. The same process using the same criteria was repeated to extract the Arabic collocations. However, unlike in the extraction of the English collocations, the nodes used in the search were mostly Arabic equivalent forms of ten false friends.

4.5 Outcome of the first step analysis

The first step analysis revealed fifteen English collocations that were identified based on the agreement of Log Dice and Log Likelihood scores that they were among the top ten collocates for each adjective.

N	Association / Measures / Collocations	T score	MI	Log Likelihood	Log Dice
01	Heinous crime	88.227	12.374	121.960.295	7.422
02	Sensible	/	/	/	/
03	Actual Cost	88.995	5.642	48.433.387	6.087
04	Concurrent enrolment	66.983	11.434	62.484.954	8.464
05	Candid camera	19.681	10.223	4.728.330	6.183
06	Comprehensive approach	137.653	6.976	148.509.366	7.326
07	Consistent manner	60.222	8.119	33.908.168	6.576
08	Eventual winner	57.106	8.987	34.373.167	6.380
09	Excited anticipation	19.354	10.792	4.866.526	6.313
10	Fastidious	/	/	/	/
11	Formidable opponent	35.638	11.458	17.690.542	7.538
12	Rude awakening	49.293	14.641	44.810.151	9.648
13	Grand prize	132.041	10.282	215.334.403	9.035
14	Inhabited Island	2.445	9.395	66.267	6.144
15	Jolly	/	/	/	/
16	Rentable	/	/	/	/
17	Secular humanism	45.481	13.153	33.780.162	8.548
18	Sympathetic ear	37.348	9.110	14.915.143	6.139
19	Ulterior motive	75.497	16.426	125.291.875	10.080
20	Petulant	/	/	/	/

Table 4: English collocations extracted by the agreement of Log Likelihood scores and Log Dice

The following paragraphs discuss the identified collocates.

4.5.1 Heinous crime

Sketch Engine concordance was used to look for the best noun collocate for the adjective ‘heinous’ within the top ten collocates at a span of one word on both sides

of the node (3 tokens around the node word). This showed that the word "crime/crimes" is the strongest collocate for "heinous". All the chosen association measures assigned the highest scores to 'crime' – as the most frequent collocate using T-Score, strongest using MI, most significant using Log Likelihood and most exclusive using Log Dice. In this case, it is legitimate to describe 'crime' as a valuable noun collocate for the adjective 'heinous'. The results have been sorted according to Log Dice scores in Table 05 and by Log Likelihood in Table 06 as shown below.

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
crime	7,815	1,468,677	88.385	12.365	122,263.83	7.427
offence	227	214,515	15.052	10.035	2,707.38	4.996
atrocity	114	98,734	10.667	10.16	1,378.83	4.993
crimes	42	43,643	6.473	9.898	492.533	4.467
deed	147	212,240	12.106	9.423	1,628.08	4.383
massacre	104	158,361	10.182	9.346	1,140.54	4.267
act	2,671	4,998,095	51.584	9.049	28,580.86	4.124
murderer	44	68,860	6.622	9.307	479.987	4.043
sin	292	575,823	17.054	8.973	3,054.48	4.008
offense	122	257,473	11.021	8.875	1,258.54	3.856

Table 5: Top ten collocates for 'heinous' sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
crime	7,815	1,468,677	88.385	12.365	122,263.83	7.427
act	2,671	4,998,095	51.584	9.049	28,580.86	4.124
most	1,807	16,233,850	42.123	6.786	13,597.27	1.865
this	2,103	87,053,008	43.943	4.582	9,556.59	-0.337
these	1,055	21,913,381	31.8	5.576	6,149.46	0.656
commit	581	1,589,096	24.037	8.501	5,707.42	3.565
a	2,221	329,105,715	40.084	2.742	4,868.21	-2.177
such	733	15,333,195	26.502	5.566	4,249.78	0.645
attack	487	2,638,617	21.947	7.515	4,118.18	2.586
the	3,032	936,353,281	37.912	1.682	3,156.04	-3.236

Table 6: Top ten collocates for 'heinous' sorted by Log Likelihood

Apart from 'crime' Log Likelihood highlighted three other nouns that are good potential noun candidate for 'heinous'. However, this association measure highlighted six function words among the list of top ten collocates. Like Log Likelihood, T-Score

assigned very high scores to these function words, while Log Dice assigned them very low and even negative scores (e.g. *this* =-0.337; *a* = -2.177) as shown in table 3. This is not noticed in the results sorted by Log Dice scores as all the words in the list are lexical words.

There is a range of different potential collocates that go with the adjective 'heinous' in English as well as the collocates of its corresponding false friend in French *haineux*. One of the shared collocates for both adjectives in the two languages is the word 'crime'. Therefore, we can say in French *crime haineux* as we can say in English "heinous crime". However, the meanings conveyed by the two collocations are not the same. The expression "heinous crime" in English means a brutal or evil crime regardless of the motives behind it. In French *crime haineux* is a crime which is driven specifically by hate so an appropriate translation of this expression would be "hate crime". The noun 'crime' fulfils the criteria set (occurring in the two lists sorted by Log Likelihood and Log Dice in the same rank). It is then legitimate to describe it as a perfect collocate.

4.5.2 Eventual winner

Among the Top ten collocates for *eventual* sorted by LL and LD the best adjectival collocate in both lists is 'winner', which ranked second in the LL list and fourth in LD list. Both lists show that 'winner'is a good candidate for a noun collocate. The collocation "eventual winner" is both a significant and exclusive according to Log Likelihood and Log Dice.

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
demise	773	79,994	27.789	10.999	10,256.53	7.246
defeat	336	354	18.33	17.617	8,100.31	6.981
ascendancy	364	11,144	19.076	12.756	5,722.89	6.928
winner	3,274	1,188,103	57.12	9.189	35,306.13	6.394
downfall	311	34,980	17.625	10.879	4,072.80	6.387
elimination	519	167,599	22.746	9.357	5,701.21	6.062
consistency	487	182,209	22.029	9.145	5,206.15	5.89
champion	1,217	715,830	34.788	8.492	11,920.37	5.634
successor	339	153,450	18.372	8.87	3,494.36	5.531
game-winner	123	2,304	11.089	13.465	2,056.96	5.499

Table 7: Top ten collocates for ‘eventual’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
the	31,524	936,353,281	152.662	2.834	79,772.19	0.141
winner	3,274	1,188,103	57.12	9.189	35,306.13	6.394
an	7,095	59,415,245	80.902	4.661	32,770.32	1.966
and	12,707	524,893,821	90.75	2.358	22,388.88	-0.334
outcome	1,644	1,501,339	40.371	7.858	14,668.40	5.084
its	3,073	24,486,193	53.35	4.732	14,350.48	2.034
goal	1,844	3,749,840	42.529	6.703	13,522.69	3.977
their	3,317	44,866,060	53.917	3.969	12,155.82	1.273
champion	1,217	715,830	34.788	8.492	11,920.37	5.634
demise	773	79,994	27.789	10.999	10,256.53	7.246

Table 8:Top ten collocates for ‘eventual’ sorted by Log Likelihood

Among the list of collocates of ‘eventual’ sorted by Log Dice scores, “eventual winner” looks to be a very common collocation in the corpus as well with the highest score given by T-Score and Log Likelihood in this list. ‘Winner’ is one of the most significant and exclusive noun collocates for the adjective ‘eventual’ according to the scores given by Log Likelihood and Log Dice respectively. Log Dice lists ‘defeat’ before ‘winner’ due to the fact that it is the strongest collocate in the list due to its rareness in the corpus with a total occurrence number of just 355 in the whole corpus, 341 of these being with ‘eventual’ exclusively. The collocation “eventual defeat” is a good example of a strong collocation and for this reason it is assigned a score of 17,649 by MI. Log Dice, in such a case, proves its bias towards strong collocates at the expense of frequent ones even if it gives a neater list without function words, which Log Likelihood often fails to do. Function words, which are assigned very high scores by Log likelihood, are assigned very low to negative scores by Log Dice. The function word *the* scored 0.141 while *and* scored -0.334. A Google search shows that “eventual defeat” gets a score of 120,000, which is far lower than the number of hits given by Google for “eventual demise” (315,000) and “eventual winner” (856,000).

'Eventual' and *éventuel* are one of the most confusing false friends between English and French. In French, the meaning of this adjective is 'possible' or 'probable' but not 'final', which is the sense of the word in English . For this reason, 'eventual winner' cannot be rendered as *gagnant éventuel*.

The first step analysis also identified ten Arabic collocates that are both exclusive and frequent as shown in the table below. The collocations obtained from this analysis might identify potential difficulties that Algerian students of translation face when translating from Arabic to English.

N	Association / Measures / Collocations	T score	MI	Log Likelihood	Log Dice
21	مزاج متعكر	9.11	18.488	1,980.11	9.147
22	خداع بصري	24.876	9.169	6,663.06	6.592
23	توجهات سياسية	41.867	10.147	21,414.35	6.020
24	عجز مالي	42.641	8.164	17,103.40	7.008
25	عقبة كؤود	24.412	15.285	11,935.70	7.208
26	احترافية عالية	34.136	9.749	13,514.97	6.255
27	مهارة فائقة	45.648	11.796	30,055.65	8.915
28	حجة دامغة	35.089	11.705	17,681.21	7.088
29	جرح نازف	37.905	14.038	25,514.64	8.47
30	اعدار واهية	28.773	13.864	14,330.54	9.151

Table 9: Arabic collocations extracted by agreement of log Likelihood and Log Dice

4.5.3 خداع بصري **xidāʕ baṣarī**

Among the top ten collocates for *xidāʕ, baṣarī* was the best collocate as shown in table 7 and 8 below. The latter occurred in both lists and was ranked first by Log Dice as shown in the table below and fourth by Log Likelihood scores.

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
بصري	621	151,645	24.876	9.169	6,663.06	6.592
اساليب	832	323,831	28.764	8.496	8,154.68	6.153
استراتيجي	295	144,153	17.115	8.168	2,754.73	5.571
استراتيجى	77	8,632	8.767	10.292	945.592	5.217
بصرى	67	6,009	8.18	10.614	852.801	5.073
ضروب	68	30,626	8.219	8.286	645.796	4.632
فنون	110	107,871	10.414	7.163	874.259	4.431
ذاتي	153	192,226	12.258	6.806	1,140.87	4.317
خطة	384	601,703	19.377	6.487	2,696.82	4.25
لفظي	50	31,327	7.039	7.81	441.917	4.178

Table 10: Top ten collocates for the noun خداع xidāʕ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
و	37,645	582,826,773	172.659	3.182	125,696.03	1.081
اساليب	832	323,831	28.764	8.496	8,154.68	6.153
,	6,440	239,679,113	59.008	1.917	8,050.62	-0.184
بصري	621	151,645	24.876	9.169	6,663.06	6.592
.	3,432	140,046,391	41.581	1.784	3,723.21	-0.317
او	1,468	27,663,111	33.179	2.899	3,385.82	0.795
من	3,470	164,164,810	39.086	1.571	3,047.13	-0.53
الذي	1,247	22,854,008	30.71	2.939	2,932.91	0.834
استراتيجي	295	144,153	17.115	8.168	2,754.73	5.571
خطة	384	601,703	19.377	6.487	2,696.82	4.25

Table 11**:** Top ten collocates for the noun خداع xidāʕ sorted by Log likelihood

The adjective بصري *baṣarī* is one of the best collocates shared in the two lists. It is ranked first in the first list, where results are sorted by Log Dice, and fourth in the second list where results are sorted by Log Likelihood. The expression الخداع البصري *xidāʕ baṣarī* is equivalent in English to ‘optical illusion’. As previously noted, function words and punctuation appearing in the top ten list sorted by Log likelihood obtained negative scored according to the Log Dice measure and very high scores according to both T-Score and Log likelihood as highlighted in table 8 above.

The second step analysis was adopted to generate collocates for the five remaining nodes. This criterion for collocate choice here is to belong to the same prominent semantic category in both languages as demonstrated by the bilingual lists of collocates.

4.6 The outcome of the second step analysis

The results obtained from applying this alternative method differed across the five collocates.

4.6.1 Sensible person

The bilingual lists pointed out a shared noun collocate for "sensible person" as shown below. Belonging to the same semantic category is evident through sharing the same collocate for 'sensible'.

sensible *(adjective)* Alternative PoS: noun (freq: 10,593)
English Web 2015 (enTenTen15) freq = 120,304 (6.54 pe

Use another candidate translation: sensé raisonnable judicieux intelligent rat
Click on collocates to access reciprocal bilingual search or find translated collocatio

nouns modified by "sensible"		61.13	nouns modified by "sensible"		43.88
way	3,333	4.19	peau	7,589	10.08
a sensible way			les peaux sensibles		
approach	2,403	4.69	zone	7,387	8.71
a sensible approach			zones sensibles		
thing	2,303	4.13	sujet	4,896	8.38
sensible thing to do			un sujet sensible		
policy	2,255	4.01	donnée	4,352	8.52
people	1,993	2.99	des données sensibles		
A lot of sensible people share our feelings			image	3,322	8
solution	1,714	4.10	image sensible au survol et		
sensible solution			information	3,276	7.10
decision	1,429	4.76	des informations sensibles		
sensible decision			point	3,089	7.07
choice	1,374	5.12	points sensibles		
a sensible choice			âme	2,954	8.94
option	1,146	3.74	âmes sensibles		
sensible option			être	2,414	7.43
person	1,089	4.56	êtres sensibles		
any sensible person			personne	2,399	6.50
reform	858	4.66	les personnes sensibles		

Figure 11: Bilingual word sketch for 'sensible'/ *sensible* in English and French

The list shows that the adjective ‘sensible’ belongs to more than one semantic category on both sides of the list (English and French). However, the fact that there is a shared semantic category in both languages, which is a human characteristic, narrows down the scope of the search. The current bilingual list not only provides the shared semantic category the adjective fits in, however; it also shows a common noun collocate for the adjective ‘sensible’ in both languages which is ‘person’/*personne* in the bilingual list. Therefore, the collocation formed will be “sensible person” in English and *personne sensible* in French.

The bilingual list demonstrates that the collocates of both sides of the list can belong to different semantic categories, one of which might be shared. This is the case with the collocates of ‘jolly’ and ‘petulant’, as ‘person’ and ‘woman’ belong to the same semantic category.

4.6.2 Jolly lady

jolly *(adjective)* Alternative PoS: noun (freq: 17,541) verb (freq: 2,009
English Web 2015 (enTenTen15) freq = 15,232 (0.82 per million)
Use another candidate translation: chef produire ! bon moins entreprise
Click on collocates to access reciprocal bilingual search or find translated colloca

nouns modified by "jolly"		47.43	nouns modified by "joli"		71.76
time	683	1.15	fille	7,515	8.25
a jolly good time			jolie fille		
man	413	2.24	femme	6,900	7.92
jolly man			une jolie femme		
fellow	381	4.66	blond	6,072	8.93
a jolly good fellow			la jolie blonde		
elf	224	7.93	maison	5,950	8.08
the jolly old elf			jolie maison		
good	151	3.13	brun	4,603	8.56
jolly good			la jolie brune		
gent	145	8.26	photo	4,391	7.97
gregarious , right jolly English gent , loved by			jolies photos		
fun	125	3.05	couleur	3,984	7.48
jolly good fun			une jolie couleur		
mood	120	5.26	robe	3,415	7.97
in a jolly mood			une jolie robe		
nick	101	6.55	vue	3,347	7.33
jolly old St. Nick			jolie vue sur		
person	100	1.18	village	3,329	7.50
A lively and jolly person who finds fun			joli village		

Figure 12: Bilingual word sketch for ‘jolly’/ *joli* in English and French

This adjective ‘jolly’ in English issued to mean ‘cheerful’ and ‘funny’. However, *joli* (fem. *jolie*) in French means ‘beautiful’. Thereby, “jolly lady” does not mean in any way a beautiful lady, which is the sense inferred from the corresponding French expression (*femme jolie)*, but, rather, a lady with a sense of humour.

The bilingual list provides insights into one shared semantic category between the two languages which is human character. On the one hand, ‘jolly’ as an adjective in English collocates mainly with ‘man’, ‘fellow’, ‘gent’ and ‘person’, while ‘joli/e’ in French collocates with fashion words (e.g. *couleur, robe, photo*) and with human character words, mainly of the female gender (e.g. *fille, femme, blond(e)*). Therefore,

the generic word ‘person’ or the more specific ‘woman’/‘girl’ can be used to form the target collocation (‘jolly person/girl’).

4.6.3 Petulant child

petulant *(adjective)* Alternative PoS: noun (freq: 70)
English Web 2015 (enTenTen15) freq = 3,430 (0.18 per million)
Use another candidate translation: irritable Hudghton Sindal humeur Chirac courage
Click on collocates to access reciprocal bilingual search or find translated collocations

nouns modified by "petulant"		31.02	nouns modified by "pétulant"		17.73
child	369	2.25	chanteur	13	2
like a petulant child			la pétulante chanteuse		
teenager	37	3.80	blond|blonde	11	2.46
she 's this petulant little teenager . I			blonde pétulante		
brat	22	6.32	jeunesse	10	1.89
petulant brat			pétulante jeunesse		
tone	20	1.46	acteur|actrice	7	2.01
a petulant tone			tube	7	1.59
reaction	19	0.26	rousse	6	4.11
petulant reaction to			imagination	6	2.41
outburst	16	4.84	lutrin	5	8.79
petulant outburst			vivacité	5	4.88
tantrum	13	5.59	comédien	5	2.19
kick	13	2.19	sexagénaire	4	7.49
a petulant kick			brunet	4	3.07
daughter	13	0.63	animateur	4	2.10
petulant daughter			épouse	4	1.70
refusal	12	3.86	infirmier	4	1.59
petulant refusal to			princesse	4	1.48

Figure 13: Bilingual word sketch for ‘petulant’/*pétulant* in English and French

‘Petulant’ and *pétulant* are adjectives used to describe human behaviour in English and French respectively. They do not, however, convey the same sense. ‘Petulant’ is always used to refer to a bad-tempered person who behaves like an angry child. It is for this reason that ‘petulant’ often collocates with ‘child’, as shown by the Log Likelihood scores above. By contrast, *pétulant* in French refers to someone who is vivacious and lively, and does not have the same disapproving connotation that its corresponding English false friend does. Therefore, *enfant pétulant* cannot be rendered as ‘petulant child’.

Most of the words appearing in the bilingual list on both sides can fit into the prominent semantic category of human character (‘child’, ‘teenager’, and ‘daughter’)

and (*chanteur, blonde, acteur/animateur*). Therefore, 'petulant' can collocate with any of these nouns pertaining to the semantic category of 'human/person'.

In the case of some other examples, particularly 'rentable' and 'fastidious', the bilingual list did not evoke any shared semantic category between collocates on both sides of the list, i.e. the collocates appearing in the two languages do not share any generic notion they both fall under. In this case, the chosen noun collocates for 'rentable' and 'fastidious' ('space', 'person') were generic nouns belonging to the prominent English semantic category in the English part of the list.

4.6.4 Fastidious person

fastidious *(adjective)* Alternative PoS: noun (freq: 35)
English Web 2015 (enTenTen15) freq = 4,032 (0.21 per millio

Use another candidate translation: attentif Traité détail séance député aussi vous
Click on collocates to access reciprocal bilingual search or find translated collocations

nouns modified by "fastidious"		24.60	nouns modified by "fastidieux"		33.29
attention	110	2.13	**travail**	782	5.23
fastidious attention to detail			un travail fastidieux		
organism	60	3.69	**tâche**	622	7.46
fastidious organisms			tâche fastidieuse		
taste	60	3.15	**recherche**	256	4.31
the fastidious taste			recherches fastidieuses		
bacterium	47	3.49	**démarche**	157	4.79
fastidious bacteria			démarches fastidieuses		
nature	43	0.72	**opération**	156	4.57
fastidious nature of the			opération fastidieuse		
microorganism	16	4.71	**étape**	104	3.69
of fastidious microorganisms			étape fastidieuse		
preparation	16	0.50	**manipulation**	103	6.24
fastidious preparation			manipulations fastidieuses		
folk	14	0.64	**lecture**	97	4.30
fastidious folk no more			la lecture fastidieuse		
dresser	13	5.07	**calcul**	94	5.38
a fastidious dresser			calculs fastidieux		
eater	12	3.59	**processus**	89	4.04
fastidious eater			un processus fastidieux		

Figure 14: Bilingual word sketch for 'fastidious'/*fastidieux* in English and French

The adjective 'fastidious'/*fastidieux* is one the strong false friends between English and French that do not share any meaning. *Fastidieux* in French means 'hard', 'tough'

and ‘tedious’. *Travail fastidieux* (‘*fastidious* work’) means hard work but not meticulous or fussy work.

All the noun collocates of the French part of the bilingual list describe a task or process. They all, therefore, fall under one semantic category of ‘process’. However, in English, the noun collocates can fall under different categories, the main one of which is human character (‘eater’, ‘dresser’, ‘folk’, ‘nature’, ‘taste’, ‘attention’). The list of the first 10 commonest collocates in the two languages shows no shared semantic category between the languages. The word that represents the predominant English semantic category, at least in the list, is ‘person’.

4.6.5 Rentable space

rentable *(adjective)* Alternative PoS: noun (freq: 105)
English Web 2015 (enTenTen15) freq = 1,894 (0.10 per mi
Click on collocates to access reciprocal bilingual search or find translated collocations

nouns modified by "rentable"			nouns modified by "rentable"		
		62.88			23.41
space	407	2.16	entreprise	2,327	6.35
rentable space			une entreprise rentable		
foot	153	3.62	investissement	1,529	7.46
million rentable square feet			un investissement rentable		
footage	29	2.44	activité	1,101	4.90
rentable square footage			une activité rentable		
pavilion	21	2.65	façon	966	4.60
a rentable pavilion			de façon rentable		
surface	21	0.07	manière	871	4.58
a rentable surface of			de manière rentable		
apartment	17	0.95	solution	859	5.27
rentable apartments			une solution rentable		
bike	14	0.83	croissance	706	6.16
rentable bikes			de croissance rentable		
shelter	14	0.56	affaire	697	5.10
bicycle	11	2.87	une affaire rentable		
rentable bicycles			site	682	3.65
cottage	11	2.09	site rentable		
Hundreds of rentable cottages , villas and			moyen	561	4.58
toilet	11	1.80	un moyen rentable de		

Figure 15 Bilingual word sketch for ‘rentable’/*rentable* in English and French

‘Rentable’ in English is always associated with space that can be rented. However *rentable* in French has a different denotation. It refers to anything which is lucrative.

Entreprise rentable does not mean a (‘rentable company’). Rather, it refers to a company with a good source of income.

The list shows that the adjective rentable collocates mostly with space. Most of the English noun collocates appearing in the list (‘footage’, ‘pavilion’, ‘apartment’, ‘cottages’) are semantically related to the first collocate ‘space’. In the French part of the list, however, ‘rentable’ collocates with words related to a source of income (*entreprise* meaning ‘company’; *investissment* meaning ‘investment’; *croissance* meaning ‘growth’). The bilingual list does not provide a shared semantic category and therefore, the semantic category the adjective ‘rentable’ belongs to in English is prioritised as the main aim is to obtain English collocations. Therefore, ‘space’ will be chosen as the main noun collocating with ‘rentable’.

4.7 Results of the two-stage Collocation Extraction

In total, the automatic extraction of collocations yielded about 30 collocations, 20 of which are English collocations while the remaining 10 are Arabic collocations, as shown in table 12 below. The collocation extraction is a two-stage process. The first and main part entails extracting collocations by means of consensus of both Log Likelihood and Log Dice. This identified 15 English collocations and ten Arabic collocations. The second stage, discussed in the previous section, identified five English collocations. All these collocations are shown in table 12 below.

N	Association / Measures / Collocations	T score	MI	Log Likelihood	Log Dice
01	Heinous crime	88.227	12.374	121.960.295	7.422
02	Sensible person	/	/	/	/
03	Actual Cost	88.995	5.642	48.433.387	6.087
04	Concurrent enrolment	66.983	11.434	62.484.954	8.464
05	Candid camera	19.681	10.223	4.728.330	6.183
06	Comprehensive approach	137.653	6.976	148.509.366	7.326
07	Consistent manner	60.222	8.119	33.908.168	6.576
08	Eventual winner	57.106	8.987	34.373.167	6.380
09	Excited anticipation	19.354	10.792	4.866.526	6.313
10	Fastidious person	/	/	/	/
11	Formidable opponent	35.638	11.458	17.690.542	7.538
12	Rude awakening	49.293	14.641	44.810.151	9.648
13	Grand prize	132.041	10.282	215.334.403	9.035
14	Inhabited Island	2.445	9.395	66.267	6.144
15	Jolly woman	/	/	/	/
16	Rentable space	/	/	/	/
17	Secular humanism	45.481	13.153	33.780.162	8.548
18	Sympathetic ear	37.348	9.110	14.915.143	6.139
19	Ulterior motive	75.497	16.426	125.291.875	10.080
20	Petulant child	/	/	/	/

Table 12: English collocations extracted using agreement of Log Dice and Log Likelihood and bilingual lists

As noted in section 2 of this chapter (the association measures) above, unlike T-Score and Log Likelihood, which are two association measures based to a greater or lesser extent on frequency, Mutual Information (MI) and Log Dice are exclusivity-based association measures to a lesser or greater extent at the expense of frequency. Finding collocates that are both exclusive and frequent is in practice difficult, if in theory possible. In theory and practice, Log Dice tries to compensate for the bias of low frequency, which is the main side effect of MI (Gablasova, Brezina, and McEnery,

2017), by highlighting exclusive collocates, though not necessarily rare ones. For this reason, Log Dice scores do not show as many technical terms as the MI scores do. Therefore, Log Dice can be regarded as the best association measure for exclusivity, and this is why some differences can be highlighted regarding the performance of MI and Log Dice although they involve exclusivity. However, Log Likelihood and T-Score proved to have very similar results due to the very similar approach (frequency) for the best-scoring collocates (e.g. they both render punctuation and function words). However, Log Likelihood does not highlight free-word combinations as T-Score does and, therefore, its scores have more credibility than those of the T-Score. Log Likelihood thereby looks to be the best measure for associations of frequency.

The findings of this study also demonstrate the effectiveness of using Log Dice for the extraction of collocations. Although Log Likelihood is said to be the best measure of all the association measures of frequency, Log Dice outperforms Log Likelihood in terms of precision, as the former performs effectively even on raw (not pre-processed) corpora (Messaoudi, Brierley, and Dickins, 2020).

The second stage analysis is used when the agreement between Log Likelihood and Log Dice cannot be reached. In this case the bilingual lists are used to highlight any shared semantic categories between collocates in English and French. Resorting to bilingual lists aims at identifying how the adjective in English and French respectively occurs in its context and any sharing of the same semantic category between the adjective and its corresponding false friend within the top ten collocates. In this regard, the noun collocate has to be chosen from the same semantic category. This can sometimes be made self-evident in the bilingual list by pointing out the same noun collocate of the English adjective and its corresponding false friend in French. In case the bilingual list does not yield any shared semantic field between the two languages, then the generic collocate, semantically encompassing most collocates and appearing in the English part of the bilingual list, was considered; i.e., noun collocates representing the prominent English semantic category were prioritised.

4.8 Research paradigms

In order to answer the questions raised in this research, a specific approach to conducting the research needs to be determined. In this light, it is worth noting the researcher' standpoint within a fog of different ontologies. The term 'ontology' here

refers to the way a researcher perceives the world and occurring social phenomena. The ontology that a researcher holds is closely related to how he or she constructs knowledge about the social world. This is referred to as their 'epistemology'. There are two main divergent ontologies, namely objectivism and constructivism. These are related to two different epistemological positions: positivism and interpretivism respectively (O'Brien and Saldanha, 2014). Positivism has given rise to a further development, known as post-positivism. These ontologies and their related epistemologies, also called 'research paradigms', are widely used in social sciences (Creswell, 2009). The positivist/post-positivist paradigm underlies the bulk of research using the quantitative approach. It seeks knowledge through objective analysis and aims to attain generalizability. By contrast, the interpretivist/ constructivist paradigm shapes the qualitative approach. This provides deeper analysis and results in hypothesis generation and construction rather than hypothesis testing, as in the quantitative approach (Lund, 2012). Lately, the mixing of quantitative and qualitative approaches has created what is called the mixed-methods approach, embracing the strengths of both approaches and avoiding their respective weaknesses (Tashakkori and Teddlie, 1998) while seeking answers to research questions, on the basis that these should be given priority over philosophical assumptions. By blending both the quantitative and qualitative methods, this research not only tries to escape their respective weaknesses, but also attempts to increase its validity and reliability.

4.9 Mixed-methods approach

A mixed-methods approach can be of two design types, either sequential or concurrent. Sequential research involves the use of qualitative investigation first, followed by quantitative, or vice versa. Sequential research can be explanatory or exploratory. A sequential explanatory design is used when the researcher employs quantitative data in the first stage and then follows up using the qualitative method to further explain the results obtained from the quantitative data. An exploratory sequential design, on the other hand, first explores data qualitatively and then analyses it quantitatively. Concurrent research, by contrast to sequential, uses both quantitative and qualitative methods simultaneously; it is a one-stage data collection process. When collecting both quantitative and qualitative data in one stage and comparing the results obtained from the two types of data after analyzing them separately, the researcher is said to follow a concurrent triangulation design. A concurrent design can

be ‘embedded’ (or ‘nested’). This is when the researcher prioritises one method over the other using the non-prioritised method to answer a different question than the primary one. A concurrent embedded design, whose suitability I shall argue for in this research, then uses predominantly either a qualitative or a quantitative method and uses the other one as a support to the dominant method.

Both concurrent and sequential designs can be transformative, that is to say, they can change the way in which the research subject is viewed. This results in either a sequential transformative design or a concurrent transformative design. The transformative paradigm is commonly used for social justice-related theories and topics (Mertens, 2012).

In the light of this discussion, the researcher espouses a pragmatic paradigm that employs a mixed-methods approach through adopting a concurrent embedded design. A concurrent embedded design is particularly appropriate to my research as it is a one-phase data collection in which the quantitative approach is prioritized in data collection and analysis. This can be demonstrated through the use of Corpus Linguistics as a statistical method to retrieve the collocations used in the translation test, given that Corpus Linguistics is viewed in this thesis as a methodology or a set of research methods rather than a theory itself. This is consistent with the view taken by Pérez-Paredes (2020, p. 9), who proposes that “orpus linguistics be seen and conceptualized both as a research methodology and a set of research methods” (p. 9). This conceptulaisation is also shared by McEnery and Wilson (2001, p. 1), who define Corpus Linguistics as “the study of language based on examples of real-life language use’, and Brezina (2018, p. 2), who refers to ‘Corpus linguistics as a scientific method of language analysis’. Brezina (2018, p.3) states that ‘in essence, Corpus Linguistics is a quantitative methodology; this means that corpus linguistics typically works with numbers which reflect the frequencies of words and phrases in corpora’ (cf. also McEnery and Hardie, 2011). The adoption of the quantitative approach was not just confined to designing the translation test. It was also extended to data analysis through the performing of statistical tests using SPSS. 26 on the data to find answers to one of the main questions and all sub-questions raised.

The qualitative approach was subsequently adopted to seek answers to the rest of the main questions by content analysis, which mainly entailed categorising the produced translations, identifying and understanding the strategies of translations adopted by

the participants, and thereby obtaining an in-depth analysis of how French as an L2 affected their translations of collocations from English to Arabic and Arabic to English.

4.10 Research instruments design

Designing an appropriate instrument for research questionnaires and tests relies on three essential factors: discrimination, reliability, and validity (Bolarinwa, 2015; Field, 2010).

4.10.1 Discrimination

Discrimination is the ability of the items involved in the instrument to assess the participants' construct and clearly discriminate between the two extremes of the scaled answers. This is to be demonstrated through how well the scores of the instrument's items reveal the difference in the construct measured between the different participants. 'Constructs are broad concepts or topics for a study. Constructs can be conceptually defined in that they have meaning in theoretical terms. They can be abstract and do not necessarily need to be directly observable. Examples of constructs include intelligence or life satisfaction' (DSouza, no date). A construct can be used to articulate the cause behind observable behaviours, attitudes, beliefs or opinions.

In our case the construct sought to be measured by the questionnaire is exposure to the English language. In our questionnaire a four-item Likert scale (none =0; less than 1h=1; 1h-2h= 2; 3h; and more=3) included in the questionnaire aims at measuring the degree of exposure to English outside the classroom through determining: 1/ how much time the respondent spends on daily basis chatting with native speaker of English, 2/ how much time is allocated to watching English TV programs daily, and 3/ how much time they spend surfing English websites.

On the basis of the three item-questions above, the following table illustrates how the scores of the exposure questions reveal the differences in exposure between three randomly chosen participants.

	Participant One				Participant Two				Participant Three			
	0h	≤1h	1h-2h	≥3h	0h	≤1h	1h-2h	≥3h	0 h	≤1 h	1h-2h	≤3h
Q1		**ü**					**ü**					**ü**
Q2	**ü**						**ü**			**ü**		
Q3		**ü**			**ü**				**ü**			
Score	2				4				4			

Table 13: Example of score discrimination on the exposure construct between three participants

In scoring 4, the participant one is said to be twice as much exposed to English, through chatting with native speakers of the language, and watching English programmes than participant one. Similarly, participant two is as exposed to English as participant three, although they differ in the manner of exposure to English. Participant three is exposed to English more through watching English programmes than chatting with native speakers. On the other hand, participant two is exposed to English through being equally exposed to both English programmes and through chatting with English native speakers. The 2-score difference between participant one and the other two participants is exactly the same difference in exposure, even if they differ in terms of the time allocated to each aspect of the exposure construct. Equal intervals on the scale thus represent equal differences in the construct measured, this being one thing which makes a good questionnaire (Field and Hole, 2002).

4.10.2 Validity

The notion of validity in research is generally related to the sufficiency of the gathered data's authenticity for drawing conclusions that correspond to the lived world. Here we do want to get into the debate over the existence of a real state, as in positivist/constructivist stances, which make the very potential of validation contested (Tymoczko, 2005). Validity of an instrument depends on whether it really measures what it intends to (Field and Hole, 2002), i.e. to what extent the gathered data can find answers to the questions raised. These questions lead us to conclude that the validation of the research is achieved through validating the instrument and thereby the items making up the instrument.

Many factors can threaten validity at different levels (Frey et al., 2000). I shall mention only the one applicable to this study, namely measurement validity, which entails sampling methods and size, and the content validity of the instrument. First the size of the sample effects the representativeness of the population and thereby the generalizability of the results. Similarly, the sampling method can affect the representativeness of the population and the fair chance of all the individuals being represented within the chosen sample. The questionnaire aims at measuring the exposure to English in and outside the setting of the class through asking about the duration of exposure to related activities in addition to students' self-perceived levels of the languages they have been taught in class. Accordingly, all the questions serve this purpose. To ensure the content validity of both the questionnaire and test, a pretest of the instruments was undertaken by 7 experts, five of them Algerian lecturers at the different institutions in which the study took place. Three of these academics are lecturers at Abu AL Kassem Saadallah Algiers 2 University, while the other two are lecturers at Ahmed Ben Bella University and the High Arab Institute for Translation respectively. The remaining two academics are English native speakers.

4.10.3 Reliability

A reliability test is meant to measure how likely the same instrument or experiment is to generate the same results if the test is repeated under the same circumstances (Field and Hole 2002; O'Brien and Saldanha 2014). Therefore, the core notion of reliability is whether the instrument developed measures consistently what we want it to measure.

Reliability tests differ according to the design and the purpose of the instrument used to collect data. There are different types of reliability: reliability of internal consistency, inter-rater reliability, and parallel form reliability. First, internal consistency reliability tries to measure how different statements within the same Likert scale all serve or contribute to assessing one construct. Therefore, this type of reliability test assesses the extent to which each item holds with the rest of the items of the construct. Cronbach's (1951) alpha is the most frequent way of measuring the internal consistency of a construct.

Second, inter-rater reliability entails administering the instrument to the same respondents at two different points in time and comparing the results to see whether they are consistent or not. Inter-rater reliability which tests whether two researchers

or raters provide the same estimate of the observed phenomenon. Third, parallel-form reliability entails administering two equivalent questionnaires, both intended to measure the same construct and then assessing the consistency of answers.

The aim of the questionnaire used in this study is to create a profile about Master one students of translation in Algeria and to determine to what extent these students are exposed to English within the context of other languages (Arabic both Standard and colloquial, French and Berber). The questionnaire seeks to collect some demographic information in the first part of the questionnaire. The other questions seek answers about length of exposure to English on a daily basis to see their correlation with the students' test scores. In this light, this questionnaire does not assess a latent construct using several items. The questionnaire uses multi-type scale questions where dichotomous/polysemous and open-ended questions are all used to get detailed information about the exposure of each student to English through related daily activities. If the student does not provide correct demographic information besides his mode and length of exposure to English, then there is no other way of getting this information. The other important point is that these kinds of questions are only asked in one possible way. For instance, there is basically only one possible way to ask about the length of exposure on a daily basis to English TV programmes or films, which is along the lines: "How much time, on average, do you spend on watching English TV programmes or films?". Similarly, we cannot resasonably ask, "How many years you have been alive?", instead of "How old you are?". If we do so in an attempt to judge how consistent the participants' answers are, we run the risk of not getting answers because participants are highly unlikely to answer such repeated questions. Therefore, neither measuring internal consistency using Cronbach's Alpha nor using a parallel-form approach would suit our questionnaire. The reliability of this questionnaire is self-evident since it asks for an exact estimate of exposure duration and investigates the shared English-language gets, within a group of other languages (Arabic both Standard and colloquial, French and Berber), inside and outside the class.

4.11 The setting of the study

The setting of this study was three main institutions: Abu AL Kassem Saadallah Algiers 2 University, the High Arab Institute of Translation in Algiers and Ahmed ben Bella Oran 2 University. There were thus two public universities, one in the

capital city and the other one in the west of the country, and one private translation institution. The initial intention was to include as well students from Annaba University to cover the eastern part of the country where the researcher did her studies before. However, due to the limited time and considering how far Annaba is from the other institutions, this was not possible. A total number of 89 Master One students majoring in English translation was targeted including those involved in the pilot study. The participants were chosen randomly based on their presence to their classes. The percentage of participants in each of the above-mentioned institutions was approximately 40% of the population. This is sufficient to generate results that are highly likely generalizable across the three targeted populations. This can be demonstrated using Qualtrics, an online sample-size calculator which identified an ideal sample size of 89 participants for a population size of 162 at 95% confidence level and .07 margin of error (Qualtrics, 2020).

4.12 The sampling technique

The sampling technique adopted is stratified random sampling. This technique entails dividing the population into subgroups which better describe the current situation of our population. This technique suits best the needs of our sampling as our sample comprises subjects from three different institutions with different-size populations. This technique is designed to ensure the accurate representativeness of the different populations especially with the two-layer stratification (groups) that our data involves. The first layer of population stratification (grouping) can be demonstrated through creating a diversified sample of Master One students from three different institutions across Algeria. The second layer of data stratification was within the same university in Algiers, Abu AL Kassem Saadallah Algiers 2 University, and the total number of Master One students was 100. The overall number of students were grouped into five different groups of the same size and the same number of subjects was extracted from each group to create a complete sample of 40% of the total population of Master One students in this university. Therefore, the number of students chosen from Algiers Two University was 40. The total population number of Master One students from the High Arab Institution was 30 students, and from Oran Ahmed ben Ahmed University 42. Thereby the sample number for each of these two institutions was 12 and 17 respectively.

4.13 Piloting of the questionnaire and the test

In devising the questionnaire, the following timing strategy was used to accurately estimate the length of the questionnaire and time needed by participants to finish. This strategy is a simple point system developed by the Versta research team (Versta Research Newsletter, 2011) to predict the survey length. Each question is given a score of points depending on its complexity and then the total number of points is divided by 8, which is the average number of questions that can be answered in one minute. The point system stipulates that a simple question with dichotomous answers gets one point, and open-ended questions are assigned three points. Questions entailing mental calculation get two points, and one point is assigned for each of the set of choices for multiple-choice questions. In order to predict the questionnaire length and average time it would take to fill it in, this point system was used. The questionnaire used in this study has a score of 48 points in total, which when divided by 8 gave 6. The estimated length time the respondents were expected to take to answer the questionnaire was, therefore, 6 minutes.

A pilot study of the questionnaire was conducted with twenty students from the University of Algiers to check the wording of the questions and to determine the suitability and clarity of the statements. The pilot respondents were asked about how clear the questions were, and how appropriate the order of the questions was. In reviewing the answers of the respondents, the researcher noted that they gave additional answers which were not included in the choices for questions 6 and 8. These additional answers appeared repeatedly across the answers provided by the respondents. Therefore, the potential answers to questions six and eight were extended to cover more answers. Questions six and eight ended up having 8 choices each rather than just 3 and 4 respectively, as had been the case prior to the piloting of the questionnaire.

The test was also administered at the pilot stage simply to determine how much time it would take the students to do it. In reviewing the answers of the students who participated in the pilot study, the researcher noted that none of the students translated ‘excited anticipation’. For this reason, this collocation was removed prior to the administration of the test in the main study.

The time for both the questionnaire and the test was recorded; it turned out that the questionnaire did not take more than 5 minutes and the test took 20 minutes.

Therefore, the total time needed for answering both the questionnaire and the test was calculated at 25 minutes. In order to further validate the questionnaire and the test, they were both revised and approved by three experts in Arabic-English translation studies.

4.14 Research ethics

The research is an ethically designed and approved research, approved by the University of Leeds Research Ethics Committee. Prior to any contact with the participants, ethical approval was sought by filling in different forms related to the emancipatory stance resulting from the prospective changes on the population targeted in this research. The application for ethical approval also explained that the subjects would fully understand the activities which they had agreed to take part in. Participants were then provided with:

1/ An information sheet that highlights the main objectives of the study and the researcher who is conducting it in addition to the supervision team and their contact details. The form contains a statement as well on the influence this research may have on the participants in the future. It invites them to voluntarily take part in it with the potential to withdraw at any time prior to the commencement of the data analysis stage and in this case their data would be discarded and destroyed. The participants are then invited to give their consent at the end of the form by signing it.

2/ Consent form which stresses that anonymity is to be maintained throughout the data analysis stage. On this basis, participants were asked to tick the boxes confirming their full understanding of the statements made and the fact that they can ask for clarification from the researcher at any point. The sheet informed the participants about the implications of their participation in the research, and the potential advantages and disadvantages that may arise, as well as the requirements of taking part in the study in addition to the procedures of data protection and confidentiality. No undue influence was exerted on the participants as the researcher has no previously established professional relationship with them. Therefore, there was no impact of power relations within this study.

4.15 Data collection procedures

The researcher contacted the head of each translation institution to get their approval to access the students during their normal class time. Through different means of contact (emails, phone calls and meeting in person), the researcher was able to get the approval from the three heads of the different institutions after explaining that this study is part of the requirements for a PhD degree at Leeds University and the approval of the Ethics Committee at Leeds University has been granted. One of the challenges that confronted the researcher in this fieldwork was the unsure availability of the students due to the political situation in Algeria at that time. To overcome this challenge the researcher got access to the timetables of the students for each of the five groups of the Master One students at Algiers University to meet the students in person and administer both the questionnaire and the test to them after taking permission from their tutors.

Similarly, the researcher directly approached a cohort of nearly forty Master One students at the University of Oran, and explained the instructions to them on the spot. Unlike the first and second universities, immediate direct contact with the participants from the High Arab Institute was not possible. Here, the researcher was asked to submit the copies of the questionnaire and tests to the administration staff to be delivered to one group of thirty students by their teachers. However, at this point the researcher was allowed to access the students personally. This was done through their teachers who had a look at both the questionnaire and the test and approved them being administered to the students. Answering the test and the questionnaire took around 25 minutes, as predicted.

4.16 Error analysis

Errors are an inevitable part of the learning process. One way of developing teaching practices stems from the premise of evaluating the performance of learners and closely examining the errors in their production. Unlike a mistake, an error is understood as a visible and frequent misconception about the target language. The learner cannot notice it and therefore cannot correct it. A mistake, by contrast, is an accidental aberration in the learner' production due to lack of attention; once noticed it can be self-corrected. Focusing on the learner as a generator of language and comparing his/her production with the target language is the essence of error analysis. In this

light, Ellis defines error analysis as: “a set of procedures for identifying, describing, and explaining errors in the learner language” (1994, p. 70). Contrastive analysis and error analysis are two complementary fields; the potential source of difficulties for EFL learners and the hypothetical errors predicted by the former are tested and verified by the results of the latter.

Error analysis is often underpinned by some procedural steps starting with collecting the data and ending with correcting them. Corder (1967) introduced the following steps for error analysis: collecting sample of learners’ writing, identification of errors, description, explaining, and evaluating and correcting of errors. Boumali (2010) reformulated these steps as: data collection, error identification, classification, identification of the areas of difficulties of the target language faced by learners, and remedy. This last is the equivalent of the evaluating and correcting step suggested by (Corder, 1967). Error analysis can be condensed into three main stages: 1. identification of errors; 2. classification and description; and 3. evaluation and interpretation. This three-stage error analysis is analogous to Elmgrab’s (2016) three possible criteria for evaluating translation: identification of errors, description, and explanation.

4.17 Classifications of the answers based on the acceptability scale

Responses to both the test and questionnaire were entered into an SPSS sheet. Then the acceptability of the answers was determined in terms of a general classification of these answers using a developed acceptability scale whereby the produced translations were classed into five separate categories: 1. good collocation and good translation; 2. acceptable translation and acceptable collocation; 3. acceptable translation but inacceptable collocation; 4. acceptable collocation but unacceptable translation; and 5. neither acceptable collocation nor acceptable translation. The five classified categories were transferred to a three-point acceptability scale: acceptable, partially acceptable, and unacceptable, as shown in table 14 below.

	Good collocations and good translations	Acceptable collocations and acceptable collocations	Acceptable translations but unacceptable collocations	Acceptable collocations but inacceptable Translations	Unacceptable collocations and Unacceptable translations
	Correct Answers			Erroneous Answers	
	Acceptable	Partially acceptable		Unacceptable	
Total	1147			960	
Overall number of attempts	2107				

Table 14: Summary of the classification of the students correct and incorrect responses based on the acceptability scale

The acceptability scale developed in this thesis was operationalised through two different methods for judging collocational acceptability: 1. English Web corpus, 2015; and 2. inter-annotator agreement of two native speakers of English. The use of these two methods reflects a hybrid approach of frequency and phraseology towards specifying the nature of collocation. The choice made here is justified on two grounds. First, these two approaches are widely known and influential in the study collocations, there being some points where the two approaches overlap. Second, the two approaches are based on different principles all of which are, from my perspective, crucial to the nature of collocations (for a detailed discussion on these principles see chapter two). I have therefore decided to amalgamate the two approaches by taking the statistical approach as a starting point and complementing it with the phraseological approach to validate our results. This is operationalized through the use of frequency in the first stage of our analysis via using 1. English Web corpus (2015) and Arabic Web (2012) and verifying that the produced translations meet the key criteria of the phraseological tradition by consulting two native speakers of English. Both methods informed my decision and helped build an acceptability scale for the collocations produced by the participants.

4.17.1 Good collocations and good translation

Good collocations are conventionalized expressions that are close equivalents of the source-text collocations, i.e., semantically (connotatively and denotatively) accurate

renditions of the source collocations. Therefore, the target collocations in the test occurred in a linguistic context which ensures the naturalness and authenticity of language use.

Most of correct answers for عقبة كؤود / *ʕaqaba ka'ūd* were considered good collocations and good translations as is the case with 'unsurmountable/insurmountable obstacle'. The rest were simply considered acceptable translations being less accurate than the former and acceptable collocations being less exclusive collocation and therefore, less acceptable like *huge obstacle.*

4.17.2 Acceptable collocation and acceptable translation

What is meant by acceptable collocations are word combinations which are less conventional and less accurate semantically (connotatively and denotatively). The students' correct Arabic translations of 'formidable opponent' were mostly of the second group (acceptable collocations and acceptable translations). These included answers like مواجه قوي / *muwājih qawī* or خصم صعب / *xaṣm ṣaʕb*, which sound less conventional and less accurate than ند لدود / *nidd ladūd*, albeit that they are acceptable Arabic equivalents of the English source collocation.

4.17.3 Acceptable translation but unacceptable collocation

Here, the given answers are denotatively and connotatively close to the source-text collocation. However, the target-text word combination is not idiomatic or conventional. Therefore, the expression may be more like a free-word combination or a clause rather than a collocation. These instances appeared in the rendition of 'fastidious person' and 'rude awakening' as يولي اهتمام لتفاصيل دقيقة / *yūlī ihtimām li adaqq at-tafāsīl* and صدمت عندما تيقنت لحقيقة / *ṣudimt ʕindamā tayaqqant li-ḥaqīqat al-amr* respectively, which were clauses rephrasing and elaborating the meaning of the source collocations. This category mostly involves instances where the paraphrasing strategy is used.

4.17.4 Acceptable collocation but unacceptable translation

The third group comprises acceptable collocations but unacceptable translations – conventional *(adjective+noun)* word combinations that are not equivalent to the source-text collocations, i.e., semantically (connotatively and denotatively) far from the meaning of the source-text collocations. مقاربة بالكفاءات / *muqāraba bil-kafā'āt* (competency-based learning) as an equivalent to 'comprehensive approach' is an

instance of a good collocation in Arabic but an unacceptable translation as it deviates semantically from the meaning of the source collocation.

4.17.5 Neither acceptable collocation nor acceptable translation

The fifth group comprises target-text word combinations that are neither conventional nor semantically (connotatively and denotatively) close to the meaning of the source-text collocations. An example of this category is translating 'rude awakening' as *istīqāḍ faḍḍ* / استيقاظ فظ or صحوة قاسية / *ṣahwa qāsiya* which are non-conventional expressions in Arabic, being the result of a literal translation of the English collocation. Besides the fact that these answers do not transfer the collocativeness of the source collocation in English, they are semantically far from the meaning of the English source collocation.

The first three categories make up 1,147 answers, which are acceptable as translations and to varying degrees as collocations. The two last categories make a total of 960 erroneous answers – these being semantically non-equivalent albeit they are good collocations for the fourth category and unacceptable in terms of both translation and collocation for the fifth category. The errors recorded in the last two categories were further categorized into three groups: lexical, grammatical, and spelling as shown in table 15 below.

Types of Errors	Examples
Lexical	حقيقة مرة / *ḥaqīqa murra* meaning 'bitter truth' for *rude awakening*
grammatical	Refuting* for *refutable argument*
spelling	'professionalisme' for *professionalism*

Table 15: Classification and examples of produced errors

4.18 Scoring procedure

For scoring (evaluative) purposes, the five classified responses were transferred to Dukhali's three-point acceptability scale: unacceptable, partially acceptable, and acceptable. Item responses belonging to the first, second and third group (*good collocation and good translation*, *acceptable collocation and acceptable translation*, and *acceptable translation but unacceptable collocation*) were marked as 'correct' and those belonging to the fourth and fifth group (*acceptable collocation but unacceptable translation*, and *neither acceptable collocation nor acceptable*

translation) were considered erroneous. By classifying the fourth and fifth group as erroneous answers, a category of 'errors' was identified. In the following section, students' attempts are analyzed along the acceptability scale developed in this thesis and their adopted strategies are identified in their endeavor to render English collocations into Arabic and vice versa.

The data collected from the two different parts of the translation test (English collocations to be rendered into Arabic, and collocations to be rendered from Arabic into English) were scored according to a two-stage analysis. As our collected data involves translating collocations, there was not one single expected correct answer for each collocation. The number of correct answers differed from one collocation to another. Some collocations had only two or three correct possible answers. Many other collocations, however, accepted a variety of both expected and non-expected correct collocations. In order to code the data, an initial coding scheme was drawn up prior to the analysis. However, to allow different responses to emerge freely from the data and to cope with multiple correct answers, the initial scoring scheme was developed in line with the evolving data. This process complicated the task of coding the data as it meant the data did not fall restrictedly into a priori established categories. However, it also allowed for a middle ground between a data-driven coding scheme, which runs the risk of being overly wide and hardly quantifiable, and a pre-set scoring scheme, which molds data to fit with the researcher's theoretical assumptions (O'Brien and Saldanha 2014). Data from the first part were scored as follows: 0 = blank, 1= correct answer, 2= incorrect because of French influence, 3= incorrect for other reasons.

4.19 Item Analysis

Item analysis is a procedure used to investigate whether the items of an exam or test are functioning as intended or not. Item analysis can help in understanding the facility value of the whole test through assessing the facility of each item. One of the statistical methods of item analysis is item difficulty index, also termed 'item easiness index'. This focuses on the proportion of correct responses of the participants to each item in the test. It is calculated as the ratio of the number of participants who answered the item correctly to those who answered the question in general. It is a simple calculation but leads to a powerful indication of how each item is treated in the test. It is referred

to as *p* value and can range between 0.0 and 1.0. The lower the *p* value of an item's difficulty, the harder the question is. By contrast, the higher the *p* value of the item is, the easier the question is.

4.20 Item Analysis Procedures

In order to obtain a general distribution of the collocations used in the translation test given to students, an item analysis was run using Excel 2013 and SPSS 26. The results obtained from the item analysis demonstrated that the frequency of the correct collocations ranged from 21% up to 90%. Generally, an average item would have a *p* value falling between .40 and .60. An item whose proportion of being correct is less than 40% is classified as a difficult item. By contrast, an item which is correct with a proportion of over 60% is classified as easier. After running the item analysis, the items of our test were classified into three categories. First, items having a difficulty index .40 or less were considered as 'very difficult'. Second, items whose difficulty index ranged from .41 up to .69 were classified as 'medium' or 'average' items. Third, items with a difficulty index of .70 or more were considered 'very easy'. The following table demonstrates the easiness index of the collocations used in the translation test.

Item number	**Collocation**	**Number of correct answers**	***P* value Item difficulty index**	**Easiness of items**
1.	Heinous crime	85	0.89	easy
2.	Sensible person	20	0.22	difficult
3.	Actual Cost	49	0.54	medium
4.	Concurrent enrolment	28	0.31	difficult
5.	Candid camera	47	0.53	medium
6.	Comprehensive approach	41	0.46	medium
7.	Consistent manner	21	0.24	difficult
8.	Eventual winner	46	0.52	medium
9.	Fastidious person	17	0.18	difficult
10.	Formidable opponent	43	0.48	medium
11.	rude awakening	8	0.09	difficult
12.	Grand Prize	53	0.60	medium
13.	Inhabited islands	31	0.35	difficult
14.	Jolly woman	49	0.55	medium
15.	Rentable space	47	0.51	medium
16.	Secular humanism	39	0.44	medium
17.	Sympathetic ear	40	0.45	medium
18.	Ulterior motive	48	0.54	medium
19.	Petulant boy	9	0.10	difficult
20.	مزاج متعكر	63	0.71	easy
21.	خداع بصري	22	0.25	difficult
22.	التوجهات السياسية	53	0.59	medium
23.	عجز مالي	23	0.26	difficult
24.	عقبة كؤود	34	0.38	difficult
25.	احترافية عالية	42	0.47	medium
26.	مهارة فائقة	47	0.53	medium
27.	حجة دامغة	31	0.35	difficult
28.	الجرح النازف	58	0.65	easy
29.	أعذارًا واهية	54	0.61	easy

Table 16: Adjectival collocations according to their easiness index

According to Thorndike and Hagen's classification (1977), just under half of our collocations (48%) fell under the 'medium' or 'average' category while the remainder were at opposite ends of the continuum ('easy'14%, and 'difficult' 38%) as shown in figure 10. This shows that our test is of a good quality as it accounts for the differing levels of competencies of the learners.

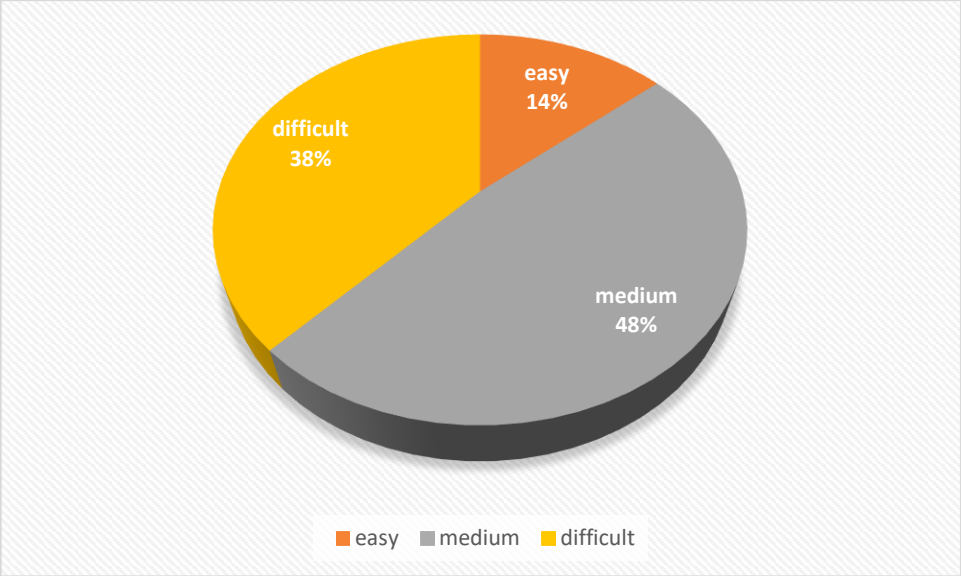

Figure 16: Classification of collocations based on difficulty level

4.21 Conclusion

This chapter specified the research methodology of this thesis. It presented the research design and explained the need for a mixed methods approach to answer the main and sub-questions of this thesis. The chapter also provided information about data collection tools and procedures from obtaining ethical approval to the quantification and classification of the students' responses according to the developed scale of acceptability. The chapter then introduced the adopted research paradigm which inspired the design of the research instruments, particularly the extraction of the collocations that were used in the translation test. The chapter gave a rigorous description and comprehensive account of the two-stage automatic extraction of collocations and their findings. This was done through proving a theoretical introduction to association measures and their practical effects, introducing the steps of the extraction of both the English and Arabic collocations to be used in the translation test. After that, the chapter presented the findings of the performance of the association measures used. It then moved on to providing detailed information about the participants in the study, the setting in which the research was conducted, the sampling technique, and the piloting of the questionnaire and the test, and finally it presented the acceptability scale developed and used to evaluate the students' translations and collocations along with the scoring of the answers.

5 Chapter five results and discussion

This chapter presents the findings of this study and related discussion. It seeks answers to the three main and four sub-questions raised in this thesis. It is divided into two parts. The first reports results for the main and sub-questions raised in this thesis using both quantitative and qualitative analytical methods. The second part follows up the results with possible interpretations of the findings. The chapter begins by summarising the demographic variables from the students' answers to the questionnaire. Then, it analyses quantitatively students' translation scores and qualitatively their responses in the light of the three main questions of this research, these being 1. the effect of students' adopted strategies, 2. French as SL in the overall performance, and 3. the difference in attainment in translating collocations from English to Arabic and from Arabic to English. Then, inferential statistical tests are conducted to examine any potential relationships between the social variables and the students' scores in the context of the four sub-questions of this study. After that, possible explanations are put forward for the findings in the first part of the chapter, and comparisons are drawn with other similar studies.

5.1 Results of demographic information questions

The following section discusses the demographic information in the pilot and main study

5.1.1 The Pilot study

In the piloting stage 20 students took part in this study. All of them were Master One English Translation majors. These participants were from Abu AL Kassem Saadallah Algiers 2 University, where part of the main study took place. Female participants made up 60% of the pilot group and male participants 40%. Their ages were between 21 and 24 years old and the mean average age was 22.5 years old. The table below visualises some descriptive statistics for the pilot stage sample.

Number	Gender				Age		
20	Male	Percentage	Female	Percentage	Minimum	Maximum	Mean
	8	40%	12	60%	21	24	22.25

Table 17: Average ages and the percentage of male and female participants in the pilot study

5.1.2 The Main study

For the purposes of the analysis, the overall sample will include the pilot stage participants as well as the main study sample. The number of females in the main study is more than three times (77.5%) the number of male participants (22.5%). This number reflects the actual proportion of males and females in the Master One class. The ages of the participants range between 20 and 56 years old with a mean average age of 23 years old as demonstrated in Table 18.

Number	Gender				Age		
89	Male	Percentage	Female	Percentage	Minimum	Maximum	Mean
	20	22.5%	69	77.5%	20	56	23.67

Table 18: Average ages and percentages of male and female participants

As displayed in the above table the overall sample had 20 male participants and 69 female participants. Therefore, the overall group chosen for the study had slightly more than 2/3 female students and slightly less than 1/3 male participants. All the participants are Algerians and their native language is Algerian Arabic 89.9%. Only 10.1% of them speak both Algerian Arabic and Berber, as shown in Table 3. They are learning English as a foreign language. All in all, they studied English over a period of 12 years. All the participants have studied English starting from middle school (year 6) for a 4-year period and continued through high school for 3 years and then to university level (3 years). The bachelor's degree type studied by all these students was English. 75% of the candidates were majoring in English, 14% in translation and the rest (10.2%) were taking a degree in either French or Arabic. These students were Master One English translation majors for the academic year 2018/2019. A Master's degree is two years long in Algeria. The rationale for choosing Master One and not

Master Two was fear of non-availability, due to final year dissertation preparations that Master Two Students were occupied with at the time this study was taking place.

	Frequency	**Percent**
Arabic	80	89.9%
A mixture of both Arabic and Berber	9	10.1%
Total	89	100%

Table 19: The first language spoken by the participants

87.6% (78 students) of these students were not taught by native English teachers, while 12.4% (11 students) were taught by native English teachers as shown below.

Native Teacher?	Frequency	Percentage
No	78	87.6
Yes	11	12.4
Total	89	100

Table 20: Participants taught by native English teacher

Among the 11particpants taught by native English teachers, 6 were taught by native English teachers for less than a year and 4 for over a year and less than two years. Only 1 student spent more than 3 years under the tutorship of a native English teacher.

A few participants (17 students) claimed that they had attended training in English in Algeria and somewhat fewer (10) gave insights into the place and the duration of this training. This length of training differed from one participant to another, and ranged between15 days and 2 years.

Training in Algeria?	Frequency	Percentage
No	72	80.9%
Yes	17	19.1%
Total	89	100%

Table 21: Attended training in Algeria

Similarly, there were a very few participants (10) who attended training abroad in English-speaking countries over differing lengths of time, the shortest being 8 days and the longest 2 months.

Training Abroad?	Frequency	Percentage
No	79	88.8%
Yes	10	11.2%
Total	89	100%

Table 22: Attending training in native English-speaking countries

When asked about the language(s) used for surfing, English was the most used language at 47.2%. This was followed by French at 25.8% and to a lesser extent by both languages, Arabic and English (9%), Arabic alone (7.9%) and then by both Arabic and French, as illustrated in table 23 below.

Languages used for surfing	Frequency	Percentage
Arabic	7	7.9
French	23	25.8
English	42	47.2
Arabic-French	5	5.6
Arabic-English	8	9.0
French-English	2	2.2
All above	2	2.2
Total	89	100.0

Table 23: Languages used for surfing

While a few participants (14.6%) were doing translation as a major, most of them were majoring in English (74.2%) as shown in table 24.

Bachelor's degree	Frequency	Percentage
Arabic	2	2.2
French	4	4.5
English	66	74.2
Translation	13	14.6
other	3	3.4
missing	1	1.1
Total	89	98.9

Table 24: Types of undergraduate degree

When asked about the language used as medium of instruction in class, the participants gave different answers. It seems that that there was not one single language used in the classroom and that the three languages, Arabic, French and English, were used to varying degrees. Arabic was chosen as the primary medium of

instruction followed by English and then French. The main language pair the students translate from and into was Arabic-English but it was not the only one as they less commonly translated the language pair Arabic-French as well. All the participants agreed that they translated texts from different fields and not a particular genre.

When asked to judge their proficiency in Arabic, French, and English, over 75% of the participants perceived their level of English as being beyond intermediate and the rest judged themselves as intermediate. None, however, described his/her level as elementary. As far as Arabic and French were concerned, very few participants chose 'elementary' to describe their proficiency level in these languages. Nearly 70% of the sample described their level in Arabic as beyond intermediate. Similarly, 60% of the sample perceived their level of French as beyond intermediate.

5.2 Results of the three main questions raised in this thesis

The following section provides answers for the three main questions raised in this thesis, which are the following:

- How was the overall performance by Algerian Masters' students affected by their adopted strategies in the translation task?
- did the students perform equally in translating English collocations into Arabic and Arabic collocations into English (both Tests)?
- How does Algerian EFL learners' L2 affect their proficiency in translating English collocations into Arabic and Arabic collocations into English?

5.2.1 How was the overall performance by Algerian Masters' students affected by their adopted strategies in the translation task?

In order to answer the above question the overall translation test scores and the adopted strategies are presented below

5.2.1.1 The Translation test scores

The overall scores obtained ranged from 2 to 25 out of 29, which is the total number of collocations. The average score in the test was 12.85. These scores have been normalised as demonstrated in table 25 below.

Number of participants 89				
	Minimum	Maximum	Mean	Standard deviation
Total Scores out of 29	2	25	12.85	5.75
Total Scores out of 100	6.90	89.21	44.32	19.83

Table 25: Total scores for the translation test

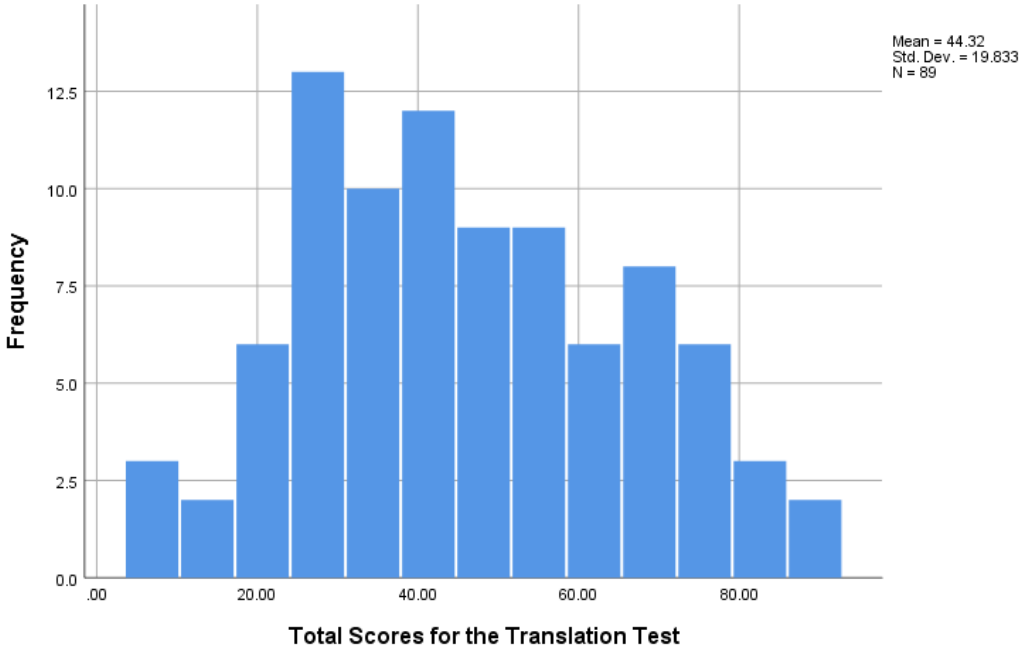

Figure 17**:** Mean percentages of the participants' total scores for the translation test

	Correct answers			Erroneous answers		
Items	Good collocations and good translations	Acceptable collocations and acceptable collocations	Acceptable translations but unacceptable collocations	Acceptable collocations but unacceptable translations	Unacceptable collocations and unacceptable translations	Blanks
1	84	1		2	2	6
2	20			70		1
3	33	16		37	1	3
4		28		16	10	35
5	47			8	11	23
6	41			26	10	12
7	21			47	3	18
8	46			27		16
9	5	10	2	57	1	19
10	2	41		8	25	13
11	6		2	8	35	38
12	52	1		17	2	17
13	31			49		9
14	44	5		23	4	13
15		46		17	7	20
16	39			2	13	35
17	34	6			24	25
18	47	1		23		18
19	7	2		65		16
20	63			3	17	6
21	20	2			47	20
22	47	6		11	20	6
23	18	5		30	23	13
24	20	14			11	45
25	23			16	24	8
26	2	45			31	11
27	30	1		2	37	19
28	58			3	20	8
29	29	25		4	11	20
Total	869	274	4	571	389	493
	1147			960		
Overall number of attempts	2107					

Table 26: Classification of the students correct and incorrect responses for each collocation based on the acceptability scale

The results of this study demonstrate that the translations produced by the students fall under all five categories identified earlier: 1. good collocation and good

translation; 2. acceptable translation and acceptable collocation; 3. acceptable translation but unacceptable collocation; 4. acceptable collocation but unacceptable translation; and 5. neither acceptable collocation nor acceptable translation. The five classified categories are transferred to a three-point acceptability scale: acceptable, partially acceptable, unacceptable.

Combined together the first three categories represent a total of 1,147 correct answers which are all accepted as translations and to varying degrees as collocations. The two last categories make a total of 960 erroneous answers due to being semantically inequivalent albeit good collocations for the fourth category and unacceptable in terms of both translation and collocation for the fifth category. The errors recorded in the last two categories were further categorized into three groups, lexical, grammatical, and spelling, as shown in table 27 below.

Types of Errors	Number	Percentage %
Lexical	874	91.04
Grammatical	52	5.41
Spelling	34	3.54
Total	960	100

Table 27: Classification of produced errors

No	Used translation techniques	Number	Percentage %
1	Literal translation	1174	45.15
2	Omission	492	18.92
3	Synonymy	308	11.85
4	Equivalence	286	11.00
5	Approximation	135	5.19
6	Substitution	126	4.84
7	Reduction	69	2.65
8	Paraphrasing	10	0.38
	Total	2600	100

Table 28: Frequency of strategies used

In this translation task, subjects used different translation strategies, as shown in table 28 above, whose causes might be subject to different interpretations, none of which is

definitive. The participants may have drawn on their first language as it is the first resort for any foreign language learner or at least for their assumptions as to what the target language norms should be. Similarly, they might have lost patience when meeting with difficult items and therefore might have avoided giving any answerers. At some points, they might have attempted to focus on the whole meaning of the sentences rather than the collocations themselves and thereby provided lengthier or shorter answers. At others, they could also have tried to compensate for their inability to recall the target equivalent and then decided to approximate the source meaning either by providing a semantically close candidate, or by recklessly giving completely irrelevant answers.

5.2.2 Did the students perform equally in translating English collocations into Arabic and Arabic collocations into English (both Tests)?

In order to answer the above question scores for translating Arabic collocations and English collocations are presented below

5.2.2.1 Scores for translating Arabic collocations

The normalised scores range from 6.90% to 89.21%. The total score represents the sum of scores for translating the English collocations into Arabic and vice versa.

Number of participants 89				
	Minimum	Maximum	Mean	Standard deviation
Total Scores out of 10	00	10	4.80	2.53
Total Scores out of 100	00	100	47.98	25.28

Table 29: Scores for translating Arabic collocations

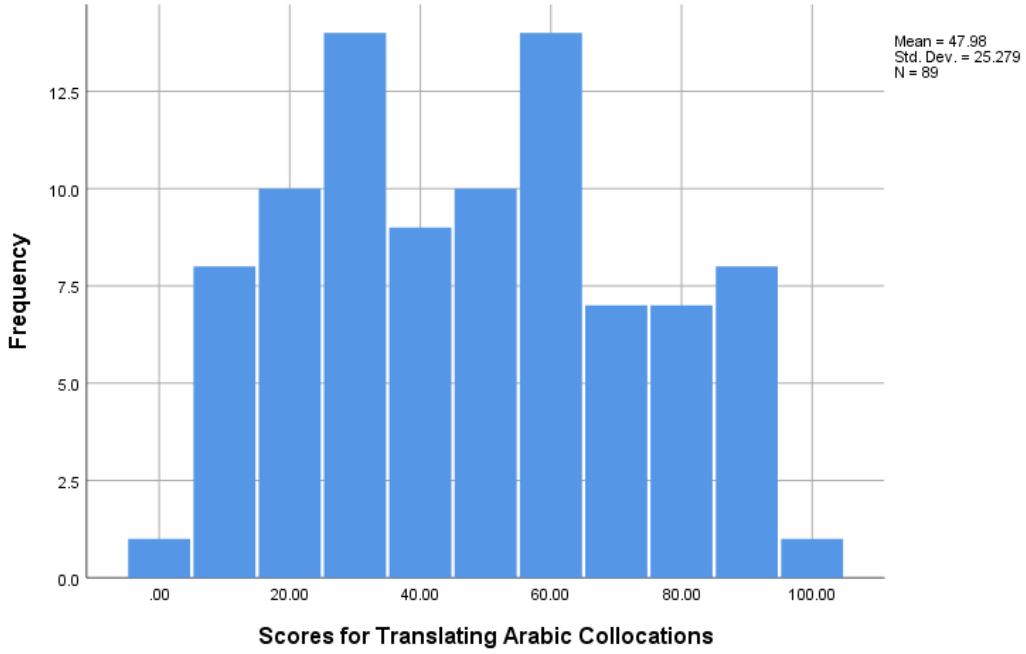

Figure 18 Mean percentages of the participants' scores for translating Arabic collocations into English

5.2.2.2 Scores for translating English collocations

Number of participants 89				
	Minimum	Maximum	Mean	Standard deviation
Total Scores out of 19	1.00	17.00	8.06	3.93
Total Scores out of 100	5.26	89.47	42.40	20.71

Table 30: Scores for translating English collocations

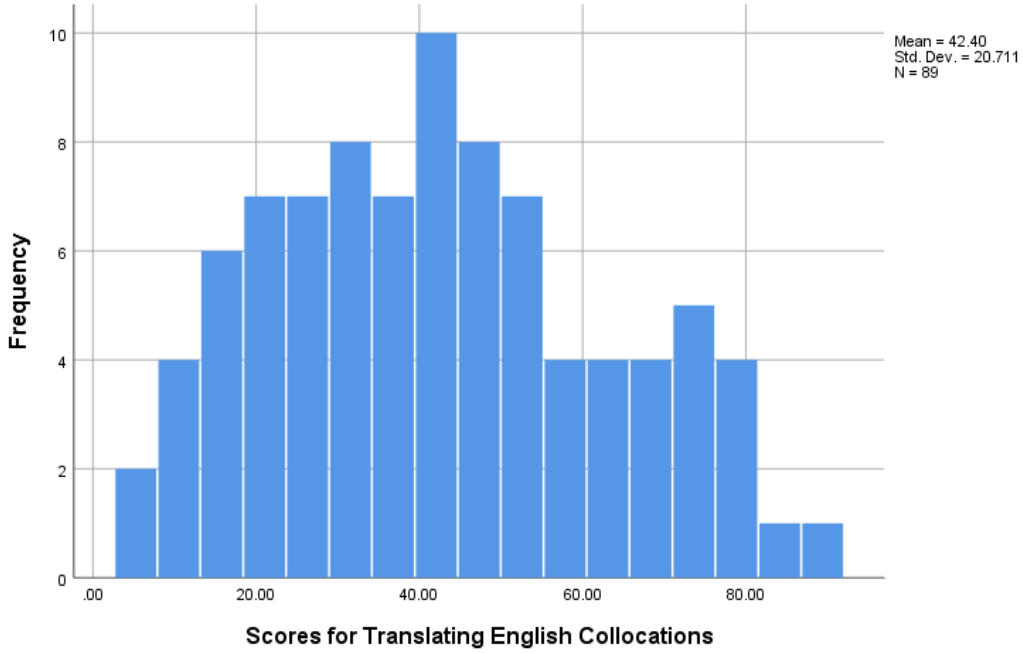

Figure 19 Mean percentages of the participants' scores for translating English collocations into Arabic

	Mean	**Correlation**	**Sig**	**Std. Deviation**
Scores for Arabic collocations	47.98	.56	.000	25.28
Scores for English collocations	42.40			20.71

Table 31: Paired samples statistics

To test the hypothesis that the mean difference in scores for translating English collocations into Arabic (47.94, SD 25.28) and translating Arabic collocations into English (M=42.40, SD 20.40) is statistically significantly different from zero, a dependent samples t-test was conducted. Prior to conducting the paired t-test, the assumption of normality for the difference in scores between the two translation tasks was examined. The assumption of normality for the difference in scores was perfectly satisfied for conducting a paired t-test, as the z-values (which are calculated by dividing absolute skewness and kurtosis values by their standard of error) were estimated at 1.03 and 1.46 for English collocation scores, and 0.59 and 1.94 for Arabic collocation scores (as shown in Appendix F). These are less than the allowable z-

values for normally distributed data (i.e., z -value < | -1.94 | | +1.94 |) (Posten, 1984). The homogeneity of variances was tested and found tenable (F (8,78) = .46 *p* =.88>.05) The correlation between the scores for the two translation tasks was .56 at *p* (.00 <.05). The null hypothesis (there is no statistically significant difference) of equal mean scores for the students in the two translation tasks is rejected, as t= -2.40 (88) at *p* = (.02<.05). Thus, the mean score for the Arabic collocations' translation task was significantly higher than the mean score for the English collocations' translation task. Cohen's *d* was estimated at .25, which is a small effect based on Cohen's (1992) guidelines.

Cohen's *d* for a paired t- test is calculated as follows:

$$d = \frac{\text{t}}{\sqrt{\text{N}}} \text{ Or } d = \frac{\text{Mean}}{SD} / d = \frac{-2.40}{\sqrt{89}} \text{ or } \frac{-5.58}{21.87} = .25$$

5.2.3 How does Algerian EFL learners' L2 affect their proficiency in translating English collocations into Arabic and Arabic collocations into English?

Out of a total of 960 errors occurred in 2,107 attempts. Around a third of theses, which were either lexical or spelling, were possibly due to French interference (309 instances (32.18%), while around two thirds involved L1 interference, overgeneralisation and other errors caused by the adopted strategies.

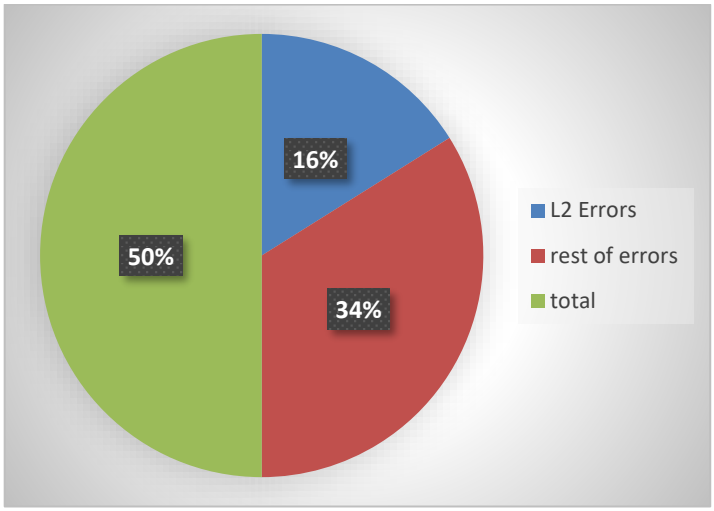

Figure 20: L2 interference errors

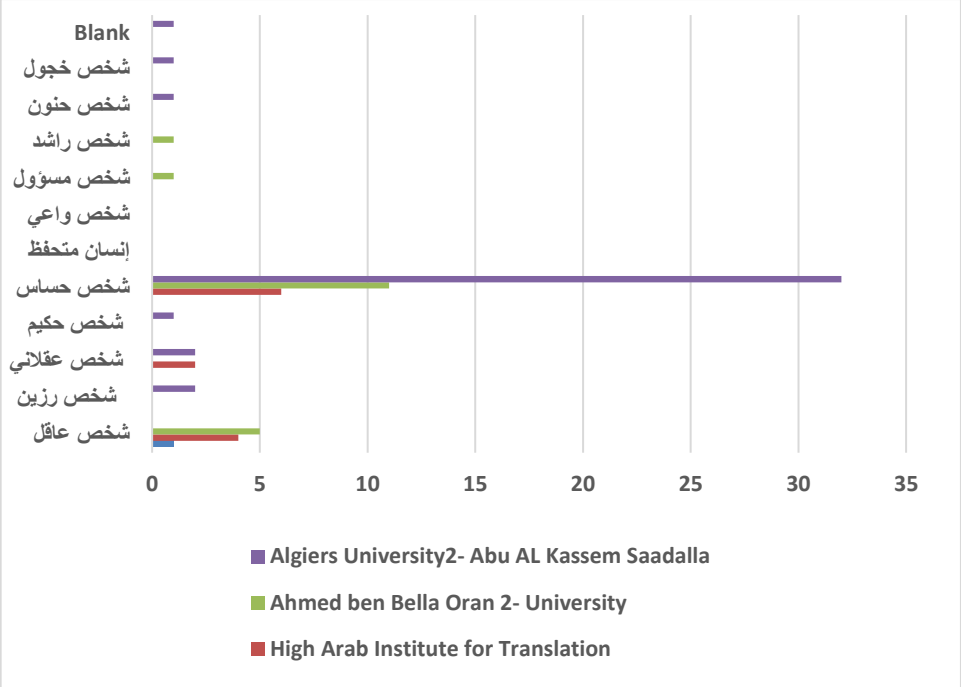

Figure 21: Distribution of translations of 'sensible person' across the three different institutions

Across all the groups from different institutions there were twenty correct answers given by the participants for 'sensible person' and 70 wrong answers. Out of these, 67 were due to French interference. As demonstrated by figure 3, the most frequent answer was شخص حساس / *šaxṣ ḥassās*, which is a result of a direct negative lexical interference from French.

5.3 Results of the Sub-questions Raised in this Thesis

The following section discusses the answers for the sub-questions raised in this thesis which are the following

- Is there a significant statistical difference between the performance of males and females in producing collocations in English?
- Did the students from the different institutions perform equally in the translation test?
- Is there a statistically significant difference in the participants' scores based on their amount of exposure to English?
- Is there a statistically significant difference in the participants scores based on their self-perceived level in Arabic, French and English?

5.3.1 Is there a significant statistical difference between the performance of males and females in producing collocations in English?

To compare the performance of the participants in relation to their gender, we measured the difference in the translation scores between females and males conducting an independent samples t-test. Before conducting the independent t-test the normality distribution of the scores of both groups female and male was examined using skewness and kurtosis levels. Additionally, the variability between the two groups was checked using Levene's test for equality of variances. For our data, the z-values (obtained by dividing the absolute skewness and kurtosis levels by their standard of error) for male scores were estimated at .42 and -.56 respectively. For female scores the z-values were estimated at .80 and -1.31 respectively. These values are less than the permissible values for z-values .42 (i.e., skew < |-1.94 | and kurtosis | +1.94 |) (Posten, 1984). Levene's test for equality of variances was estimated at F=.85, (p = .77>.005). Thus, the homogeneity of the variances is assumed.

Gender	N	Mean scores of the translation test	Std. Deviation
Male	20	45.69	20.68
Female	69	43.93	19.72

Table 32: Mean scores for males and females

The table shows that there is no significant difference between the mean scores of the females (M=45.69, SD=20.68) and males (M=43.93, SD=19.72). Similarly, the independent T-Score test yielded t= .35 (87) at (p=.73 >.05). This proves that gender has no relationship with the performance of the participants. The conducted independent t-test demonstrated that there is no significant difference between the performances based on gender. In order to specify, the power of the t-test in detecting non-significant difference and to determine how real is the difference we calculate the Cohen's d effect size for the independent t-test, which is given as follows:

$$d = \frac{\mathrm{M1-M2}}{\sqrt{S^2 pooled}} = \frac{\mathrm{M1-M2}}{\sqrt{\frac{(n1-1)SD^2 1+(n2-1)SD^2 2}{(n1+n2)-2}}} = \frac{\mathrm{M1-M2}}{\sqrt{\frac{(n1-1)SD^2 1+(n2-1)SD^2 2}{(n1+n2)-2}}} = \frac{45.69-43.93}{\sqrt{\frac{(20-1)20.68^2+(69-1)19.72^2}{(69+20)-2}}} = .09$$

Cohen's d was estimated at 0.09 which is, according to Cohen's (1988) conventions, less than even a small effect size difference between the average scores of the female and male participants. Equally this can be understood as follows: 9% of the variability in scores of the participants can be explained by their gender. Thus, this proves that the females' scores were not associated with statistically significantly larger mean scores than the males'.

5.3.2 Did the students from the different institutions perform equally in the translation test?

To answer the question of whether the participants from the different institutions performed differently, a one-way ANOVA was conducted after checking the normality of the scores in each institution and homogeneity of variances. The homogeneity of variances was tested and found tenable (F(3,85)= .42, p=.74>.05) and the normality of the participants scores from the different institutions were assumed (see Appendix Appendix F).

	Sum of squares	df	Mean Square	F	Sig.
Between Groups	8858.360	3	2952.787	9.74	.000
Within Groups	25755.936	83	303.011		
Total	34614.295	85			

Table 33: Tests of between-subjects effects for one-way ANOVA (for institution)

The table above shows the output of one-way ANOVA analysis and whether there is a statistically significant difference between our group means. We can see that the significance value is .001 (i.e., p = .001), which is below .05. Therefore, Results of the between subjects from the one-way ANOVA test demonstrated that there is a statistically significant difference among these three groups at (F (3) = 9.74 p = .000).

Institution	Number	Mean	Minimum	Maximum	Std. Deviation
Pilot Sample (Algiers 2 University)	20	30.34	6.90	72.41	16.35
Abu AL Kassem Saadallah Algiers 2 University	40	44.13	13.79	79.31	17.57
High Arab Institute for Translation	12	64.37	31.03	86.21	16.06
Ahmed ben Bella Oran 2 University	17	47.06	13.79	86.21	19.04
Total	89	44.32	6.90	86.21	19.83

Table 34: Mean scores for the translation test across the different institutions

To find out whether the difference in mean scores across the different institutions the Tukey post-hoc test was run. It showed that the difference in mean scores for the students at the High Arab Institute and the pilot students was estimated at (34.02 p =.05) ,which means that the difference was statistically significant. The mean score for students of the High Arab Institute was also higher than the mean scores obtained from both the universities of Algiers and Oran with the following means of difference (20.33 p=.004), and Oran (17.31 p = .05). By contrast, there was no statistically significant difference between the mean scores of Algiers' students and those of Oran University (2.92p = .94).

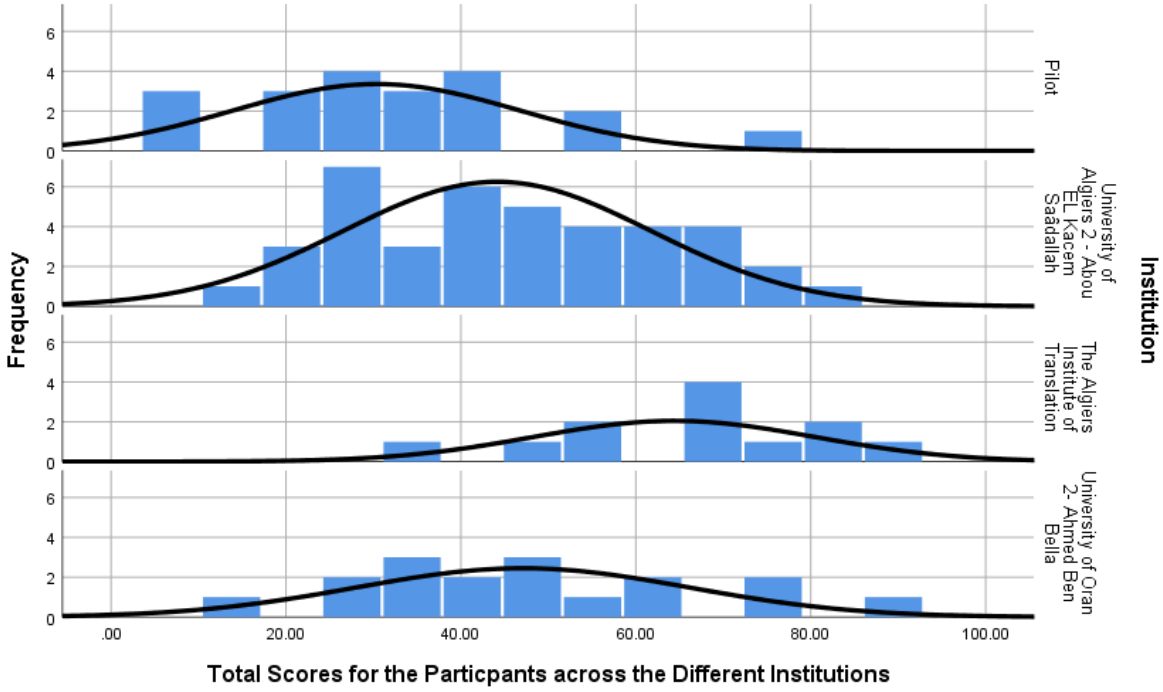

Figure 22 The distribution of the total translation scores across the different institutions

F	Sig.	Partial Eta Squared
9.74	.000	.26

Table 35:Cohen's d

The effect size for the one-way ANOVA test conducted is estimated at .26. Thus 26% of variability in the scores of participants can be explained by the institution they belong to.

5.3.3 Is there a statistically significant difference in the participants' scores based on their amount of exposure to English?

		Frequency	Mean-Score	Levene's test
Time watching English programmes	None Less than 1h From 1h to 2h 3h and more	6 25 30 28	31.61 48.28 44.60 43.23	F (3,85) =1.53/ p=.21>.05
Time chatting with native English speakers	None Less than 1h From 1h to 2h 3h and more	12 32 28 2	43.39 44.29 44.95 65.52	F (3,70) =.096/ p=.96>.05
Time surfing the internet	None Less than 1h From 1h to 2h 3h and more	6 32 37 14	40.23 45.91 42.68 46.80	F (3,85) =.17/ p=.92>.05

Table 36: Categorical responses for self-perceived amount of exposure to English

To answer the question of whether the participants' self-reported amount of exposure to the English language was related to their performance on the collocation test, the number of respondents in each category of questions (10,12, and 14) and their mean scores for each category is shown in Table 36 above. The normality of the scores in each category was checked using Shapiro-Wilk and Kolmogorov-Smirnov tests and the homogeneity of the variances was tested and found tenable using Levene's test (as shown in the table below). Therefore, it is appropriate to use analysis of variance (ANOVA) to test for a significant difference in mean scores between the groups in relation to their self-perceived time of exposure to English. The results indicated that there was no statistically significant difference in scores according to how much time respondents spent watching English programmes, browsing English websites, and chatting online in English.

The first performed one-way ANOVA revealed that there was not a statistically significant difference in the scores of the participants between at least two groups based on time spent watching English programmes (F (3, 85) = 1.20, p =.32>.05) as shown below in table 37.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1396.881	3	465.627	1.191	.318
Within Groups	33217.414	85	390.793		
Total	34614.295	88			

Table 37: Tests of between-subjects effects for one-way ANOVA for watching English TV

A one way-ANOVA was conducted to investigate the effect of chatting online with native speakers of English on their test scores and revealed no statistically significant difference between groups (F (3, 85) = .71, p = .55>.05, as shown in table 38.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	889.360	3	296.453	.709	.550
Within Groups	29251.657	70	417.881		
Total	30141.017	73			

Table 38: Tests of between-subjects effects for one-way ANOVA for chatting online

A one way-ANOVA was performed to examine the effect of the duration spent in surfing internet in English on the participants scores. This revealed that there was no statistically significant difference between the groups with different durations engaging in surfing English websites (F (3,85) = .30p = .82>.05), as shown below in table 39.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	365.607	3	121.869	.302	.824
Within Groups	34248.689	85	402.926		
Total	34614.295	88			

Table 39: Tests of between-subjects effects for one-way ANOVA for surfing in English

5.3.4 Is there a statistically significant difference in the participants scores based on their self-perceived level in Arabic, French and English?

		Frequency	Mean-Score	Levene's test
Level in Arabic	Elementary Intermediate Pre-advanced Advanced Very advanced	4 16 25 33 11	55.17 37.50 41.93 47.02 47.65	F (4, 84) =2.74/ p=.03>.05
Level in French	Elementary Intermediate Pre-advanced Advanced Very advanced	10 19 31 23 6	44.83 49.55 43.72 42.43 37.36	F (4, 84) =.86/ p=.49>.05
Level in English	Elementary Intermediate Pre-advanced Advanced Very advanced	0 13 17 51 8	/ 38.73 36.31 47.94 47.41	F (3,85) =.45/ p=.72>.05

Table 40: Categorical responses for self-perceived level in Arabic, French and English

To examine whether the participants' performance in the translation test was affected by self-perceived level in Arabic, French and English, one-way ANOVA tests were conducted after checking the normality of scores of the participants (see Appendix F) in each group and the homogeneity of variances as shown in table 40. Generally, the tests revealed that there was no statistically significant difference in means according to the different groups representing the respondents' self-perceived level in Arabic, French and English (F (4,84) = 1.10p = .36>.05); (F (4,84) = .56p = .69>.05); (F (3,85) = .20p = .13>.05) respectively, as shown in (tables 41,42,43).

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1720.924	4	430.231	1.099	.363
Within Groups	32893.371	84	391.588		
Total	34614.295	88			

Table 41: Tests of between-subjects effects for one-way ANOVA for self-perceived level in Arabic

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	905.796	4	226.449	.564	.689
Within Groups	33708.499	84	401.292		
Total	34614.295	88			

Table 42: Tests of between-subjects effects for one-way ANOVA for self-perceived level in French

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2241.763	3	747.254	1.962	.126
Within Groups	32372.532	85	380.853		
Total	34614.295	88			

Table 43: Tests of between-subjects effects for one-way ANOVA for self-perceived level in English

5.4 Discussion of the three main research questions of this thesis

This section discusses the results in light of the three main research questions raised in 5.2

5.4.1 Discussion of the results of the first question of the study regarding overall performance of the students and how it was affected by the translation strategies they used

	Good collocations and good translations	Acceptable collocations and acceptable collocations	Acceptable translations but unacceptable collocations	Acceptable collocations but unacceptable Translations	Unacceptable collocations and unacceptable translations	Blanks
	Correct answers			Erroneous answers		
	869	274	4	571	389	
Total	1,147			960		493
Overall number of attempts	2,107					

Table 44: Summary of the classification of the students' correct and incorrect responses based on the acceptability scale

As shown in table 44, results revealed that there was a total of 1,147 correct answers. 869 of these were both good collocations and good translations, while 274 were classified as acceptable collocations and acceptable translations, leaving only 4 as acceptable translations but unacceptable collocations. One of the collocations for which most of the correct answers were of the first group was 'heinous crime'. This received 84 answers, all of which were good collocations and good translations. The students' correct Arabic translations of 'formidable opponent' and مهارة فائقة / *mahāra fā'iqa* were mostly of the second group (acceptable collocations and acceptable translations). For the former, answers like مواجه قوي / *muwājih qawī* or خصم صعب / *xaṣm ṣaʕb*, which when compared with ند لدود / *nidd ladūd* sound less conventional and less accurate, albeit that they are acceptable Arabic equivalents of the English source collocation. Similarly, 'highly skilled' as an equivalent of the latter is considered more conventional and accurate than 'high skills' or 'excellent skill'. Only 4 answers were considered acceptable translations but unacceptable collocations. These only occurred in the rendition of 'fastidious person' and 'rude awakening' as يولي اهتمام لتفاصيل دقيقة / *yūlī ihtimām li adaqq at-tafāsīl* and صدمت عندما تيقنت لحقيقة الأمر / *ṣudimt ʕindamā tayaqqant li-ḥaqīqat al-amr* respectively و which were clauses rephrasing and elaborating the meaning of the source collocations.

Out the total number of attempts, there were 960 errors. While 571 of these were acceptable collocations and unacceptable translations, 389 were neither acceptable collocations nor acceptable translations. Most of the wrong answers produced as equivalents for 'inhabited island' were acceptable collocations but unacceptable

translations, such as جزر مهجورة, جزر نائية جزر خالية / *juzur mahjūra, juzur nā'iya, juzur xāliya* or جزر غير مقطونة, جزر غير مأهولة / *juzur ġayr maqtūna* and *juzur ġayr ma'hūla*, as the adjective 'inhabited' was erroneously rendered into its French false friend 'inhabité'. In the rendition of خداع بصري / *xidāʕ baṣarī*, all errors were unacceptable collocations and unacceptable translations, literal translation being the most commonly adopted technique. This latter played an obstructive role as the collocation in question was a non-transparent collocation resulting in answers like: 'deceptive/ false/deceiving/manipulative/tricky pictures' which are all direct transfers from Arabic L1.

5.4.1.1 Translation strategies adopted by the students

Both correct and incorrect produced answers can be linked to eight distinctive strategies adopted by the participants in the rendition of English collocations into Arabic and vice versa as shown in figure below.

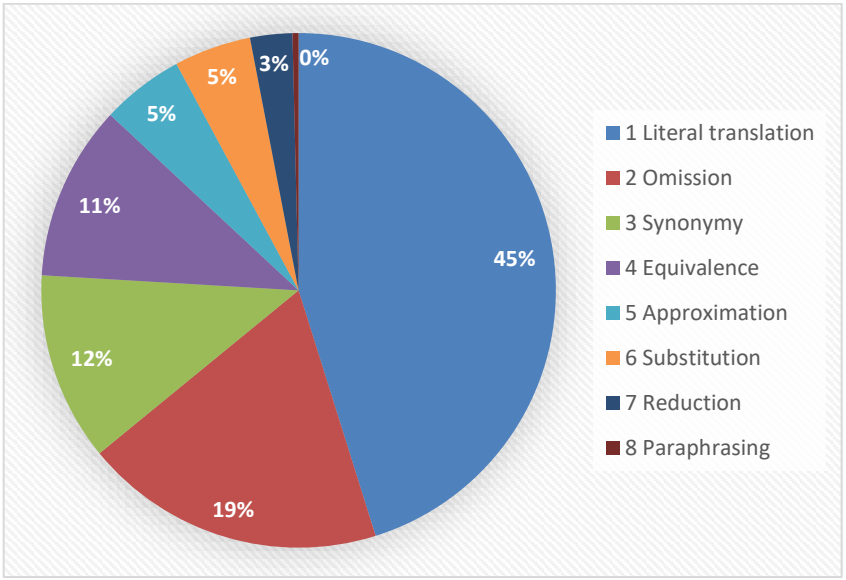

Figure 23: Strategies adopted by the participants in the translation task

5.4.1.1.1 Literal strategy

This is the most conspicuous strategy in the students' responses with 1,174 occurrences (45.15%). Unlike Zughoul and Abdul-Fattah's (2001) results showing literal translation as the third most commonly adopted technique and Dweik and Abu Shakra's results (2010) reporting literal translation as the least used technique with

Quranic collocations, this study found that literal translation is the most frequent technique.

Although this served as a facilitating technique in rendering a few transparent collocations, it equally led to inaccuracies and mistakes due to language interference instances particularly second language interference from French. One of the best examples showing the effectiveness of literal translation was translating 'heinous crime' into its Arabic equivalents *جريمة شنعاء /مروعة /فظيعة /شائنة/بشعة /شنيعة jarīma šanʕā'/ šanīʕa bašiʕa / šā'ina/ nakrā'/ faḍīʕa / murawwiʕa*, which resulted in 84 correct answers**.** The successful rendition of this collocation may also emanate from students' non-exposure to the French word *haineux* 'full of hatred', just as much as from their awareness of the English adjective *heinous*. In contrast, an example of the ineffectiveness of literal translation was noticed with the rendition of 'sensible person', where out of 70 wrong answers, 67 were possibly due to French interference. This is because *شخص حساس šaxṣ ḥassās*, meaning 'sensitive person', which is a result of direct negative lexical interference from French, was suggested as an equivalent by these respondents for 'sensible person'. This suggests that literal translation more often plays an obstructive role with transparent collocations (although it may be facilitative in a few cases) rather than a merely facilitative role, as suggested by (Bahumaid, 2006). This accords with Dukali's results (2018), which demonstrate that L1 interference is one of the main reasons why students produce unnatural collocations. Similarly, it corresponds to Almakhtary (2017), who attributed the incompetence of EFL learners mainly to overgeneralization, and L1 (Arabic) and L2 (French) interference, and the study of Jabak, Abdullah and Mustapha (2016), which demonstrated that participants struggled in rendering collocations from Arabic to English mainly due to literal translation. As in other studies, the negative role of L1 interference was also noted in Ahmed (2012) results.

5.4.1.1.2 Omission

participants refrained from translating 492 times (18.92%), making omission the second most frequently adopted technique. This can be ascribed to either lack of knowledge, or unwillingness to undertake a voluntary task of half an hour. This corresponds to Abdul-Fattah & Zughoul (2001), who found that omission, which they referred to as 'avoidance' was the second most adopted technique with 28.24 % after paraphrasing. It also partially accords with Dweik and Abu Shakra's (2010) deletion

strategy, involving eliminating either a part or whole answer, which they found to be the commonest technique used in rendering Biblical and Hadith collocations and the third most commonly adopted technique for rendering Quranic collocations. In this test, omission was mostly adopted with the two collocations 'rude awakening' and *عقبة كؤود / ʕaqaba ka'ūd*, which only received 8 and 34 correct answers respectively and accordingly had low index of easiness scores as well. Depending heavily on omission, then, may be a sign of an item's difficulty.

5.4.1.1.3 Synonymy

The use of synonymous words of either one or both constituents of the collocation of the expression is an attempt to provide an equivalent target collocation. What might cause erroneous or less accurate translations when adopting this technique is that full synonymity is not always attainable especially if the collocation in question is clearly conventionalised. Synonymy was the third most frequently used technique in the translation task at 11.84%, representing 308 instances. Learners' reliance on this technique may be due to: 1. the availability of many synonymous alternatives that were close enough in particular contexts allowing choices to be made between them without affecting the idiomaticity and acceptability of the collocation in the target language, as is the case with translating *jolly woman* being translated as *بشوشة / مبتهجة / مسرورة / سعيدة / ضحوكة امرأة / مرحة imra'a' bašūša / mubtahija/ masrūra // sayyida / saʕīda / ḍaḥūka;* 2. the inability of respondents to choose from among different options the precise collocate out of unawareness of the selectional restrictions that collocations are subject to, as in translating *candid camera* as *كاميرا مغافلة / مخبئة/ مخفية kamirā muġāfila / muxabba'a / maxfiyya*, i.e. as synonymous with *كاميرا خفية kamirā xafiyya*. Synonymy here includes also used exaggerated synonymous[9] forms of the constituent parts of the collocation. Exaggeration instances totalled 19 cases, 18 of which were made when rendering مهارة فائقة *mahāra fā'iqa* into English. The participants in this case gave many exaggerated synonymous equivalents for the adjective *fā'iqa* 'high' in Arabic. This technique resulted in most cases in adjectives

---

9 Exaggerated synonyms/ hyperbolic near-synonymous TT translation entails using a hypernym instead of using hyponym as in translating 'small' as 'minuscule' or 'good' as 'excellent', where the latter is semantically properly included within the former.

like ‘outstanding’, ‘terrific’, ‘extraordinary’, or the prefix ‘super’. This concords with the heavy use of synonymy that was also noted in Dweik and Abu Shakra (2010), where this was the commonest technique adopted by the subjects in rendering religious lexical collocations, but contrasts with Abdul-Fattah & Zughoul (2001), which showed a minimal use of this technique at just 2.68% of overall occurrences. A dependence on exaggeration particularly may emanate from the impression that when choosing big-sounding adjectives, the translations become impressive.

5.4.1.1.4 Equivalence

With 286 instances in this analysis, equivalence makes up 11% of total used strategies and accounts for all the correct answers generally and particularly those of all the participants from the High Arab Institute, who successfully rendered *candid camera* as *الكاميرا الخفية / al-kamirā al-xafiyya*. In translating *خداع بصري / xidāʕ baṣarī,* out of 22 correct answers, 20 were the result of adopting the equivalence strategy. Equivalence always results in producing correct answers (acceptable collocations and acceptable translations). Accounting for 286 out of 1,147 correct answers, equivalence constitutes almost a quarter (24.93%) of the correct answers produced in this study. Although equivalence necessarily leads to correct answers, students did not mainly adopt it in the rendition of any of the collocations, as they either failed to identify equivalent collocations in the target language or, because there is no such equivalence between Arabic and English (non-equivalence) in relation to a particular collocation as exemplified by the collocation ‘concurrent enrolment’, for which there is no exact equivalent in Arabic.

The lack of equivalence may be due to cultural differences between Arabic and English which this expression involves. As concurrent enrolment is not an allowed option in the Algerian higher educational system and in Arabic universities in general, neither the term nor the notion it expresses are familiar to students. This leads us to think of this term as involving cultural untranslatability. Therefore, although *الانخراط المشترك / التسجيل المتزامن /الدراسة المتزامنة ad-dirāsa al-mutazāmina / at-tasjīl al-mutazāmin / al’inxirāṭ al-muštarak* (all referring to simultaneous registration) were considered linguistically correct translations for ‘concurrent enrolment’, these translations do not mean the same to Arabic speakers as their English counterpart means to English native speakers as the notion is directly related to their educational reality.

5.4.1.1.5 Approximation

Approximation was the fifth most frequent strategy with 5.19% (135 instances of total attempts). This was one rank behind the corresponding technique, generalisation, in Abdul-Fattah & Zughoul (2001) but three behind generalisation in Dweik and Abu Shakra (2010), where generalisation was the second most frequent strategy. Approximation was also the main adopted technique for translating عجز مالي / *ʕajz mālī* into English, students proposing 'fiscal deficit' and financial crisis' as equivalent collocations. These were considered good collocations but unacceptable translations since they deviate from the meaning of the source collocation in that both are more specialized terms. While both *fiscal deficits* and *financial crises* hit governments, banks and companies, a *financial deficit* is a nonspecific term that can apply to individuals as well. In providing these two answers, students seem to adopt the approximation strategy where other terms with more specific meanings were recalled. 'Consistent manner' was the second collocation where students depended mainly on approximation. In their rendition of this, there were 50 errors of different types and severity. The closest incorrect translations to the correct ones were أسلوب ثابت / *uslūb ṯābit,* صفة ثابتة / *sifah ṯābita,* شكل متوافق / *šakl mutawāfiq,* صفة مستمرة / *sifah mustamirra,* and شكل دائم / *šakl dā'im.* These translations were classified as good collocations but unacceptable translations as they only captured the semantic aspect of 'doing something in the same way over time'.

Approximation leads in most, if not all, cases to wrong answers. Students may use this technique when they cannot recall or produce the exact equivalent target collocation, so they try their best to approximate the whole meaning communicated by the source collocation by reconstructing an optimal meaning, as suggested by Dweik and Abu Shakra (2010). This approximation may, at the minimum, be manifested in the given translations through referring to the defining features of the source language item. This was the case in rendering 'candid camera' into Arabic where some answers referred to the fact that a candid camera is a hidden camera producing 'hidden camera' مصورة مخفية / *musawwira maxfiyya* as an equivalent form, its small size resulting in answers like كاميرا صغيرة/ *kamirā saġīra,* or even its function of recording as in كاميرا مراقبة / *kamirā murāqaba.*

5.4.1.1.6 Substitution

Substitution resulted in good collocations but unacceptable translations in this translation task. This strategy was the fourth most frequent technique with 126 instances, accounting for 4.84% of the total answers produced. This was mostly used by the subjects when rendering *ulterior motive* as *نوايا مضمرة / nawāya muḍmara*, *دافع قوي / dāfiʕ qawi*, *نية سيئة / niyya sayyi'a*, and *نوايا خفية / nawāya xafiyya*, which back-translate as 'hidden intentions', 'strong motive', 'bad intention', and 'implicit/planned intention'. Although these substitutes are idiomatic Arabic collocations, they fail to convey the intended meaning of the source collocation.

Other instances of substitution include completely irrelevant answers. An example of this is the rendering of 'sensible person' as *شخص حنون / šaxṣ ḥanūn*, *إنسان متحفظ/ insān mutaḥaffiḍ*, and *شخص خجول / šaxṣ xajūl*, which may be due to random guessing on the students' part. Using substitution, particularly giving completely irrelevant answers, may be due to reckless decisions or random guessing.

5.4.1.1.7 Reduction

Reduction was the second least frequent strategy adopted by students with 69 instances accounting for 2.65% of total attempts. This accords with Dweik and Abu Shakra (2010), where reduction was also the second least used technique in translating Quranic lexical collocations. Resorting to this technique may be an indicator of students' inability to find the exact target collocation, their carelessness in omitting an integral part of the collocation, or equally their focus on the general meaning of the sentence rather than conveying the specific collocation.

5.4.1.1.8 Paraphrasing

Unlike in Dweik and Abu Shakra (2010), where paraphrasing was the third most commonly employed strategy in translating lexical collocations and the second most common in rendering semantic collocations, in our study, paraphrasing was the least commonly used strategy by the students at 0.38%, accounting for ten instances. Translating the collocation 'secular humanism' as *حركة انسانية غير مقيدة / ḥaraka 'insāniyya ġayr muqayyada* suggests the use of linguistic simplification via paraphrasing to explain what the collocation means. Similarly, rendering *مزاج متعكر / mizāj mutaʕakkir* as 'was not in the mood' in English involves lexical simplification via paraphrasing, as is the case with *الاستماع اليه بشفقة / al-'istimāʕ 'ilayh bišafaqa* when rendering "sympathetic ear" into Arabic.

Apparently paraphrasing is opted for in cases where participants cannot produce accurate equivalent collocations and therefore provide either an explanation, definition or reformulation for the source collocation. The minimal use of this strategy can be associated with its non-conformity with the aim of the task, which requires target collocations, as it "involves filling a one-slot item with an explanation consisting of several items" (Baker 2018, p. 43).

5.4.1.2 Types of errors

While translation strategies can be linked to lexical errors, there is no clear relationship between these strategies and lexical, grammatical and spelling errors, which I will consider in the following sections.

5.4.1.2.1 Lexical errors

This type of error was the most frequent with 92.10% in the students' produced translations of *adjective+noun* collocations including wrong choice of both adjective and noun. Most, if not all, of the lexical errors that occurred in this study were the result of the wrong choice of the adjective, which can be often due to adopting a literal translation strategy or other strategies, mainly substitution and approximation, and to a lesser extent synonymy and reduction. Most, if not all, instances categorized under acceptable collocations, but unacceptable translations are directly linked to the use of substitution and approximation strategies.

5.4.1.2.2 Grammatical errors

This type of error was less common than lexical errors in the students' answers, with only 52 errors (5.41%). These include either a change in the order of the *adjective+noun* or mostly translations with wrong patterns. There were three common wrong-pattern translations in the respondents' answers consisting of *adverb+noun* instead of *adjective+noun*, *adjective+adjective* instead of *adverb+adjective*, and wrongly formed adjectives or nouns via adding suffixes ('ing', 'ed'). An example of the first type is 'highly professionalism' or 'high professional' for 'highly professional'. Examples of the second erroneous pattern are 'financial vulnerable' and 'financial weak' for 'financially vulnerable' and 'financially weak' when trying to render عجز مالي / *ʕajz mālī* into English. Instances of erroneous patterns due to the adding of 'ing' and 'ed' suffixes are noted in the use of 'refuting' for 'irrefutable' in 'irrefutable argument', 'bleeding wounding' and 'bleeding injuring', or 'blooded

injury' for 'bleeding wound'. Grammatical errors including a change in the order of the *adjective+noun* pattern were also noticed in misplacing the noun الانسانية / *al-'insāniyya* before the adjective العلمانية / *al-'ilmāniyya*, this being the most frequent error in translating 'secular humanism' into Arabic. Apart from this error, all instances of grammatical errors were found with the English translations of Arabic collocations. This may well be explained by the better proficiency of the learners in Arabic, which is their first taught language.

5.4.1.2.3 Spelling errors

This type of error was the least observed at only 3.54% (34 instances) of the total errors recorded and they all occurred in the English translations of the Arabic collocations in the test. Among the words that were often misspelt in the respondents' answers were 'proficiency' and 'professionalism', which were spelt as follows: 'proeftiоncy'*, 'professioncy'* 'Preffwsionalism'*, and 'professionalisme'*.

In recording more lexical errors, 91.04%, than grammatical errors, this study demonstrates that participants have more difficulties in choosing the correct node or collocate in their produced translations than forming a grammatically correct collocation. French L2 interference is the main reason behind the dominant presence of lexical errors. These facts can be also explained by the following two other points: 1/ the level of the subjects and 2/ the nature of the task.

First, the students in this study (2nd year master's) are academically more advanced than those of the studies reviewed above (Abdul-Fattah and Zughoul, 2001; Ahmed 2012; Almaktari 2017; Alsakran 2011; Boumali 2010; Hammadi, 2015). This can be related to their better English proficiency including collocational proficiency and grammatical accuracy. This is evidenced by Almaktari's (2017) results, which demonstrate that oral and writing skills of students correlate positively with their general collocational proficiency, and Ahmed's (2012) study, in which he claims that the year of study influenced the outcome of the test; third-year students performed better than second-year students.

Second, the fact that this study adopted a translation test for a pre-determined grammatical pattern *adjective+noun*, rather than a free written production task, facilitated the task for students and therefore minimized the potential occurrence of grammatical errors. This can be further exemplified by Dukhali's study (2016), which

used a free written production task that ultimately led to the dominant occurrence of grammatical errors.

5.4.2 Discussion of the second question related to the performance of the Students in translating English collocations into Arabic and Arabic collocations into English

The second research question addressed in this study asked whether there was a significant difference between Algerian students' performance in rendering English collocations into Arabic and vice versa. In response to this question this study found that there were more errors in the rendition of English collocations into Arabic than in the rendition of Arabic collocations to English. This is demonstrated by the statistically significant difference in the mean score between the English collocations task (42.40%) and the Arabic collocations task (47.98%), as evidenced by the dependent t-test result (t= -2.40 (88) at p = .02<.05).

The better performance of the students in Arabic-English translation of collocations can be explained by the very few instances of French interference represented through spelling mistakes in this part of the test in comparison with the English-Arabic translations where almost all the cases of lexical French interference were recorded.

It is worth noting that while there are a substantial number of studies that dealt with translating English to Arabic (Almaktari 2017; Alsakran 2011; Dukhali, 2016), relatively few have dealt with Arabic-English translation (Ahmed, 2012; Abdul-Fattah & Zughoul, 2003; Dweik & Abu Shakra, 2010; Jabak, Abdullah and Mustapha,2016; Hammadi, 2015), and even fewer have compared students' performance in translating English-Arabic and Arabic-English collocations (one example is Shammas, 2013).

There is a consensus among all the reviewed studies related to English-Arabic translation of collocations on the poor collocational performance of the students. Dukali (2016) found that 61.9% of the produced answers were acceptable English collocations while 38% were not (errors). Similarly, but slightly worse, in Almaktari (2017), the mean score for acceptable collocations was 29.30% while in Alsakran (2011) the respondents' performance lagged far behind with 16.07% as a mean score.

With a mean score of 42.40% the findings of this study corroborate previous studies in terms of the poor collocational performance in rendering English collocations into

Arabic. The higher mean score in this study as compared to previous studies can be explained by the advanced level of the participants.

In most of the reviewed studies related to Arabic-English translation of collocation in this thesis, the students' Arabic-English translation scores were unsatisfactory with a percentage of errors above 60%. In Jabak, Abdullah and Mustapha's study (2016), errors occurred in 71.86% of the total given answers, and in Dweik Abu Shakra's (2010) in 62.5%, while Zughoul and Abdullah (2003) found that only16.6 % of students total attempts with the translation task were correct. Ahmed's study (2012) was the only one in which the students' scores were close to the mean score recorded in the study at hand, with a mean score of 42%.

With a mean score of 47.98%, the overall performance of the Algerian students in translating the target collocations from Arabic to English was better than the mean scores in other studies. However, it still lower than might be expected from Masters' students.

While Hammadi's questionnaire (2015) included both English and Arabic collocations for translation, its findings did not compare the English-Arabic translations with the Arabic-English ones and just focused on the collocational errors and their sources. By contrast, in Shammas' study (2013), and in congruence with this study, the performance of the participants of the Higher Arab Institute, from which data was collected for this study as well, showed that the errors in English-Arabic translations (87.92%) outnumbered those in Arabic-English translations (84.80%). However, the overall number of errors across all the different institutions in the same study showed the opposite; errors in Arabic-English collocations (76.98%) outnumbered those in English-Arabic collocations (63.44%).

5.4.3 Discussion of the third question in relation to how Algerian EFL learners' L2 affects their proficiency in translating English collocations into Arabic and Arabic collocations into English

Most of the lexical errors that occurred in this study were the result of the wrong choice of adjective or noun, which, in turn, was in many cases can be because of L2 interference, or L1 interference via the literal translation strategy, and in others because of the adoption of other strategies mainly substitution and approximation, and to a lesser extent synonymy and reduction.

When translating جريمة شنيعة / *jarīma šanīʕa* into English, only 3 errors occurred, all of which were lexical. Of these three errors, two were because of L2 interference, the given answers being جريمة كراهية / *jarīmat karāḥiyya*, and جريمة كره / *jarīmat kurḥ*. Students in this case misunderstood 'heinous crime' for its corresponding false friend in French 'crime haineux', which means in French a crime driven specifically by hate rather than a brutal or evil crime regardless of the motives behind it, as in English. Also, in rendering 'sensible person' into Arabic, there were 70 wrong answers out of which 67 were possibly due to French interference. The most frequent answer was شخص حساس / *šaxṣ ḥassās*, which was a result of direct negative lexical interference from French. Similarly, out of 38 errors made in translating 'actual cost' into Arabic, 25 were due to French interference, as the adjective 'actual' was mistakenly understood as meaning the same as French 'actuel', leading to answers like: الثمن الحالي / *aṯ-ṯaman al-ḥālī* and التكلفة الحالية / *at-taklufa al-ḥāliyya*.

In all the above-mentioned examples French interference resulted in acceptable collocations but unacceptable translations as they deviate semantically from the meaning of the source collocations.

French interference also resulted in expressions that are neither acceptable collocations nor acceptable translations as in the rendition of 'concurrent enrolment'. Out of 26 erroneous answers, the ten noted French instances caused both inacceptable collocations and unacceptable translations in Arabic. In these instances, the adjective 'concurrent' was erroneously rendered as التسجيل التنافسي / *at-tasjīlāt al-muʕādiya* and التسجيلات المعادية / *at-tasjīl at-tanāfusī*. Concerning spelling, some spelling mistakes/ errors like 'professionalisme'*, 'pertinante'* and 'politic'* could be ascribed to French interference as they are French spellings for 'professionalism', 'pertinent', and 'political' respectively. Apart from 1/ the level of the subjects and 2/ the nature of the task, the dominant presence of lexical errors can also be linked to L2 interference, which mostly led to wrong choices of collocates/nodes rather than grammatical errors.

In this research, while spelling errors can be mostly linked to L2 interference, grammatical errors can be due to either overgeneralization by generalizing grammatical rules like the use of the 'ing' suffix, for example, to form nouns, or L1 interference when an Arabic structure is transferred into English, resulting in misplacing the adjective after the noun in the rendition of English collocations into Arabic. Lexical errors, the most frequent type of error, can be related to L1

interference, L2 interference, and the reliance on certain translation strategies like substitution, approximation, synonymy and reduction The results of this study correspond to the reviewed studies' findings about the impact of L2 on learning L3. These include Sadouni (2016), which demonstrates that the main reason behind Algerian students translation spelling mistakes is that they "mistake English words for French ones" (p. 133), Almaktary (2017), which attributes the incompetence of EFL learners mainly to overgeneralization, and L1 (Arabic) and L2 (French) interference, and Boumali (2010), which shows that most (64%) of the respondents struggled when trying to find the right equivalent between English and French when the entry has an identical spelling in both languages (p. 28).

The findings of this study also accord with the findings of the other reviewed studies about collocations such as Nesselahauf (2003), Ahmed (2018) Dukhali (2016) and Jabak, Abdullah & Mustapha (2016), which all demonstrated that L1 interference had a negative effect on the collocational competence of EFL learners leading to unnatural collocations, these being based on collocations found in the native language of the learners. In addition to L1 interference, Hammadi (2015) and Dukhali (2016) attribute the occurrence of the different errors to overgeneralization and use of synonymy.

5.5 Discussion of the four sub-questions of this thesis

This section discusses the results in light of the four sub-questions raised in 5.3.

5.5.1 Discussion of the first sub-question related to how gender affected the translation test scores

In relation to the difference in the performance between the participants in this study based on their gender, the independent T-Score test demonstrated that gender has no relationship with the performance of the participants. The overall sample had 20 male participants and 69 female participants. Therefore, the overall group chosen for the study had more than 2/3 female students and less than 1/3 male participants. Gender disproportionality in joining English or translation departments, as shown in this sample, is due to societal and motivational factors. First, societal gender-stereotyping jobs favour language learning degrees for females and thereby females think that doing languages-related degrees can enhance their occupational opportunities. Second, females show more interest and confidence in learning languages than their male peers, which in turn can be due to their different motivational factors. While

females' motivations are more integrative in learning languages, as manifested in liking their related cultures and speakers, males' motivations tend to be more instrumental.

The literature points out that females are more efficient in verbal skills while males do better when spatial abilities are involved. This might be due to neurological differences between females and males manifested by the approaches they take towards learning with conscious analytical approaches for males and nativist approaches for females. Swaminathan (2008) suggests that there are measurable differences in terms of brain activity when learning words. This is also supported by (Rua, 2006), who claims that: "In a classroom setting... boys need to be taught language both visually (with a textbook) and orally (through a lecture) to get a full grasp of the subject, whereas a girl may be able to pick up the concepts by either method" (p. 6). However, the result of our study is not in congruence with the biological difference hypothesis.

This insignificant difference in mean scores based on gender in this study can be probably explained by the small number of males in our sample and the similar educational and linguistic background of the learners in terms of sharing the first language and having the same curricula and probably also the use of various teaching strategies that cater for the needs of both genders (using both visual and strategies). This corresponds to Ahmed's (2012) results where the mixed ANOVA showed that there was no significant statistical difference between the performance of male and females in terms of their lexical collocation knowledge.

5.5.2 Discussion of the second sub-question in relation with the difference in the participants scores based on the faculties in which they are enrolled

The participants in this study were enrolled in three different institutions, Abu AL Kassem Saadallah Algiers 2 University, the High Arab Institute of Translation, and Ahmed ben Bella Oran 2 University. To determine whether and how the groups differed from one another, we ran a post-hoc test. The results of the Tukey post-hoc test showed multiple comparisons between the different groups.

There was a statistically significant difference between the groups as determined by one-way ANOVA ($F(3) = 9.74$ $p = .000$). A Tukey post-hoc test revealed that the mean score difference for the students at the High Arab Institute was higher than the mean scores obtained from both the universities of Algiers and Oran, with the means

of difference (20.33 ±17.57 marks at p=.004); (17.31 ±19.04 marks at p = .05) respectively. There was not however, a statistically significant difference between the mean scores of Algiers' students and those of Oran university (2.92p = .94).

One possible explanation for the statistically significant difference between the High Arab Institute and Abu AL Kassem Saadallah Algiers 2 University, and Ahmed ben Bella Oran 2 University is that students from the High Arab Institute are more exposed to the English language than their colleagues at other institutions, as indicated by their answers related to exposure to the English language in table 29. Over a third of the High Arab Institute of Translation students attended English training programmes in Algeria and a similar percentage had a native English tutor. In addition to 50% of these students spent 3 hours and more in watching English programmes daily. By contrast, only 15% of the students from Abu AL Kassem Saadallah Algiers 2 University and just under 6% from Ahmed ben Bella Oran 2 University attended training programmes in learning English, these being of varying lengths from 2 weeks to 2 years. The same percentages were recorded from both universities for those who were taught by native English speakers. Also, roughly similar percentages were observed when it comes to students who travelled to native English-speaking countries, staying from 8 days to 2 months (7.5% from Abu Al Kassem Saadallah Algiers 2 University, and 5.88% from Ahmed ben Bella Oran 2 University), time spent watching English programmes (32% of Abu Al Kassem Saadallah Algiers 2 University students, and 29% of Ahmed ben Bella Oran 2 University students spent 3 hours or more daily watching English programmes). Their type of bachelor's major was also slightly different, with approximately 86% of the students from Abu Al Kassem Saadallah Algiers 2 University and 94% from Ahmed ben Bella Oran 2 University majoring in English language. This slight difference may explain the slightly better performance from Ahmed ben Bella Oran 2 University students on the ground that those who majored in English had more exposure to English than those who majored in translation.

Institutions	Trainings	Native-Eng. Teacher	Travel to NESC	Watching Eng. Programmes	English Bachelor's degree
High Arab Institute for Translation	33%	33%	25%	50%	41.66%
Abu Al Kassem Saadallah Algiers 2 University	15%	15%	7.5%	32.5%	87.5%
Ahmed ben Bella Oran 2 University	5.88%	5.88%	5.88%	29.41%	94.11 %

NESC: Native English-Speaking Countries

Table 45: Summary percentages of students' answers to Questions 4,6,8,10 and 15 across the different institutions

5.5.3 Discussion of the third sub-question in relation to how the performance of the students was affected by their exposure to English activities

This question asked if there was a statistically significant difference between the participants proficiency in producing collocations and their reported amount of exposure to the English language through shows, friends and browsing. To this end, analysis of variances (ANOVA) tests were used to see if there was a significant difference in mean scores between the groups in relation to their self-perceived time of exposure to English. The results indicated that there was no statistically significant difference in scores according to how much time respondents spent watching English programmes, browsing English websites, and chatting online in English.

The first one-way ANOVA test revealed that there was not a statistically significant difference in the scores of the participants between at least two groups based on time spent watching English programmes ($F(3, 85) = 1.20$, $p = .32 > .05$), as shown above in table 37, although the first part of the graph (24) below shows that the scores of the participants increased drastically from the first group with no exposure to English programmes to the second group with up to one hour watching English programmes before slowly decreasing as the time of exposure increased.

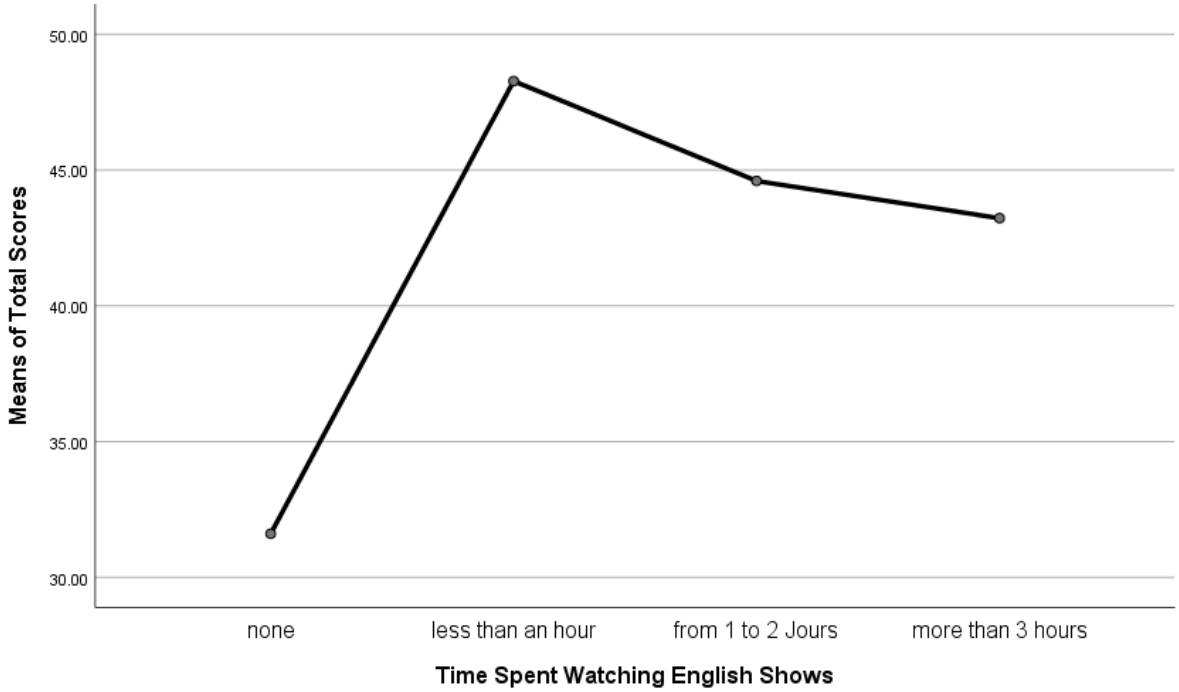

Figure 24: Mean percentages of the participants based on the time spent watching English shows

The one way-ANOVA conducted to investigate the effect of chatting online with native speakers of English on their test scores revealed that was no statistically significant difference between groups ($F(3, 85) = .71$, $p = .55 > .05$), as shown in table 38. The graph demonstrates that the scores of the participants steadily increased from the first group (none) to the third group with up to 2 hours engaging in chatting online before rising steeply for those who spent longer than 2 hours.

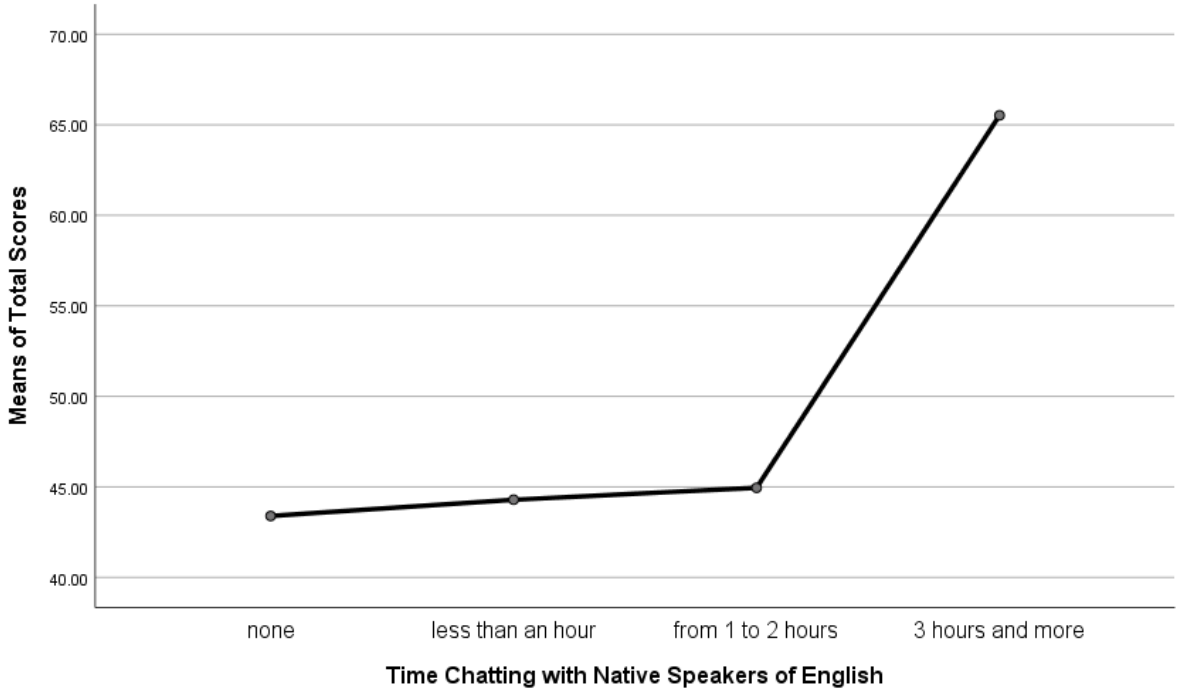

Figure 25: Mean percentages of the participants based on the time spent chatting online with native English speakers

A one way-ANOVA was performed to examine the effect of the duration spent on surfing the internet in English on the participants scores. This revealed that there was no statistically significant difference between the groups regarding the time spent in surfing English websites (F (3,85) = .30p = .82>.05), as shown above in table 39.

The mean scores of the participants fluctuated as shown in the graph below. These increased from the first group (none) to the second group (up to 1 hour), then dropped sharply and swiftly rose again starting from 2 hours onwards.

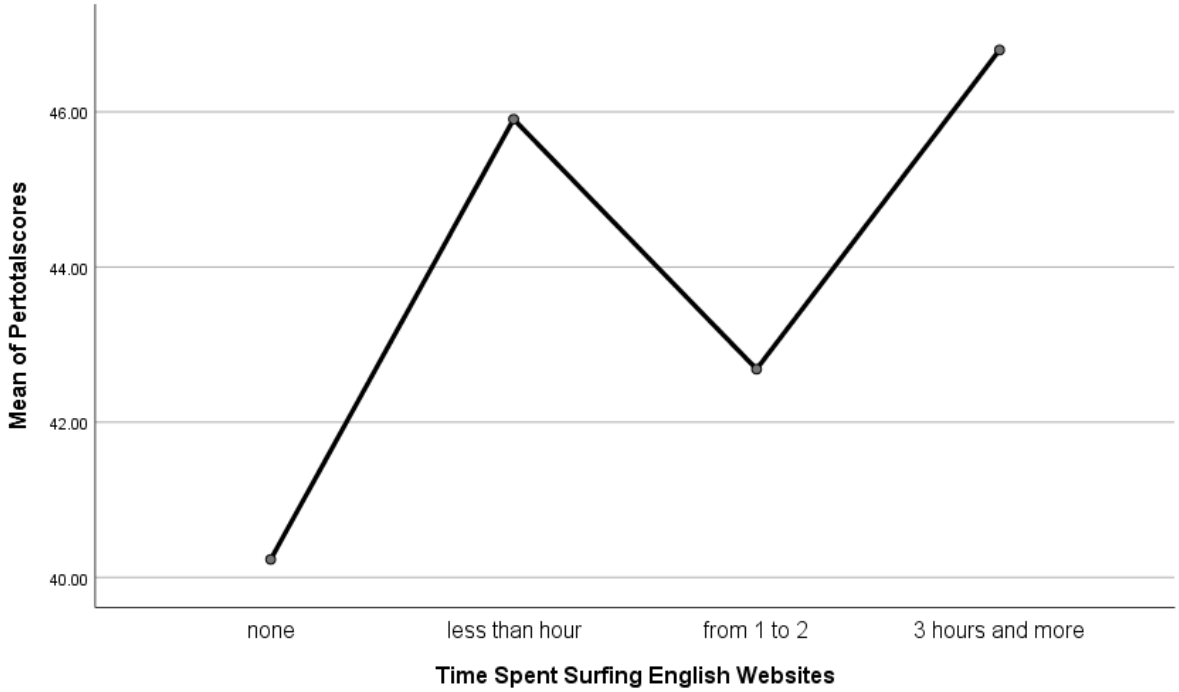

Figure 26: Mean percentages of the participants based on the time spent surfing the internet in English

It is worth noting, however, that this non-significant statistical result does not eliminate the possibility of the influence of the exposure time on students' scores. Rather it merely shows that the difference in students' scores belonging to the different groups representing different exposure times is not sufficient to rule out the possibility that it might be due to chance. To reiterate, it cannot be confirmed that the difference between the groups is due to chance, only that we do not have evidence that it is not due to chance. The difference in scores between the groups then might still be caused by their different exposure time to English, or it might be just caused by chance.

This does not agree with Ahmed (2012) and Shehata (2009), who showed that in their studies that there was a statistical significance between the average scores for the groups representing different lengths of exposure to English. Ahmed's (2012), and Shehata's (2008) findings demonstrated that the more time students spent on activities involving exposure to English, the higher their scores became. Therefore, both works concluded that the time of exposure to English though real-life experiences impacts positively on the students' scores. Ahmed (2012) goes further, stating that the longer

Libyan EFL students are exposed to English, the better their collocational knowledge of English becomes.

5.5.4 Discussion of the fourth sub-question regarding differences in the participants' scores based on their self-perception of their levels in English, Arabic, and French

The purpose of this question is to consider whether there is a statistically significant difference between the participants in the translation test according to their self-perceived level in Arabic, French, English. Generally, the tests revealed that there was no statistically significant difference in means according to the different groups representing the respondents' self-perceived level in Arabic, French, and English (F (4,84) = 1.10p = .36>.05); (F (4,84) = .56p = .69>.05); (F (3,85) = .20p = .13>.05) respectively as shown in (tables 41,42,43).

The graphs for both Arabic and English demonstrate that the mean score increased hand in hand with the self-perceived level for both languages. This increase started from the intermediate level group onwards for Arabic and from the pre-advanced level group for English, as shown in the following graphs.

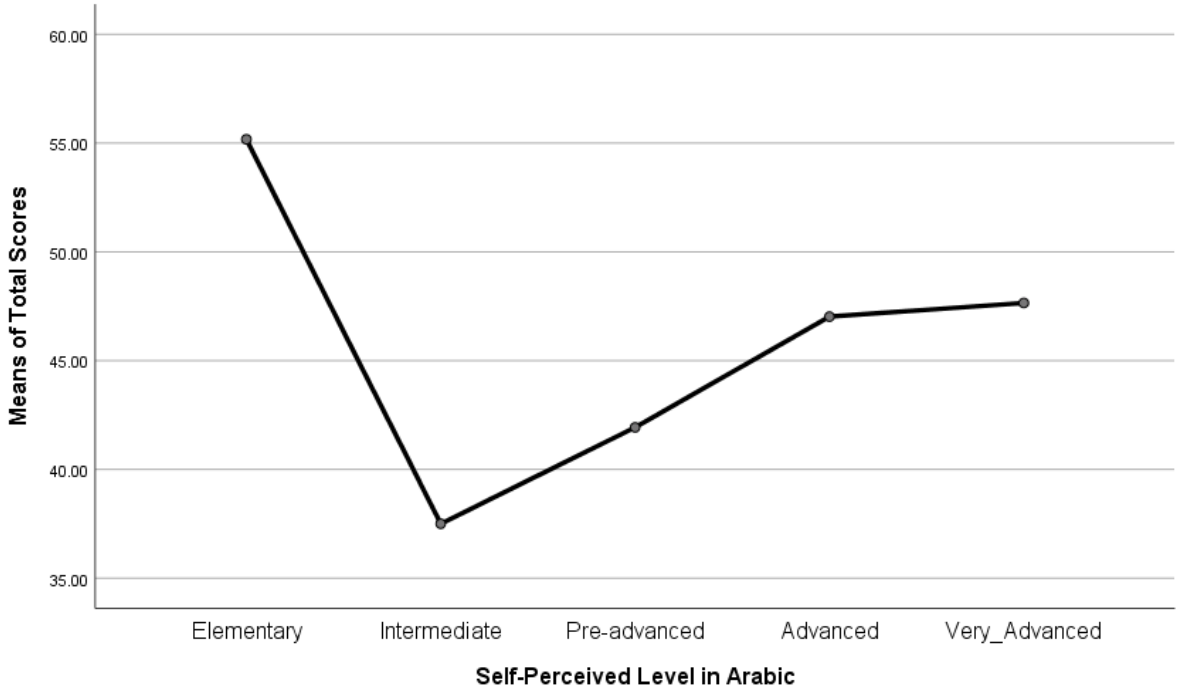

Figure 27: Mean percentages of the participants based on their self-perceived level in Arabic

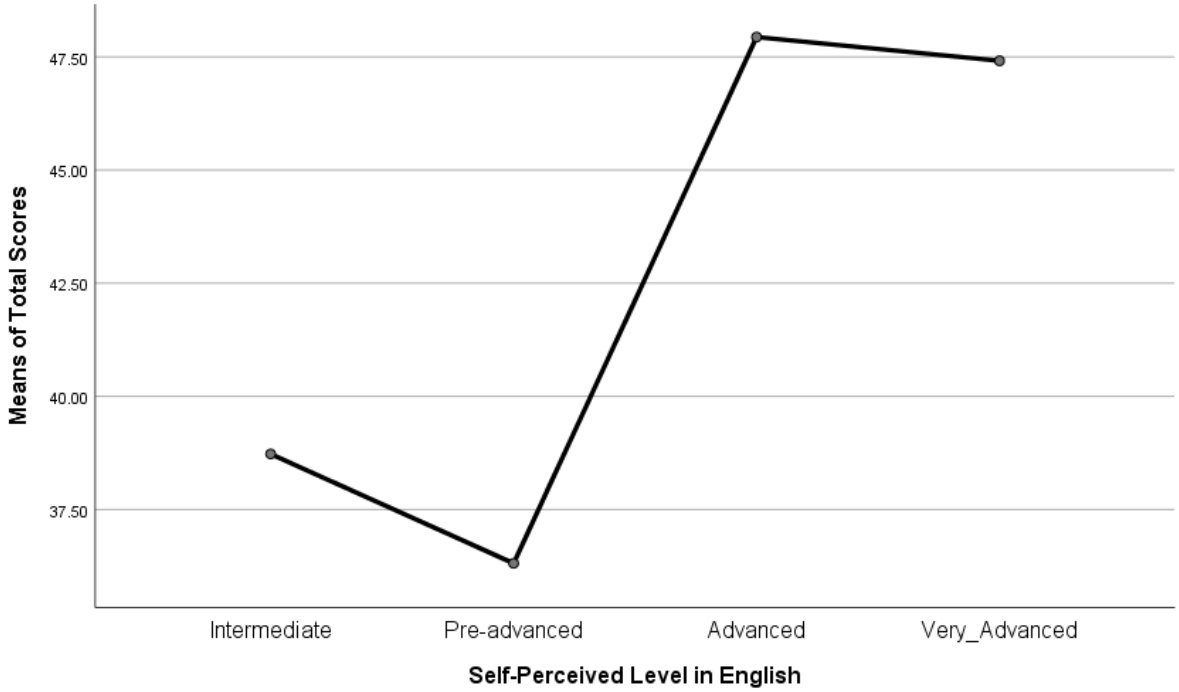

Figure 28: Mean percentages of the participants based on their self-perceived level in English

However, the graph for the French language, below, showed that respondents' scores increased starting from the intermediate level group, and

decrease when self-perceived levels are higher. Literature in educational psychology points towards the strong correlation between self-perception and academic achievement (Ghazvini, 2011). Studies like Mathew (2017) and Guay & al. (2010) found that that high academic self-perception in students massively contributes to a better performance.

Although the results showed in graphs 27 and 28 might support the plausibility of the claim that students with a higher self-perception of their levels in Arabic and English scored higher than those with a lower self-perception levels in the two languages, there is no statistical significance to the results. One explanation for the statistically insignificant difference between participants according to their self-perceived levels in English and Arabic, is that this difference might be due to chance.

Similarly, graph 29 representing students' self-perception of their level in French supports the hypothesis that students with a higher self-perceived level in French tend to commit more errors related to L2 interference, thereby achieving lower scores. However, the statistically insignificant difference in mean scores according to the

different groups of self-perceptions suggests also that these differences might be due to chance.

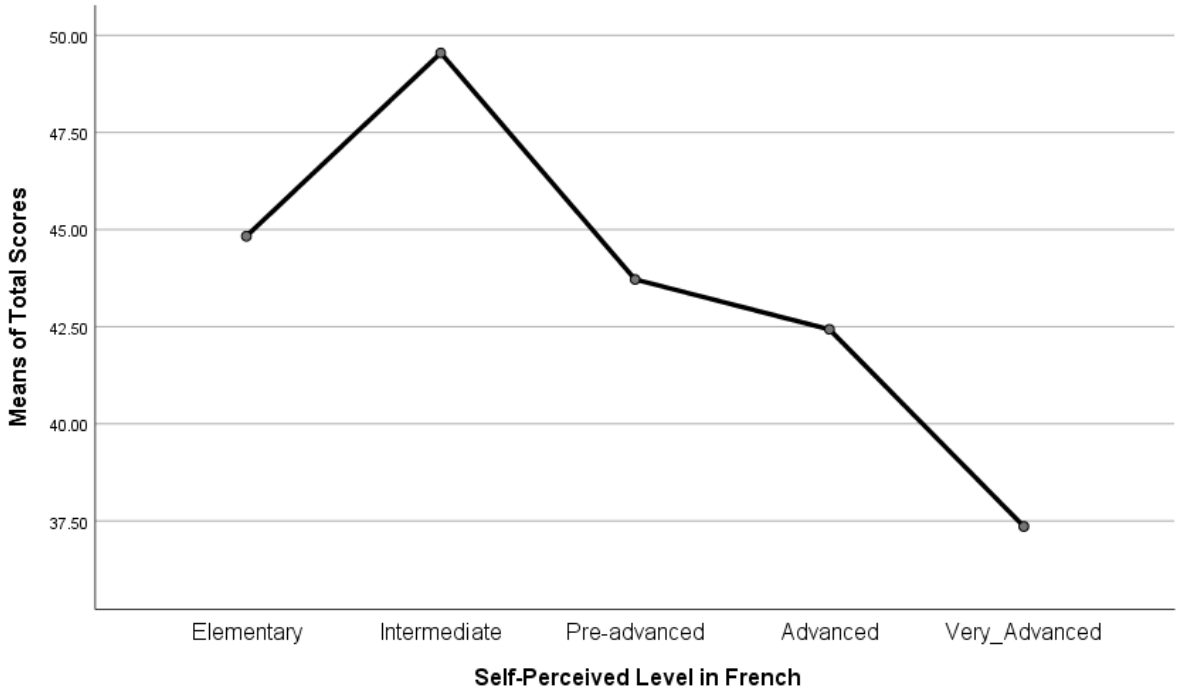

Figure 29: Mean percentages of the participants based on their self-perceived level in French

5.6 Conclusion

Both correct and incorrect produced answers can be linked to eight distinctive strategies adopted by the participants in the rendition of English collocations into Arabic and vice versa. They depended firstly on literal translation, which in most cases revealed instances of French interference in their rendition of English collocations into Arabic and rarely contributed to the production of correct equivalent answers. Omission was the second most commonly used strategy; this might be the quickest solution in undertaking a voluntary task for half an hour. Synonymy was the third most frequent strategy, due to the availability of many synonymous alternatives that were close enough in particular contexts allowing choices to be made between them without affecting the idiomaticity and acceptability of the collocation in the target language or equally maybe due to the participants' unawareness of the selectional

restrictions that collocations involve. Equivalence was the fourth most commonly employed strategy. This led in all cases to producing both acceptable collocations and acceptable translations. Substitution and approximation were the fifth and sixth most frequently used strategies. They most often led to incorrect answers as the latter always led to unacceptable translations, though acceptable collocations, and the former led to unacceptable translations that may or may not be acceptable collocations. Reduction and paraphrasing were the least opted for strategies with the latter being the least used. The minimal adoption of these two strategies can be linked to their non-conformity with the aim of the task, which was to produce equivalent target collocations to the source ones rather single words or lengthy sentences.

In this research, three types of errors were identified, as made by the 89 Algerian translation students in their attempts to produce accurate translations for English/Arabic collocations. These are :1/ lexical errors; 2/ grammatical errors; and 3/ spelling errors. Lexical errors were the dominant error type at 874 (91.04%). Spelling errors and grammar-related errors totalled just 34 (3.54%) and 52 (5.41%) respectively. The error analysis also revealed that a third of the errors were possibly due to French interference, most of them being lexical errors. These were the most frequent errors and can be mostly related to L2 interference, and the reliance on particular translation strategies like substitution, approximation, synonymy and reduction. While spelling errors can be often linked to L2 interference, grammatical errors may be due to either overgeneralization by generalizing grammatical rules, or L1 interference when an Arabic structure is transferred into English, such as misplacing the adjective after the noun in the rendition of English collocations into Arabic. In this study more errors occurred in the translation of the English collocations into Arabic than in translating from Arabic to English. This is illustrated by the difference in the mean error scores for the rendition of the English collocations into Arabic, which is 42.40%, as compared to that for Arabic collocations into English, which is 47.98%.

The better performance of the students in Arabic-English translation of collocations can be explained by the very few instances of French interference represented through spelling mistakes in this part of the test in comparison with the English-Arabic translations, where almost all the cases of French lexical interference were recorded.

6 Chapter six Conclusion

This chapter concludes this research, which investigated the interference of the French language as a Second Language (SL) in the translation of English collocational false friends into Arabic and vice versa by Algerian master's students. This was operationalised through automatically extracting collocations including false friends between English and French. In the following paragraphs, I will not go through the contents of thesis chapter by chapter, as this is already covered in chapter 1 (section 1.9). Rather I will identify key aspects of the thesis.

This concluding chapter firstly provides some insights into its results as a mixed-methods study as covered throughout chapter 5. It then discusses the challenges and reflections on the experience of this research. Then, it presents its contributions, limitations and, lastly, provides recommendations and suggestions for future studies.

The study has made a number of contributions to the fields of language transfer, corpus linguistics, and translation, which reflect the strengths of its theoretical, methodological and analytical framework respectively. These contributions are centred on different aspects. First is the novelty of the study's theoretical framework, which adopts a hybrid statistical and phraseological approach towards defining collocation as a phenomenon. This framework helped developed an automatic empirical extraction process for English and Arabic collocations, which made up the bulk of the translation test. Second is the originality of its two analytical frameworks, the first of which introduces an acceptability scale for collocations and translation, while the second creates a typology of translation strategies based on synthesising existing typologies and emerging patterns of solutions adopted by the participants of this study. Third is its methodology, and its use of a mixed-methods approach to understanding second language interference in third language learning through collocational false friends in the Algerian context.

6.1 Findings and implications

The task of the automatic extraction of collocations revealed the following:

- Log Dice and MI seem to share the same approach.
- Log Dice tries to compensate for the bias of low frequency, which is the main side effect of MI.

- Log Dice scores do not show as many technical terms as do MI scores.
- Log Dice can be regarded as the best association measure for exclusivity.
- Log-Likelihood and T-Score proved to have very similar results due to their very similar approach.
- Log-Likelihood looks to be the best measure among the associations of frequency.
- Although Log Likelihood is said to be the best of all the association measures of frequency, Log Dice outperforms Log Likelihood in terms of precision, as the former performs effectively even on raw (not pre-processed) corpora.

Both correct and incorrect answers can be linked to eight distinctive strategies adopted by the participants in the rendition of English collocations into Arabic and vice versa. The most frequent strategy was literal translation. While equivalence always led to correct answers by resulting in both acceptable collocations and translations, other strategies mostly led to wrong ones.

In the translation task, subjects used different translation techniques whose causes might be subject to different interpretations, none of which is definitive. The participants may have drawn on their first language, as it is the first resort for any foreign language learner or at least for their assumptions as to what the target language norms should be. Similarly, they might have lost patience when meeting with difficult items and therefore avoided giving any answers, as they did in 492 (18.92 %) of cases. At some points, they might have attempted to focus on the whole meaning of the sentence rather than the collocations themselves and thereby provided lengthier or shorter answers, as they did in 79 (3.03 %) of cases. At others, they could have tried to compensate for their inability to recall the target equivalent and then decided to approximate the source meaning either by providing a semantically close candidate, or by recklessly giving completely irrelevant answers, as they did in 261 (10.03 %) of cases.

As noted, respondents depended firstly on literal translation (45.15 % of cases), which in most cases revealed instances of French interference in the rendition of English collocations into Arabic and rarely contributed to the production of correct equivalent answers. Omission was the second most commonly used strategy (18.92 % of cases);

this might be the quickest solution in undertaking a voluntary task for half an hour. Synonymy was the third most frequent strategy (11.85 % of cases), due to the availability of many synonymous alternatives that were close enough in particular contexts allowing choices to be made among them without affecting the idiomaticity and acceptability of the collocation in the target language or equally maybe due to the participants' unawareness of the selectional restrictions that collocations involve. Equivalence was the fourth most commonly employed strategy (11 % of cases). This led in all cases to producing both acceptable collocations and translations. Substitution and approximation were the third and fourth least frequently used strategies (4.84 % and 5.19 % of cases respectively). They most often led to incorrect answers with the latter always leading to unacceptable translations but acceptable collocations, and the former leading to unacceptable translations that may or may not be acceptable collocations. Reduction and paraphrasing were the least opted for strategies, with the latter being the least used altogether (2.65% and 0.38 % of cases respectively). The minimal adoption of both strategies can be linked to their non-conformity with the aim of the task, which was to produce equivalent target collocations to the source ones rather single words or lengthy sentences.

Error analysis revealed that nearly a third (32.18%) of the errors were possibly due to French interference. Additionally, more errors occurred in the rendition of the English collocations into Arabic than of Arabic collocations into English. This is illustrated by the difference in the mean scores of the rendition of the English collocations into Arabic (42.40%) and Arabic collocations into English (47.98%). Lexical errors were the most frequent type of error (91.04 % of cases). This may well be related to L2 interference, and the reliance on certain translation strategies like substitution, approximation, synonymy and reduction, as explained in chapter 5 (section 5.4.2).

In this research, while spelling errors may be mostly linked to L2 interference, grammatical errors can be due to either overgeneralization by generalizing grammatical rules like the 'ing' suffix, for example, to form nouns, or L1 interference when an Arabic structure is transferred into English, as in the misplacing of the adjective after the noun in the rendition of English collocations into Arabic, as discussed in chapter 5 (section 5.4.3)

6.2 Challenges and reflections

The challenges of this research were related to its nature, as well as the application of the methodology. Being interdisciplinary in nature, this research posed a challenge in connecting three different research aspects related to different fields namely: (i) language transfer, particularly SL language interference in third language learning, (ii) corpus linguistics, and (iii) translation. Learning how to use Sketch Engine was a challenge, as was the use of the SPSS statistical package for social sciences, which I employed for both statistical tests and data visualization. Training was required for understanding how to run a variety of statistical tests and what purpose they serve. Gaining the necessary background in linguistics, how association measures work and choosing the best ones for the automatic extraction of collocations was a time-consuming undertaking. This is because it was necessary to develop an understanding of the reasoning underpinning each of these associations to select the one/ones that best highlight/s the collocational dimension this research is interested in. This was a steep learning curve. Moreover, the journey of searching for a theoretical framework that combines two different facets of language transfer, i.e. collocations and false friends, to comprehensively investigate non-positive language transfer was enriching for the researcher. Furthermore, the categorisation of translations for evaluative purposes was not straightforward. This is because the produced translations did not fit within the initial acceptability scale and therefore the coding scheme. As a result, I needed to work out how to alter the acceptability scale, and the classification of the data were revised accordingly multiple times to fit the emerging patterns of data.

Prior to the analysis, an initial coding scheme was developed. As the data evolved, the scoring scheme developed as well, to cope with multiple correct answers and to allow multiple responses to emerge naturally. Coding the data was complicated since it did not fall neatly into previously established categories. Although the data coding in this study was laborious, it provided a suitable middle ground between a complex, unquantifiable data-driven classification scheme and a predetermined score system, which would adjust the data according to the researcher's preconceived assumptions, and therefore undermine its own validity.

6.3 Contributions

This study has made the following contributions. I shall mention them first and proceed to explain each of them in this section:

1. Investigating SL interference in L3 learning in the Algerian context

The study of learners' collocation competency is not novel in itself, but its study in an L3 context is underrepresented in the literature. There are a number of limitations of previous research which this research seeks to address. First, little attention has been paid in previous research to the impact of L2 on L3 collocational competence. This study, therefore, fills a research gap regarding learners' difficulty in producing collocations in general and testing SL interference in L3 learning in particular. One major original feature of this study is thus that it provides a thorough investigation of the impact of L2 on L3 in using collocational false friends specifically in the Algerian context.

2. Exploring the difficulties of translating *adjective+noun* pattern

According to previous research on different collocational patterns, *adjective+noun* collocations are one of the main sources of difficult for learners, so further study of this pattern is essential (Zughoul and Abdul-Fattah, 2001; Alsakran, 2011).

3. A new conceptualisation for collocations and false friends

In terms of theoretical framework, one of the contributions of this research is that it provides a hybrid definition of collocations combining both statistical and phraseological approaches by including key defining aspects belonging to both approaches. These collocational relationships are later on mapped onto the existing association measures of collocations to automatically identify the collocations to be used for the study. Furthermore, unlike previous studies, this study employs false friends within their wider natural linguistic context of appearance as parts of collocations.

4. A systematic approach to the items chosen for the translation test

In previous research, collocational competence in ESL and EFL settings (e.g. Abdul-Fattah & Zughoul, 2003; Ahmed, 2012; Almaktari 2017, Alsakran 2011, Dukhali,

2016; Dweik & Abu Shakra, 2010; Jabak, Abdullah and Mustapha,2016; Hammadi, 2015; and Shammas, 2013) has been studied using corpora of students' writing or translation tests. Most of the studies that used translation tests did not adopt a systematic approach to the choice of test items. Thus, it is not clear how the test items were selected and sometimes the whole process of item selection is barely described, if at all. By contrast, this study adopts a rigorous and systematic approach to extracting collocations to draw sound, conclusive results.

5. Using the stratified sampling technique

This study uses a particular sampling technique, known as the stratified strategy, for deciding the sample representing the population. The study ended up creating a balanced test of good quality which can account for the differing levels of competence of the learners within the targeted sample as evidenced by the item test procedure and easiness index.

6. The efficiency of LD and LLD as association measures for extracting collocations.

This study presents a practical experiment about how four association measures function in extracting English and Arabic collocations. It then compares LD as being the best measure for exclusivity and LLD being the best for significant frequency. The study argues in favour of the efficiency of LD as the best measure for extracting collocations in both languages.

7. Creating an eightfold typology of translation strategies

Another contribution of this study is that it identifies an eightfold typology of translation strategies based on students' translation of collocations, on the basis of a review and critique of other typologies found in the literature.

8. Creating new linguistic resources

The research creates a comprehensive list of false friends between English and French. This list can be used in English, French and translation classes to raise awareness about false friends. This might benefit English learners who have considerable knowledge of French or have been exposed to French before they learn English. The

false friends in the compiled list can be used in teaching English as well as translation. Equally, this list can serve as a seed list for extracting bilingual collocations or for automatic training models.

9. Designing a balanced collocation test

The collocations chosen for the translation test vary across the scale of difficulty in that they range from easy to difficult items. Around half of the test items belong to the medium category. The remaining items belong to either the easy or difficult categories. This shows that our test is of good quality as it can account for different levels of competence among learners.

10. A range of different translation possibilities

For each of the collocations used in the test, the different answers have been recorded in the appendix. This can be used for computational purposes where these answers serve as input to train automatic models.

6.4 Limitations

Like all research, this research work has various limitations. Here, I will point to future research areas based on a discussion of these limitations. To gain a deeper understanding of learners' difficulties with lexical collocations in general, and false friends in particular , the limitations of the present study should be addressed in future research. The sample and instruments of data collection are included in these limitations, as is discussed below.

The first limitation of this study is related to the questionnaire not providing a precise evaluation of the levels of proficiency of participants in Arabic, French and English. The 89 participants come from different universities and are enrolled on a Master One programme with different bachelor's degrees in French, English, Arabic, and translation. In this light, a standardised test such as TOFEL or IELTS, and DELF, DALF would have been effective in getting a clearer understanding of their English and French proficiencies rather than asking students about their own perceived level in these languages. However, this was not possible due to the limited time and cost-ineffectiveness of administering such tests.

Second, the data of this study covered only 89 Master One students majoring in translation in the 2018-2019 academic year from only three institutions: Abu AL Kassem Saadallah Algiers 2 University, Oran Ahmed ben Ahmed University, and the High Arab Institution of Translation. This limits the generalizability of the outcome of this research to Algerian Master One translation students.

The stratified random sampling aimed at the accurate representativeness of the different populations especially with the two-layer stratification (groups) that our data involves. The first layer of population stratification (grouping) in our research is a sample of Master One students from three different institutions across Algeria. The second layer of data stratification is within the same university. However, the fact that the fieldwork took place in 2019, at a time when Algeria was subject to political instability due to a struggle for political change, led students from other cities who were attending the aforementioned institutions to head back to their home cities. This might have affected not only the randomness of our sample but also its size and to a lesser extent its gender balance. The overall group chosen for the study had more than 2/3 female students and less than 1/3 male participants. Although gender disproportionality in English and translation departments in our sample can be explained by societal and motivational factors, as mentioned in chapter 5, the political conditions under which the study was undertaken might have reduced the proportion of female participants.

Another limitation of this study is related to the number of test items in the translation test used to measure the learners' collocational knowledge. It seems that having more collocations to translate, especially Arabic collocations in the second part of the test, could have given more insights into the problems encountered by learners when rendering Arabic collocations into English. However, this was not possible as it was challenging to strike a balance between the time needed to do the test and the level of interest of participants, given that it was a voluntary task. In conclusion, these limitations will hopefully inspire researchers to improve this study or devise better measurements in the future. In order to further this study and provide possible directions for future research in this field, suggestions for further study are presented below.

6.5 Recommendations and suggestions for future studies

The acceptability scale of translations and collocations proposed in this study can be used as a training exercise for teaching students to avoid false friends, in translation and in Arabic and English classes. In addition to this, the theoretical framework for defining collocations adopted in this thesis has the capacity to inform future collocational studies that consider the benefit of the statistical approach, backing it up with the linguistic approach. This could also be extended to applying the relevant method of extraction of collocations using exclusive and frequently significant association measures. In addition, the eightfold typology of translation strategies can give insights into Arabic English translation in general and better inform English-Arabic translation teaching in general and in Algeria in particular.

Furthermore, the compiled list of false friends can serve as a linguistic resource that can be helpful for educational purposes. Thus these false friends can be useful for English or French language learners regardless of whether English or French is their first language; they can equally be helpful for English learners.

Algerian students, having learnt French as a SL and English as a L3, are highly likely to find false friends confusing, as are English learners who have considerable knowledge of French or have been exposed to French before they learn English. Therefore, it is vital to incorporate the false friends in the compiled list in teaching English as well as translation. Additionally, some flaws in language teaching and learning may be revealed by the students' use of false friends. Teachers do not seem to pay enough attention to the peculiar characteristics and collocations of these words, which are lacking in textbooks and other didactic materials (Roca-Varela, 2015).

Based on the outcome of this study, false friends and collocations should be more prominently discussed in English language and translation departments in Algerian universities by both teachers and curriculum designers. To this end, designing curriculums that effectively integrate collocations into a balanced course at each level of teaching English as a foreign language is of paramount importance.

Tutors of English-Arabic translation at Algerian universities do not seem to follow a unified curriculum in their teaching (Saadouni 2016). They should therefore co-operate to put together a unified curriculum that includes false friends and their typical lexical and grammatical relationships including collocational relationships.

Developing vocabulary generally and raising learners' awareness of false friends in particular, and collocational competence should begin at earlier stages. In this regard, more activities on collocations and false friends between English and French should be incorporated into EFL course books for different levels. For example, introducing false friends and collocations in the teaching materials allows for authentic language use and gives ample opportunities to practice using collocations in the classroom (El-Masharwi, 2008). The introduction of collocations should come in the context of covering a variety of topics and activities that are of interest to learners.

It is therefore recommended that language-teaching materials should provide explicit instructions on how false friends and collocations should be targeted in teaching and learning. Clearly, such materials should offer instructors teaching ideas for implementing the appropriate techniques to further enhance students' lexical proficiency. Additionally, such materials should be provided to instructors for the purpose of assessing how collocations should be addressed in the curriculum. Furthermore, it is vital to enhance the quality of teacher training and support to improve the effectiveness of their teaching, which, in turn, will have a direct impact on learners' levels of achievement and their learning experience.

It is important to incorporate corpora in language teaching and learning. For example, teachers can make use of concordances to contextualise false friends and collocations and provide more authentic contexts of occurrence for them in English. Concordances might also be used to provide information and statistics about language patterns across different genres and registers. Corpora are not only helpful in terms of teaching and contextualizing vocabulary, but they also give insights into language development and change across time and can thereby monitor recent developments in English. They can thus encourage learners to be independent have a sense of ownership over knowledge which they discover for themselves. Corpora therefore have huge potential to facilitate the learning of English. As Brezina and Flowerdew put it:

> Corpora can play a crucial role in Second Language (L2) research and pedagogy…these insights into L2 production that corpora offer can help make the process of language learning and teaching more effective and enjoyable (2017, pp. 1-2).

Equally, insights from this research and the linguistic resources identified above can be used for future research. For example, the compiled list of false friends can serve

as a seed list to automatically build comparable corpora, which can, in turn, be used for automatic training models. Moreover, the corpus-assisted use of false friends to automatically extract their respective collocations can be applied with all the false friends on the list which can be used as nodes to extract bilingual collocations.

References

Ahmed, Z.A. A. 2012. English lexical collocation knowledge of Libyan university students. PhD thesis, Bangor University.

Al-Athwary, A.A., 2021. False friends and lexical borrowing: A linguistic analysis of false friends between English and Arabic. *Journal of Language and Linguistic Studies, 17(S1*), pp.368-383.

Al-Barakati, M., 2013. Translation of the Sex-Related Qur'ānic Euphemism into English. Ph.D Thesis. Leeds: University of Leeds.

Albir, A.H., 2001. *Traducción y traductología*. Madrid: Cátedra.

Al-Kattan, A. B., 2007. The Notion of Collocation in English with Reference to Arabic. *Buhuth Mustaqbaliya Scientific Periodical Journal*, 4(1), pp.7-17/عدد18.

Almaktary, H., 2017. English collocations in the Tunisian EFL context. *International Journal on Studies in English language and literature*, 5(8), pp.4-20.

Alsakran, R. 2011. *The productive and receptive knowledge of collocations by advanced Arabic-speaking ESL/EFL learners*. Unpublished Master's thesis, Colorado State University.

Baisa, V., Jakubíček, M., Kilgarriff, A., Kovář, V. and Rychlý, P., 2014. Bilingual word sketches: the translate button. In *Proceedings of the 16th EURALEX International Congress* pp. 505-513

Baker, M., 2018. *In other words: A coursebook on translation*. London and New York :Routledge.

Baker, P., Hardie, A. and Mcenery, T. 2006. *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Barfield, A. 2003. Collocation Recognition and Production: Research Insights. Tokyo: Chuo University.

Beltrán, M. R. C. 2006. Towards a typological classification of false friends (Spanish-English). Revista Española de Lingüística Aplicada. (19), pp.29-40.

Benson, M. B. and Benson, E. E., &Ilson, R.1986. *The BBI Combinatory Dictionary of English. A Guide to Word Combinations. Amsterdam/Philadelphia :* John Benjamins Publishing

Bolarinwa, O.A., 2015. Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. *Nigerian Postgraduate Medical Journal*, 22(4), pp.195-201.

SBoumali, A., 2010. False friends: A Problem Encountered in Translation, the case of 3rd year students of Translation at Mentouri University. *Unpublished Master's Dissertation*, Mentouri University, Constantine.

Brenders, P., Van Hell, J. G. and Dijkstra, T. 2011. Word recognition in child second language learners: Evidence from cognates and false friends. Journal of Experimental Child Psychology, 109(4), pp.383-396.

Brezina, V. 2018. *Statistics in corpus linguistics: a practical guide*. Cambridge: Cambridge University Press.

Brezina, V. and Flowerdew,L., (eds), 2017. *Learner Corpus Research: New Perspectives and Applications*. London: Bloomsbury Academic.

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.

Bybee, J. L. and Hopper, P. J. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins Publishing.

Cela Gutiérrez, C., 2019. The translation of actually as an expression of evidentiality: a corpus-based analysis. *LFE. Revista de Lenguas para Fines Específicos.*25(1), pp.28-39

Chamizo-Domínguez, P. J. (2006). False friends. In Keith Brown (eds.), Encyclopedia of Language and Linguistics (2nd edn). Oxford: Elsevier, pp 426-429.

Chamizo-Domínguez, P. J. C. and Nerlich, B. 2002. False friends: their origin and semantics in some selected languages. Journal of Pragmatics, 34(12), pp.1833-1849.

Chesterman, A., 1997. *Memes of translation*. Amsterdam: John Benjamins.

Church, K.; Gale, W.; Hanks, P.; and Hindle, D. 1991. Using statistics in lexical analysis. In Lexical Acquisition: Using On-Line Resources to Build a Lexicon, edited by Uri Zernik. Hillsdale: Lawrence Erlbaum. pp115-163

Corder, S.P. 1967. *The Significance of Learners' Errors. International Review of Applied Linguistics in Language Teaching*, 5, pp.161-170.

Cowie, A. P. 1998. *Phraseology: Theory, analysis, and applications*. Oxford University Press: Oxford.

Creswell, J. W. 2009. *Research design: Qualitative, quantitative, and mixed methods approaches, 3rd ed. Research design: Qualitative, quantitative, and mixed methods approaches, 3rd ed*. Thousand Oaks, CA, US: Sage Publications, Inc.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 197-334.

Cruse, D. A. 2000. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press

Dickins, J. 2020. 'An ontology for collocations, formulaic sequences, multiword expressions, compounds, phrasal verbs, idioms and proverbs'. Linguistica Online. 23. pp. 29-27.

DSouza, J.[no date]. *Understanding the Differences Between Constructs, Variables, and Operational Definitions*.[online].[Accessed 17/11/2021].Available from https://www.statisticssolutions.com/understanding-the-differences-between-constructs-variables-and-operational-definitions/

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.

Dunning, T. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*(19)1, pp.61-74.

Dweik, B., & Abu-Shakra, M. 2010. Strategies in Translating Collocations in Religious Texts from Arabic into English. Atlas Global Journal for Studies and Research, pp.1–38. Eric, ED510365.

El-Dakhs, D.A.S., 2015. The lexical collocational competence of Arab undergraduate EFL learners. *International Journal of English Linguistics*, 5(5), p.60.

Ellis, R. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.

EI-Masharawi, A. 2008. Collocation Errors Made by English and Journalism Majors at the Islamic University of Gaza. Unpublished Master's dissertation. The Islamic University of Gaza.

Elmgrab, R.A. 2016. Possible criteria for evaluating students' translation errors. *International Journal of Humanities and Cultural Studies,* 1(3), pp.131-145.

Evert, S. 2005. The statistics of word cooccurrences: word pairs and collocations. Germany: Institute for Natural Language Processing University of Stuttgart.

Evert, S. 2008. Corpora and collocations. *Corpus linguistics. An international handbook*, 2, Germany: Institute of Cognitive Science, University of Osnabrück. pp.1212-1248.

Evert, S. and Kermes, H. 2003. Experiments on candidate data for collocation extraction. *In: Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 2: Association for Computational Linguistics*, pp.83-86.

Evert, S. and Krenn, B. 2003. Computational approaches to collocations. Introductory course at the European Summer School on Logic, Language, and Information (ESSLLI 2003), Vienna. Available at: http://www.collocations.de/EK/Articles/ESSLLI-1.ppt. (Website accessed on 11 May 2020).

Field, A. and Hole, G. 2002. How to design and report experiments. Los Angelos, London, New Delhi, Singapore: Sage.

Field, A.P. and Gillett, R., 2010. How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology,* 63(3), pp. 665-694.

Firth, J. R. 1957. A synopsis of linguistic theory 1930–55. In Studies in linguistic analysis, 24 pp1–32. The Philological Society, Oxford. Reprinted in Palmer (1968), pages 168-205

Fišer, D. and Ljubešić, N., 2013. Best friends or just faking it? Corpus-based extraction of Slovene-Croatian translation equivalents and false friends. *Slovenščina* 2.0, 1, pp. 50-77.

Frey, L., Botan, C.H. and Kreps, G., 2000. Investigating communication. New York: Allyn & Bacon.

Frunza, O. and Inkpen, D. 2010. Identification and disambiguation of cognates, false friends, and partial cognates using machine learning techniques. International Journal of Linguistics, 1(1), pp. 1-37.

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.

Gabrielatos, C. 2019. Lexicogrammar: Lexical Grammar or Construction Grammar? Two corpus-based case studies. Invited talk, UCREL Research Seminar, Lancaster University, 25 April 2019.

Ghazvini, S.D., 2011. Relationships between academic self-concept and academic performance in high school students. Procedia - Social and Behavioral Sciences, 15, pp.1034-1039.

González-fernández, B. and Schmitt, N. 2020. Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied linguistics*. **41**(4), pp.481–505.

González-Fernández, B. 2022. Conceptualizing l2 vocabulary knowledge: An empirical examination of the dimensionality of word knowledge. *Studies in second language acquisition*. **44**(4), pp.1124–1154.

Gouws, R., Prinsloo, D. and de Schryver, G.M., 2004. Friends will be friends: true or false: lexicographic approaches to the treatment of false friends. In *11th EURALEX International Congress (EURALEX-2004)* (pp. 797-806). Faculté des Lettres et des Sciences Humaines.

Granger, S. and Swallow, H. 1988. False friends: a kaleidoscope of translation difficulties. Langage et l'Homme, 23(2), pp.108-120.

Gries, S. T. 2010. Useful statistics for corpus linguistics. *A mosaic of corpus linguistics: Selected approaches*, 66, pp.269-291.

Gries, S. T. 2013. 50-something years of work on collocations. *International Journal of Corpus Linguistics*, 18(1), pp.137-166

Guay, F., Ratelle, C.F., Roy, A. and Litalien, D., 2010. Academic self-concept, autonomous academic motivation, and academic achievement: Mediating and additive effects. Learning and Individual Differences, 20(6), pp. 644-653.

Gutiérrez, C.C., 2018. *An analysis of the translation strategies of evidential adverbs in a corpus-based study* (Doctoral dissertation, Universidad de Las Palmas de Gran Canaria).

Gyllstad, H. 2007. Testing English Collocations : Developing Receptive Tests for Use with Advanced Swedish Learners. Språk- och litteraturcentrum, Lunds universitet.

Halliday, M. 1966. *Lexis as a linguistic level. In memory of JR Firth, edited by CE Bazell et al.*, 148-162. London: Longmans, Green & Co. Ltd.

Harouni, Z., 1998. Language transfer in translation: The case of lexico-semantic transfer and its impact on lexical correspondence/non-correspondence between languages. *مجلة العلوم الإنسانية*, pp.49-54.

Heid, U. and Prinsloo, D.J., 2008. Collocational false friends: description and treatment in bilingual dictionaries. In: Bernal, E. and DeCesaris, J.(eds.) *Proceedings of the 13th EURALEX International Conference, Barcelona*, 15-19 July 2008 pp. 1357-1364

Higgins, I. and Hervey, S.G., 1992. *Thinking Translation: Course in Translation Method-French to English*. London and New York: Routledge.

Hilu, A.L.J.N., 2018. False friends and translation: A case study of Iraqi students of translation. *Mustansiriyah Journal of Arts (MuJA)*, (82), pp.1-30.

House, J. 1977. A Model for Translation Quality Assessment. Tübingen: Gunter Narr.

House, J., 1986. Acquiring translational competence in interaction. Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies, Tübingen: Narr, pp.179-191.

Howarth, P. 1996. Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making. Lexicographica Series Maior 75. Tübingen: Max Niemeyer.

Ibrahim, Y., 2003. *The translation of collocation into Arabic problems and solutions* (Doctoral dissertation, University of Leeds).

Inkpen, D., Frunza, O. and Kondrak, G. 2005. Automatic identification of cognates and false friends in French and English. In: Proceedings of the International Conference Recent Advances in Natural Language Processing, pp.251-257.

Jääskeläinen, R., 1999. Tapping the process: An explorative study of the cognitive and affective factors involved in translating. Joensuun yliopisto.

Jabak, O., Abdullah, S.N.S. and Mustapha, N.F., 2016. The Difficulty of translating collocations from Arabic into English encountered by a sample of Arab students.

Jarvis, S. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in them interlanguage lexicon. Language Learning, 50(2), pp. 245–309.

Jarvis, S. 2009. Lexical transfer. In A. Pavlenko. eds.*The Bilingual mental lexicon: Interdisciplinary approaches*, pp. 99–124. Clevedon: Multilingual Matters.

Johansson, S. 2008. Contrastive analysis and learner language: A corpus-based approach. Oslo: University of Oslo.

Kerge, K. 2016. *Word-formation in the individual European languages: Estonian.– Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, Franz Rainer (Eds.), Word-Formation. An International Handbook of the Languages of Europe. Vol. 5. Handbooks of Linguistics and Communication Science 40.* Berlin, New York: De Gruyter.

Kilgarriff, A., Rychlý, P., Smrž, P. & Tugwell, D. 2004. The Sketch Engine. Proc. Euralex. Lorient, France.

Kjellmer, G. 1994. A Dictionary of English Collocations. Oxford: Clarendon Press.

Klégr, A. and Saldová, P., 2006. Collocational false friends. In: F. Čermák, M. Šulc (eds). *Collocations: studies in corpus linguistics.*pp. 168-177. Praha, NLN

Kovář, V., Baisa, V. and Jakubíček, M., 2016. Sketch engine for bilingual lexicography. *International Journal of Lexicography,* 29(3), pp.339-352.

Kuiper, K. and Allan, S. 1996. An Introduction to English Language: Sound, Word and Sentence. Hong Kong: Macmillan.

Lado, R.1957. Linguistics across cultures: Applied linguistics for language teachers. Michigan: The University of Michigan Press.

Langacker, R. W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites.* Stanford: Stanford University Press.

Larson, M. L. 1984. *Meaning-based translation: A guide to cross-language equivalence*. Lanham, MD: University Press of America.

Lewis, M. 1997. Pedagogical implications of the lexical approach. In Coady, J. and Huckin, T. (eds.), Second Language Vocabulary Acquisition, 255-270. Cambridge: Cambridge University Press.

Lörscher, W., 1991. Translation performance, translation process, and translation strategies: A psycholinguistic investigation, 4, Tübingen: Gunter Narr Verlag.

Lund, T. 2012. Combining qualitative and quantitative approaches: Some arguments for mixed methods research. *Scandinavian journal of educational research,* 56(2), pp.155-165.

Magno, C. 2009. Assessing the level of English language exposure of Taiwanese college students in Taiwan and the Philippines. *Asian EFL Journal.*

Makayev, K., Takhtarova, S., Gulkanyan, M., Kuznetsova, I. and Kazymova, T. 2020. The urgency of considering “false friends of a translator” for construction specialists. IOP Conference Series: Materials Science and Engineering. 890(1), pp.1-9.

Malblanc, A., 1968. Stylistique comparée du français et de l’allemand (= Bibliothèque de stylistique comparée 2), 5, Paris: Didier.

Manning, C. D. and Schütze, H. 1999. *Foundations of statistical natural language processing.* Cambridge: MIT press.

Martins, C. and NUNES, M.P., 2013. L2 to L3 transfer: Learner corpora analyses. Learner corpus studies in Asia and the world,**1**,pp.271-281

Masini, F. 2005. Multi-word expressions between syntax and the lexicon: The case of Italian verb-particle constructions. *SKY Journal of Linguistics*, 18(2005), pp.145-173.

Mason, I., 1994. Techniques of translation revisited: a text-linguistic review of" borrowing and modulation". In: Hurtado, A. eds. Estudis sobre la traducci. Castelló: Universitat Jaume I. pp.61-72.

Mathew, J.S., 2017. Self-perception and academic achievement. *Indian Journal of Science and Technology*, *14*(14), pp.1-6..

Maurer-Stroh M. 2004. *Towards a bilingual adjective-noun collocation dictionary of English and German*. PhD thesis, (English and American Studies)-Universität Klagenfurt, Klagenfurt.

McCarthy, M. J. and O'Dell, F. 2005. English Collocations in Use. Cambridge: Cambridge University Press.

McEnery, T. and Hardie, A. 2012. *Corpus linguistics: method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T. and Wilson, A. 2001. *Corpus linguistics: an introduction (2nd ed.).* Edinburgh: Edinburgh University Press..

McKeown, K.R. and Radev, D.R., 2000. Collocations. *Handbook of Natural Language Processing*. pp.507-523. New York: Marcel Dekker.

Meara, P. and Miralpeix, I. 2016. *Tools for researching vocabulary*. Multilingual Matters.

Mel'čuk, I. 1998. Collocations and lexical functions. *Phraseology. Theory, analysis, and applications*,Oxford: Clarendon Press. pp.23-53.

Memišević, A. and Margić, B.D. 2011. Impact of Language Proficiency on Error Rate in Croatian-English False Friends. In: M. B. Vukanovic and I. Vodopija-Krstanovic, eds. *The global and local dimensions of English: Exploring issues of language and culture* , 4, pp.69–82. Münster, Germany: Lit Verlag.

Mertens, D. M. 2012. Transformative mixed methods: Addressing inequities. *American Behavioral Scientist*, 56(6), pp.802-813.

Messaoudi. S. Adjectives In Modern Standard Arabic and English: A contractive study. Unpublished Master's dissertation. Annaba University

Messaoudi, S., Brierley, C. and Dickins, J., 2020, July. Evaluation of Association Measures in Automatic Extraction of Collocations in Arabic. In: *Arabic Linguistics Forum Conference* June 30th -July 2nd, pp.1-78 Leeds:University of Leeds.

Mitchell, T. F. 1971. Linguistic ‘goings-on’: Collocations and other lexical matters arising on the syntagmatic record. *Archivum Linguisticum*, 2(1), pp.35-69.

Mitkov, R., Pekar, V., Blagoev, D. and Mulloni, A. 2007. Methods for extracting and classifying pairs of cognates and false friends. Machine translation, 21(1), p29.

Mohammed, S. a. S. 2012. *A corpus-based study of the collocational behaviour and idealogical usage of political terms in the Arabic news discourse of pre-revolutionary Egypt*. PhD thesis, Lancaster University.

Molina, L. and Hurtado Albir, A., 2002. Translation techniques revisited: A dynamic and functionalist approach. *Meta: Journal des Traducteurs/Meta: Translators' Journal*, 47(4), pp.498-512.

Nation, P. 2001. Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

Nation, P. 2019. The different aspects of vocabulary knowledge *In*: *The Routledge Handbook of Vocabulary Studies*. London, England: Routledge, pp.15–29.

Negadi, M. N. 2015. Learning English in Algeria through French-based background proficiency. Procedia-Social and Behavioral Sciences, 199, pp.496-500.

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. 2004. What are collocations? In Allerton, D.J, Nesselhauf, N. and Skandera, P. (eds.), Phraseological Units: basic concepts and their application, 1-21. Basel: Schwabe.

Nesselhauf, N. 2005. Collocations in a Learner Corpus. Amsterdam: John Benjamins.

Newmark, P., 1988. *A textbook of translation*, 66, New York: Prentice hall.

Nida, E.A., 1964. Toward a science of translating: with special reference to principles and procedures involved in Bible translating. Leiden: Brill .

Nord, C., 1991. Text Analysis in Translation: Theory. *Methodology, and Didactic Application of a Model for Translation-Oriented Text Analysis*, 94, Amsterdam/Atlanta GA: Rodopi.

Nuccorini, S., 2006. In search of ‘phraseologies’ Discovering divergences in the use of English and Italian ‘true friends’. *European Journal of English Studies*, 10(1), pp.33-47.

Nunberg, G., Sag, I. A. and Wasow, T. 1994. Idioms. *Language*, 70(3), pp.491-538.

O'Brien, S. and Saldanha, G. 2014. Research methodologies in translation studies. London: Routledge.

Pastor, G. C. 2017. Collocational Constructions in Translated Spanish: What Corpora Reveal. In: R. MITKOV, ed. *Computational and corpus-based phraseology*. pp.29-40. London: Sprigner

Pawley, A., and Syder, F. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Richards, J. and Schmidt, R. (eds.), Language and Communication. pp203-239. Harlow: Longman

Pecina, P. 2010. Lexical association measures and collocation extraction. *Language Resources and Evaluation*, 44(1), pp.137-158.

Pérez-Paredes, P., 2020. Corpus Linguistics for Education: A Guide for Research. London: Routledge.

Petrović, S., Šnajder, J. and Bašić, B. D. 2010. Extending lexical association measures for collocation extraction. *Computer Speech & Language,* 24(2), pp.383-394.

Qualtrics.2020.Calculating sample size: a quick guide. [Online]. Accessed [17 November 2021] available from https://www.qualtrics.com/blog/calculating-sample-size/

Ringbom, H. 1987. The role of the first language in foreign language learning. Clevedon: Multilingual Matters.

Ringbom, H. 2001. Lexical transfer in L3 production. In J. Cenoz, B. Hufeisen, & U. Jessner .eds. Crosslinguistic influence in third language acquisition: Psycholinguistic perspectives, pp. 59–68. Clevedon: Multilingual Matters.

Roca-Varela, M.L., 2014. Faux amis in speech and writing: a corpus-based study of English false friends in the production of Spanish students. In: Kimps, D. ed. *Recent Advances in Corpus Linguistics: Developing and Exploiting Corpora*, pp. 293-307. Brill, New York: Rodopi.

Roca-Varela, M.L., 2015. False friends in learner Corpora: A corpus-based study of English false friends in the written and spoken production of Spanish learners. Bern: Peter Lang.

Rua, P.L., 2006. The sex variable in foreign language learning: an integrative approach. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (6), pp. 99-114

Rychlý, P., 2008, December. A Lexicographer-Friendly Association Score. In P. Sojka & A. Horák (Eds.), *Proceedings of recent advances in Slavonic natural language processing, RASLAN* (pp. 6– 9). Brno: Masaryk University.

Sadouni,R. 2016. Arabic-English Translation and French Influence in Algerian Universities: University of Algiers 2 as a Case Study. ASJP *(Algerian scientific journal platform*), 3, (1) pp.130-140.

Saeed, J. 2003. Semantics. Oxford: Blackwell.

Scott-Tennent, C. and Torras, F.R., 2000. Translation Strategies and Translation Solutions: Design of a Teaching Prototype and Empirical Study of its Results. In *Investigating Translation: Selected Papers from the 4th International Congress on Translation, Barcelona,* 1998 (Vol. 32, pp. 107-116). Amsterdam-Philadelphia: John Benjamins Publishing.

Séguinot, C., 1991. A study of student translation strategies. In: Tirkkonen-Condit,S. (ed.), *Empirical research in translation and intercultural studies*.. pp.79-88. Tübingen: Gunter Narr Verlag

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.

Shehata, A.K., 2008. *L1 influence on the reception and production of collocations by advanced ESL/EFL Arabic learners of English*. Master's Dissertation, Ohio University.

Schmitt, N. 2019. Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language teaching*. **52**(02), pp.261–274.

Siepmann, D. 2005. Collocation, colligation and encoding dictionaries. Part I: lexicological aspects. International Journal of Lexicography 18(4). Pp. 409-443.

Sikogukira, M., 1993. Influence of Languages Other than the L1 on a Foreign Language: A Case of Transfer from L2 to L3. *Edinburgh working papers in Applied Linguistics*, 4, pp.110-132.

Sinclair, J. 2004. *Trust the text: Language, corpus and discourse.* London: Routledge.

Sinclair, J. M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Soglasnova, L. 2018. Dealing with False Friends to Avoid Errors in Subject Analysis in Slavic Cataloging: An Overview of Resources and Strategies. Cataloging & Classification Quarterly, 56 (5-6), 404-421.

Swaminathan, N., 2008. Girl talk: Are women really better at language? *Scientific American*. Available from https://www.scientificamerican.com/article/are-women-really-better-with-language/.

Tashakkori, A., & Teddlie, C. 1998. *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Thody, P., Evans, H. and Rees, G. 1985. Faux Amis and Key Words: *A Dictionary-guide to French Life and Language Through Lookalikes and Confusables*. London:Athlone press

Thorndike, R.L. and Hagen, E.P. 1977. *Measurement and evaluation in psychology and education* 4th ed. Nashville, TN, USA: John Wiley & Sons.

Trimasse, N., 2019. The source of lexical transfer in L3 production in a diglossic context. *International journal of multilingualism*, 16(4), pp.398-410.

Tymoczko, M., 2005. Trajectories of research in translation studies. Meta: journal des traducteurs/Meta: *Translators' Journal*, 50(4), pp.1082-1097.

Van Roey, J., Granger, S. and Swallow, H., 1998. *Dictionnaire des Faux Amis/Dictionary of Faux Amis Français-Anglais English-French*. Paris: Duculot.

Vázquez-Ayora, G., 1977. *Introducción a la traductología: curso básico de traducción*. Washington D.C.: Georgetown University Press.

Versta Research Newsletter. 2011.How to estimate the length of a survey. [Online]. [Accessed 15 January 2019]. Available from https://verstaresearch.com/newsletters/how-to-estimate-the-length-of-a-survey/

Vinay, J.P. and Darbelnet, J., 1958. Stylistique compare du français et de l'anglais (Stylistic comparison between French and English). Paris: Didier.

Vinay, J.P. and Darbelnet, J., 1995. *Comparative stylistics of French and English: A methodology for translation.* Amsterdam-Philadelphia: John Benjamins Publishing.

Wagner, J., 2004. A False Friends Exercise with Authentic Material Retrieved from a Corpus. In: Proceedings of InSTIL/ICALL 2004 Symposium on NLP and Speech Technologies in Advanced Language Learning Systems. Venice (Italy), June 17–19,2004. Retrieved from https://www.iscaspeech.org/archive_open/archive_papers/icall2004/iic4_027.pdf Accessed on 17/02/2022

Wang, X. 2017. *Quotation and Truth-Conditional Pragmatics*. London and New York: Routledge.

Wang,X. 2018. Quotation and truth-conditional pragmatics. London and New York: Routledge.

Wilss, W., 1996. *Knowledge and skills in translator behavior* (Vol. 15). Saarbrucken:

John Benjamins Publishing.

Woll, N., 2018. Investigating dimensions of metalinguistic awareness: what think-aloud protocols revealed about the cognitive processes involved in positive transfer from L2 to L3. *Language Awareness*, 27(1-2), pp.167-185.

Wotjak, G., 1981. Técnicas de translación. Aspetos fundamentales de la teoría de la traducción. La Habana: Ediciones Pueblo y Educación, pp.197-229.

Wray, A. 2002. Formulaic Language and the Lexicon. Cambridge: Cambridge University Press.

Wulff, S. 2008. Rethinking Idiomacity: A Used-based Approach. In: W. T. A. M. MAHALBERG, ed. London and New York: Continuum, p.238.

Zabalbeascoa, P., 2000. From techniques to types of solutions. In Beeby, A., Ensinger, D. & Presas, M. (eds.) *Investigating translation*. 32, pp.117-128. Amsterdam-Philadelphia: John Benjamins,

Zughoul, M.R. and Abdul-Fattah, H.S. 2001. Collocational Competence of Arabic Speaking Learners of English: A Study in Lexical Semantics. ERIC, 19. Retrieved from: http://files.eric.ed.gov/fulltext/ED479650.pdf.

Appendix A: The questionnaire in English

1. Gender F ☐ M ☐
2. Age:
3. The first language spoken at home

A/ Darija ☐ B/ Berber ☐ C/ a mix of both ☐

For how long have been studying English ……. Years.

2. Have you been taught by native English language teachers

Yes ☐ No ☐

3. If yes, for how long:

A/ Less than a year ☐ B/ 1 to 2 years ☐

C/ 2 to 3 years ☐ D/ more than 3 years ☐

4. Did you attend any English training programs in Algeria?

5. Yes ☐ No ☐

6. If yes, name the institution and for how long it ……

8. Have you ever travelled to an English speaking country?

Yes ☐ No ☐

9. If yes, specify where and for how long ……..

10. How much time on average you spend daily on watching English programs?

None ☐ Less than 1 hour ☐

From 1 to 2 hours ☐ More than 2 hours ☐

11. Have you ever had friends who are native English speakers?

Yes ☐ No ☐

12. If yes, how much time do you spend chatting with them in English Daily?

None ☐ Less than 1 hour ☐

1 to 2 hours ☐ More than 2 hours ☐

13. What is the commonest language you use for surfing the internet?

Arabic ☐ French ☐ English ☐ Arabic-French ☐
Arabic-English ☐ French-English ☐ Arabic-French-English ☐
Others……

14. How much time do you spend, on average, surfing the net in English Daily?

None ☐ Less than 1 hour ☐

1 to 2 hours ☐ More than 2 hours ☐

15. What is your educational background (Bachelor degree)?

English language degree ☐ French language degree ☐

Arabic language degree ☐ Translation studies degree ☐
Other…….

16. What is the medium of instruction inside the class used by the teacher?

Arabic ☐ French ☐ English ☐ Arabic-French ☐
Arabic-English ☐ French-English ☐ Arabic-French-English ☐
Other ……

17. What is the most popular language pair or pairs you work with in class (tick all that apply)?

English Arabic ☐ French Arabic ☐

Both pairs ☐ Both +others ☐

18. On a 5. Point scale, how do you evaluate your language proficiency for each of these languages

	Elementary	intermediate	Pre-advanced	Advanced	Very advanced
my proficiency in English					
my proficiency in French					
my proficiency in Arabic					

Thank you for your cooperation

Appendix B: The questionnaire in Arabic

1. الجنس: ذكر انثى
2. العمر:.....
3. اللغة الاولى التي نطقت بها:

ا/ الدارجة ☐ 2/ الامازيغية ☐ 3/ مزيج من اللغتين معا ☐

4. المدة الزمنية التي امضيتها في تعلم اللغة الانجليزية:..... سنوات
5. هل قام بتدريسك اساتد ة ناطقين باللغة الانجليزية " الانجليزية لغتهم الام "

ا/ نعم ☐ ب/ لا ☐

6. ادا كان الجواب نعم كم من الوقت امضيت تحت اشراف هؤلاء الاساتدة

أ/ اقل من سنة ☐ ب/ من1 الى2 سنتين ☐

ج/ من 2 الى 3 سنوات ☐ د/ اكثر من ساعتين ☐

7. هل حضرت دورات تعليمية للغة الانجليزية في مراكز متخصصة في تعليم الانجليزية في بلدك

ا/ نعم ☐ ب/ لا ☐

8. ادا كان الجواب نعم كم كانت مدة الدورة و اين.....
9. هل سافرت الى احد البلدان الناطقة باللغة الانجليزية

ا/ نعم ☐ ب/ لا ☐

10. ادا كان الجواب نعم الى اين سافرت و كم دامت مدة سفرك هناك .....
11. كم من الوقت تمضي يوميا في مشاهدة برامج باللغة الانجليزية

ا/لا اشاهد على الاطلاق ☐ ب/ اقل من ساعة ☐

ج/ من 1 ساعة الى 2ساعتين ☐ د/ اكثر من ساعتين ☐

12. هل لديك اصدقاء ناطقين باللغة الانجليزية على مواقع التواصل الاجتماعي ؟

ا/نعم ☐ ب/ لا ☐

13. كم من الوقت تمضي يوميا في محادثتهم باللغة الانجليزية

ا/ على الاطلاق ☐ ب/اقل من ساعة ☐ ج/ من 1 الى2 ساعة ☐ د/ اكثر من ساعتين ☐

14. ما هي اللغة التي تستخدمها عادة اكثر عند تصفح المواقع على شبكة الانترنت

ا/ اللغة العربية ☐ ب/ اللغة الفرنسية ☐ ج/ اللغة الانجليزية ☐

15. ما نوع شهادة الليسانس البكالوريوس التي تحصلت عليها قبل التحاقك بقسم الترجمة

لغة عربية و ادابها ☐ لغة فرنسية و ادابها ☐ لغة انجليزية و ادابها ☐

شهادة ترجمة ☐ شهادة اخرى .....

16. ما هي اللغة سائدة الاستعمال في المحاضرة و الدروس من قبل الاستاذ

ا/ اللغة العربية ☐ ب/ اللغة الفرنسية ☐ ج اللغة الانجليزية ☐ د/اخرى.....

17. ما هي اكثر لغتين تترجم منهما و اليهما في القسم تستطيع ان تختار اكثر من خيار؟

عربية فرنسية ☐ انجليزية عربية ☐ اخرى .....

18. على مقياس تقييم من خمسة نقاط كيف تقيم اجادتك لكل من هده اللغات

	مستوى اساسي	مستوى متوسط	فوق المتوسط	مستوى متقدم	مستوى ممتاز
مستواي في اللغة العربية					
مستواي في اللغة الفرنسية					
مستواي في اللغة الانجليزية					

شكرا على تعاونكم

Appendix C: Translation Test

1. Translate only the underlined expression from each sentence into Arabic

01 This **Heinous crime** must not go unpunished…….

02 Nathan is very **Sensible person**. He never does anything wild and spontaneous……..

03 Maintenance fees must be considered to determine the **Actual Cost**…….

04 **Concurrent enrolment** allows high school students to be enrolled in college……..

05 He looked around as if he is trying to spot a **Candid camera** or something……

06 Educators must launch a **Comprehensive approach** to language teaching and learning……

07 They need to ensure that the products are working in a **Consistent manner**……

08 The **Eventual winner** will earn 100.000 euros and a diploma……

10 She is a very calm and **Fastidious person** ……..

11 He is a **Formidable opponent** at chess……

12 It was a **rude awakening** to learn after I left home that I wasn't so special anymore…….

13 The lucky winners of the highly anticipated **Grand Prize** Draw have been announced…….....

14 The five **inhabited islands** had no buses…….

15 Your mother is an easy - going and **jolly woman/lady/**……..

16 The province will assume the lease of all the **Rentable space** in the building…….

17 **Secular humanism** is a philosophy that embraces human reason and rejects religion……..

18 He deserved more than the **sympathetic ear** she had been lending him lately…….

19 She definitely had an **ulterior motive** in offering to help…….

20 When the movie started, the **petulant child** would not stop crying……

2. Translate only the underlined expression from each sentence into English

1. كانت السيدة في **مزاج متعكر** لا يسمح لها بالدخول في النقاش.
……..……..……
2. انتشرت مؤخرًا صور **خداع بصري** فى مواقع الانترنت.
…… ……..……
3. تثير **التوجهات السياسية** الجديدة في الشرق الاوسط جدلا واسعا……..……
4. تعاني الأنروا من **عجز مالي** كبير وقد تلجأ للتقشف……..……
5. كان موقفه **عقبة كؤود** في وجه الحل الشامل والإصلاح.
……..……..
6. يملك **احترافية عالية** في تسيير مؤسسته……
7. كان عمر محققاً ذو **مهارة فائقة** في حل القضايا الشائكة……..……
8. قدم المتهم **حجة دامغة** تدل على برائته……
9. يبدو أن هذا **الجرح النازف** لن يلتئم بسرعة
……..……..……
10. قدم المسؤول **أعذارًا واهية** لإخفاقاته
……

Appendix D:The automatic extraction of collocations by Log Likelihood and Log Dice agreement

The Automatic Extraction of the English Collocations by Log-Likelihood and Log Dice Agreement

Sensible person

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
eminently	679	14,616	26.053	12.672	10,605.91	7.256
precaution	642	95,408	25.31	9.884	7,522.19	6.538
safeguards	174	5,589	13.187	12.094	2,575.19	5.384
heat	1,625	1,315,260	40.079	7.439	13,549.60	5.202
economically	364	213,319	18.999	7.905	3,265.71	5.114
perfectly	587	439,965	24.098	7.55	4,979.69	5.074
footwear	216	89,820	14.653	8.4	2,085.50	5.003
drinking	409	337,655	20.104	7.41	3,390.46	4.838
sentencing	166	70,339	12.845	8.373	1,596.45	4.756
shoe	694	724,621	26.148	7.071	5,430.47	4.732

Table 1**:** Top ten collocates for ‘sensible’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
a	17,964	329,105,715	123.347	3.649	61,568.91	0.838
way	3,281	15,194,862	56.126	5.633	19,324.69	2.815
approach	2,181	4,606,887	46.272	6.765	16,194.72	3.93
heat	1,625	1,315,260	40.169	8.149	15,155.68	5.254
thing	2,326	8,510,451	47.46	5.973	14,749.86	3.149
more	3,093	40,350,816	52.458	4.139	12,021.20	1.325
most	2,188	16,233,850	45.266	4.953	10,846.72	2.135
solution	1,513	4,034,816	38.446	6.429	10,523.64	3.59
choice	1,243	2,459,835	34.952	6.859	9,375.92	4.003
any	1,849	16,049,912	41.376	4.726	8,598.11	1.909

Table 2: Top ten collocates for ‘sensible’ sorted by Log likelihood

When sorting the results by Log Dice and Log Likelihood, none of the words occurred in the two lists apart from ‘heat’ which when associated with ‘sensible’ gives us ‘sensible heat’, which is a kind of energy. The noun ‘heat’ will not be considered in this case.

Because no agreement is obtained between the Log likelihood and the Log dice lists, an alternative method of using bilingual word sketch (in section 4.6.1) is used to find a noun collocate for the adjective ‘sensible’.

Actual cost

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
cost	14,091	5,534,321	116.081	5.499	80,084.07	6.135
fact	9,455	4,410,237	94.683	5.251	50,517.56	5.83
physical	5,399	2,281,764	71.729	5.393	29,866.33	5.737
implementation	4,452	1,780,933	65.22	5.472	25,102.40	5.695
an	87,983	59,415,245	285.343	4.717	413,364.50	5.575
expenditure	1,705	360,253	40.8	6.393	11,753.47	5.323
number	12,883	9,589,357	108.747	4.576	57,219.83	5.312
situation	4,432	2,719,885	64.273	4.855	21,298.46	5.273
content	5,113	3,507,869	68.743	4.694	23,474.78	5.204
size	4,079	2,598,393	61.576	4.801	19,306.89	5.201

Table 03**:** Top ten Collocates of ‘actual’ Sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
the	473,613	936,353,281	611.607	3.167	1,451,765.08	4.049
an	87,983	59,415,245	285.343	4.717	413,364.50	5.575
cost	14,091	5,534,321	116.081	5.499	80,084.07	6.135
number	12,883	9,589,357	108.747	4.576	57,219.83	5.312
fact	9,455	4,410,237	94.683	5.251	50,517.56	5.83
physical	5,399	2,281,764	71.729	5.393	29,866.33	5.737
no	10,399	17,910,918	92.088	3.366	29,841.82	4.168
or	17,514	61,023,201	106.384	2.35	29,098.39	3.209
a	1,358	329,105,715	-465.857	-3.769	27,526.61	-2.891
value	6,466	5,329,161	76.68	4.429	27,420.96	5.057

Table 04: Top ten Collocates of ‘actual’ Sorted by Log Likelihood

As the tables above show, the collocates given the highest scores by Log Likelihood are function words. However, most of the commonest collocates for “actual” sorted by Log Dice are nouns. A closer look at the results also demonstrates that the Log Likelihood and t-test scores are positively correlated to a great extent. This can be noticed as well with the scores given by log Dice and MI, but to a lesser extent. Not

all the collocates given a high score by Log Dice are given high scores by MI as well. Therefore, sorting the collocates by Log Dice and MI does not result in the same order for collocates. This, however is not the case with the lists generated by Log Likelihood and T-Score which give nearly the same result with some exceptions.

The noun ‘cost’ appears in both lists as a candidate collocate for the adjective ‘actual’. This list is sorted by Log Dice scores. ‘Actual cost’ means the real non-virtual cost existing in fact and does not convey the meaning of the current cost which might be inferred from the French sense of the word *actuel*.

Concurrent enrolment

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
Enrollment	4,490	324,763	66.983	11.434	62,484.95	8.464
Sessions	778	87,083	27.877	10.804	10,112.02	7.154
Validity	744	153,304	27.248	9.924	8,759.42	6.635
chemoradiation	250	2,013	15.81	14.602	4,593.83	6.447
chemoradiotherapy	245	1,623	15.651	14.883	4,604.78	6.424
Storyline	330	58,323	18.149	10.146	3,985.20	6.17
jurisdiction	912	372,825	30.137	8.936	9,489.06	6.007
Session	5,258	3,211,534	72.29	8.356	50,744.62	5.704
resolution	2,115	1,302,997	45.847	8.344	20,301.42	5.634
chemotherapy	322	132,790	17.907	8.923	3,342.61	5.553

Table 05: Top ten collocates for ‘concurrent’ sorted by Log Dice

collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
enrollment	4,490	324,763	66.983	11.434	62,484.95	8.464
Session	5,258	3,211,534	72.29	8.356	50,744.62	5.704
With	10,119	119,656,805	94.653	4.081	39,248.84	1.469
resolution	2,115	1,302,997	45.847	8.344	20,301.42	5.634
User	2,306	5,302,300	47.469	6.444	16,100.62	3.808
Multiple	1,420	1,985,430	37.419	7.162	11,301.51	4.485
programming	1,173	855,989	34.124	8.1	10,851.04	5.341
Sessions	778	87,083	27.877	10.804	10,112.02	7.154
jurisdiction	912	372,825	30.137	8.936	9,489.06	6.007
Or	3,101	61,023,201	50.214	3.347	8,882.88	0.733

Table 06: Top ten collocates for ‘concurrent’ sorted by Log Likelihood

The adjective “concurrent” collocates exclusively and significantly with the noun “enrollment”, using both Log Dice or Log Likelihood. The noun “sessions” emerged as the second exclusive candidate collocate of ‘concurrent’. One of the most interesting things about the results is that T-Score gives the word “session” the highest score among the list. Log Likelihood, however, does not assign the highest score to “session” although T-Score and Log Likelihood are similar in principle. A Google search yields about 769,000 results for “concurrent sessions”. The collocation “concurrent enrollment”, which is assigned the highest score by Log Likelihood and Log Dice yields about 992,000 results on Google. Accordingly, “concurrent enrollment” is more frequent in the language than “concurrent sessions”. The T- score in this case is affected by the corpus and the raw frequency and Log Likelihood gives a more valid result.

While the adjective *concurrent* in French language means ‘competitor’ or ‘contestant’, the meaning of ‘concurrent’ in English is ‘simultaneous’. Translating ‘concurrent sessions’ or ‘concurrent enrolment’ are examples where the interference of French can be tested.

Candid camera

collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
Photo	4,056	3,054,312	63.609	9.679	46,905.22	5.429
conversation	1,872	1,308,013	43.217	9.787	21,781.90	5.518
Interview	1,425	2,257,769	37.652	8.606	14,228.06	4.351
A	4,810	329,105,715	61.667	3.173	13,266.11	-1.062
discussion	1,404	2,976,463	37.341	8.185	13,201.76	3.935
shot	880	864,495	29.617	9.295	9,609.93	5.01
camera	719	1,450,310	26.726	8.257	6,815.22	3.992
feedback	588	977,958	24.183	8.535	5,797.18	4.256
frame	595	1,110,827	24.318	8.369	5,729.17	4.095
moment	508	2,220,840	22.379	7.141	4,029.39	2.886

Table 07: Top ten collocates for ‘candid’ sorted by Log Likelihood

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
sundress	61	2,551	7.809	13.849	1,050.78	5.949
conversation	1,872	1,308,013	43.217	9.787	21,781.90	5.518
photo	4,056	3,054,312	63.609	9.679	46,905.22	5.429
shot	880	864,495	29.617	9.295	9,609.93	5.010
interview	1,425	2,257,769	37.652	8.606	14,228.06	4.351
memoir	96	128,900	9.776	8.844	985.901	4.309
portrait	226	352,713	14.995	8.627	2,254.10	4.275
feedback	588	977,958	24.183	8.535	5,797.18	4.256
snap	166	275,110	12.849	8.541	1,635.44	4.157
camera	719	1,450,310	26.726	8.257	6,815.22	3.992

Table 08: Top ten collocates for ‘candid’ sorted by Log Dice

The noun ‘camera’ is one of the strongest and most exclusive collocates for the adjective ‘candid’. However, it is not the most frequent and this is the reason for which Log Likelihood does not assign it the highest score in the list. The strength of this collocation can be demonstrated through the highest score given to ‘camera’ by MI (9.333) as pictured below, given that MI is the association measure of strength. The noun ‘photos’ is the second most exclusive candidate by log Dice, given that Log Dice is the best measure of exclusivity. ‘Candid photos’, by contrast, looks to be the most significant collocate for “candid” because it is the most frequent. A Google search gives 9,290,000 results for ‘candid photos’. It also shows however, that ‘candid camera’ gets a score of 3,980,000.

Comprehensive approach

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
overview	13,443	736,092	115.619	8.48	131,675.96	8.037
immigration	10,441	618,256	101.871	8.367	100,593.36	7.778
coverage	7,918	986,437	88.416	7.294	64,467.29	7.073
examination	7,020	825,845	83.281	7.377	57,952.37	7.025
approach	21,213	4,606,887	144.029	6.493	149,558.67	6.969
assessment	11,261	2,072,674	105.119	6.731	82,973.61	6.936
guide	14,094	2,969,737	117.439	6.536	100,106.11	6.884
understanding	10,680	2,135,482	102.287	6.612	76,929.52	6.83
plan	30,959	8,670,354	173.432	6.126	203,005.42	6.721
cancer	9,646	2,264,195	97.035	6.38	66,404.18	6.624

Table 09: Top ten collocates for ‘comprehensive’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
A	366,166	329,105,715	577.315	4.443	1,710,510.16	5.184
plan	30,959	8,670,354	173.432	6.126	203,005.42	6.721
most	33,762	16,233,850	179.228	5.346	185,605.85	6.009
approach	21,213	4,606,887	144.029	6.493	149,558.67	6.969
overview	13,443	736,092	115.619	8.48	131,675.96	8.037
provide	25,933	18,103,719	155.29	4.808	123,562.07	5.479
list	18,178	5,125,404	132.882	6.116	118,710.44	6.617
review	16,982	5,025,451	128.343	6.046	109,257.10	6.543
more	29,657	40,350,816	160.234	3.845	103,764.21	4.556
immigration	10,441	618,256	101.871	8.367	100,593.36	7.778

Table 10: Top ten collocates for ‘comprehensive’ sorted by Log Likelihood

Searching the adjective ‘comprehensive’ in the concordancer and sorting results by Log Dice assigns ‘overview’, and ‘approach’ the highest scores and chooses them as the most exclusive collocates of ‘comprehensive’. Comparing the two lists, sorted by Log Dice and by Log Likelihood, demonstrates that the list sorted by Log Likelihood prioritises function words and assigns ‘approach’ a higher score than ‘overview’ and ‘plan’. The list sorted by Log Dice, however, eliminates the function words and assigns ‘overview’ and ‘plan’ respectively the highest scores; then comes ‘approach’ as the fifth on the list. Searching the three collocations ‘comprehensive approach’,

‘comprehensive overview’, ‘comprehensive plan’ on Google gives the following results respectively: 6,220,000; 5,500,000; 4,850,000. Google shows that ‘comprehensive approach’ is the commonest collocation. Thereby, ‘Comprehensive approach’ is both a frequent and exclusive collocation.

Consistent manner

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	log Dice
manner	9,580	1,026,771	97.708	9.177	103,163.69	7.89
basis	5,330	1,949,418	72.576	7.407	44,243.22	6.281
nationally	1,249	328,851	35.191	7.881	11,168.81	6.032
pattern	3,110	1,580,386	55.31	6.932	23,758.39	5.763
condom	589	81,764	24.215	8.804	6,019.34	5.674
theme	2,288	1,707,730	47.257	6.378	15,729.34	5.225
approach	5,458	4,606,887	72.873	6.2	36,253.24	5.188
message	3,374	2,822,166	57.303	6.213	22,447.90	5.147
internally	453	139,339	21.178	7.656	3,909.11	5.09
performer	647	337,170	25.222	6.896	4,904.98	5.064

Table 10: Top ten collocates for ‘consistent’ sorted by Log Dice

The first two noun collocates that the list exhibits are ‘manner’ and ‘basis’ using Log Dice scores. Log Likelihood prefers ‘basis’ over ‘manner’ and this preference is shared by T-Score. However MI chose “manner” to be one of the strongest collocates of the adjective “consistent” and therefore ranked it before “manner”. This shows again the similar approach shared by Log Dice and MI on the one hand, and by Log Likelihood and T-Score on the other.

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
manner	9,580	1,026,771	97.708	9.177	103,163.69	7.89
basis	5,330	1,949,418	72.576	7.407	44,243.22	6.281
and	33,420	524,893,821	136.559	1.982	44,164.48	1.06
approach	5,458	4,606,887	72.873	6.2	36,253.24	5.188
more	8,730	40,350,816	86.477	3.747	29,417.59	2.815
pattern	3,110	1,580,386	55.31	6.932	23,758.39	5.763
message	3,374	2,822,166	57.303	6.213	22,447.90	5.147
quality	3,823	5,164,318	60.484	5.522	21,836.72	4.519
most	5,255	16,233,850	68.883	4.328	21,634.19	3.38
way	4,953	15,194,862	66.899	4.338	20,451.82	3.389

Table 11: Top ten collocates for ‘consistent’ sorted by Log Dice

In English ‘consistent’ seems to have two close senses one being more common than the other. The first sense is ‘systematic’ or ‘coherent’. This is also apparent in French (*manière consistante, façon consistante* “consistent manner”). The second, however, which is less frequent, is ‘constant’ as in ‘consistent support’. There are two common senses found in French dictionaries for *consistant*. One of them is ‘filling’ like in *repas consistant* (hearty meal), and the other sense is ‘thick’ when speaking about soup or ‘heavy’ when speaking about pastry. There are, however, other shared senses for the adjective ‘consistent’ in the two languages. One of them comes from the verb in English ‘consist of’ (‘be composed of’ or ‘include’). The other shared meaning in the two languages is ‘entailing’, as in *stratégie consistent à* (‘strategy consistent with’). One thing to be noticed here is that in the last two shared meanings of the word “consistent” do not fit into the *noun+adjective* combination and still stand on their own because they need to be followed by a preposition (‘of” and ‘with’ in English, and ‘à’ or ‘de’ in French).

Excited anticipation

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
delirium	363	14,108	19.049	12.751	5,702.14	7.167
super	3,421	792,279	58.438	10.176	41,612.05	7.024
chatter	350	42,590	18.699	11.104	4,692.93	6.688
soooo	186	16,769	13.633	11.537	2,605.57	6.157
vibrationally	140	475	11.832	16.269	2,923.85	6.052
anticipation	377	122,858	19.392	9.683	4,310.97	6.011
sooooo	142	9,024	11.913	12.041	2,089.07	5.904
singlet	134	5,903	11.573	12.57	2,070.49	5.88
triplet	133	17,444	11.526	10.996	1,762.88	5.661
electron	433	234,923	20.766	8.947	4,510.41	5.546

Table 12: Top ten collocates for ‘excited’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
state	4,959	15,309,911	69.608	6.439	34,825.47	3.401
an	4,222	59,415,245	61.564	4.25	17,118.72	1.217
people	3,049	22,451,783	53.7	5.185	16,117.55	2.149
to	7,662	428,400,642	69.266	2.26	12,451.79	-0.771
child	1,840	13,420,276	41.727	5.199	9,729.35	2.16
very	1,828	13,821,514	41.548	5.147	9,537.51	2.108
crowd	879	788,163	29.548	8.223	8,280.39	5.071
fan	905	1,486,869	29.898	7.349	7,434.09	4.252
sooo	457	18,578	21.374	12.686	7,137.74	7.423
anticipation	377	122,858	19.392	9.683	4,310.97	6.011

Table 13: Top ten collocates for ‘excited’ sorted by Log Dice

The above two tables represent the collocates of ‘exited’ sorted by Log Likelihood and Log Dice, respectively. The noun ‘anticipation’ fulfils the collocation criterion because it appears in the top ten collocates sorted by Log Likelihood and Log Dice, proving that it is both a significant and exclusive collocate for the adjective ‘excited’. This is apparent from the high rank this noun enjoys in both lists. Although ‘state’ could be a shared collocate for ‘exited’ and its corresponding false friend *excité,* it only appears within the top ten collocates sorted by Log Likelihood. Therefore, ‘state’ does not fulfil the criterion.

Fastidious person

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
daintiness	4	208	1.999	17.313	88.091	5.783
self-disgust	3	217	1.732	16.837	64.07	5.363
prokaryote	6	4,315	2.449	13.523	100.513	4.921
deadhead	4	2,163	1.999	13.935	69.288	4.918
distaste	6	9,181	2.449	12.434	91.449	4.114
groomer	4	6,996	1.999	12.241	59.893	3.837
personage	5	9,754	2.235	12.084	73.777	3.78
dresser	12	27,428	3.463	11.855	173.301	3.731
eater	13	38,398	3.604	11.485	181.084	3.392
microorganism	17	57,259	4.121	11.296	232.37	3.228

Table 14: Top ten collocates for ‘fastidous’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
attention	110	2,110,081	10.464	8.786	1,126.02	0.771
A	341	329,105,715	16.362	3.133	922.057	-4.88
most	119	16,233,850	10.733	5.956	754.876	-2.057
taste	67	865,578	8.172	9.356	737.365	1.339
bacterium	34	319,877	5.824	9.814	395.199	1.79
nature	43	2,805,786	6.506	7.02	333.99	-0.994
care	46	6,966,696	6.661	5.805	280.806	-2.208
his	73	35,760,988	8.049	4.111	280.775	-3.902
and	201	524,893,821	9.806	1.697	204.334	-6.316
the	280	936,353,281	10.126	1.34	195.869	-6.673

Table 15: Top ten collocates for ‘fastidous’ sorted by Log Likelihood

None of the nouns appeared in the two lists sorted by Log Dice and Log Likelihood. The list in which the results were sorted by Log Likelihood demonstrated the inclination of Log Likelihood towards function and stop words. Down the list, however, the noun ‘attention’ makes a significant collocate with the adjective ‘fastidious’. By contrast, the list in which the results were sorted according to the scores of Log Dice exhibited prevailing technical words, but down the list, the nouns ‘dresser’ and ‘personage’ make good noun collocate candidates for the adjective ‘fastidious’. In this case, there is no noun collocate that fulfils any of the criteria set above. Therefore, we cannot make a choice. ‘Fastidious’/*fastidieux* is one of the strong

false friends between English and French that does not share any meaning. *Fastidieux* in French means 'hard', 'tough' and 'tedious'. 'Travail fastidieux' ('*fastideux* work') means hard work but not meticulous or fussy.

Formidable opponent

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
foe	954	80,636	30.879	12.085	14,104.96	7.901
opponent	1,750	436,142	41.804	10.525	22,105.36	6.881
adversary	330	65,024	18.156	10.864	4,314.28	6.553
obstacle	1,054	359,688	32.435	10.071	12,635.68	6.397
holmes	300	108,725	17.303	9.984	3,555.84	5.952
intellect	167	67,169	12.908	9.834	1,944.17	5.544
competitor	449	324,872	21.147	8.987	4,702.65	5.294
reputation	560	491,200	23.607	8.709	5,651.18	5.083
substitute	310	296,567	17.56	8.584	3,073.08	4.873
barrier	684	766,479	26.073	8.356	6,569.87	4.778

Table 16: Top ten collocates for 'formidable' sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
A	21,263	329,105,715	139.751	4.587	104,550.37	1.081
challenge	3,504	4,959,106	58.969	8.038	32,319.87	4.518
opponent	1,750	436,142	41.804	10.543	22,151.31	6.883
task	1,656	2,226,300	40.546	8.112	15,380.41	4.575
foe	954	80,636	30.879	12.103	14,129.82	7.908
most	2,203	16,233,850	46.006	5.657	13,056.38	2.148
force	1,694	5,192,640	40.819	6.923	12,956.66	3.404
obstacle	1,054	359,688	32.435	10.09	12,663.15	6.399
barrier	684	766,479	26.074	8.375	6,587.59	4.779
reputation	560	491,200	23.608	8.728	5,665.68	5.085

Table 17: Top ten collocates for 'excited' sorted by Log Likelihood

Both lists show that 'opponent' is a good match for a noun collocate candidate. The noun 'opponent' received the same ranking as well in the two different lists. The

collocation ‘formidable opponent’ is both a significant and exclusive candidate according to Log Likelihood and Log Dice.

The confusion between ‘formidable’ in English and ‘formidable’ in French is that ‘formidable’ in English has a negative connotation and it is a typical adjective that describes hardship, opponent, challenge, obstacle, or adversary. On the other hand, ‘formidable’ in French has completely the opposite connotation because it always conveys a positive view about something. The collocation ‘Une aventure formidable’ means a great adventure in the positive sense of the word but not an adventure that brings about huge difficulty or hardship.

Rude awakening

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
awakening	2,639	76,345	51.368	14.488	48,079.77	9.699
curveball	368	4,264	19.182	15.808	7,366.45	8.561
mechs	247	647	15.716	17.953	5,765.11	8.159
mechanicals	182	382	13.49	18.273	4,352.81	7.732
shock	746	521,565	27.284	9.893	8,762.49	5.476
awakenings	40	2,228	6.324	13.543	671.788	5.454
gesture	325	228,819	18.008	9.883	3,807.89	5.375
britannia	46	12,210	6.779	11.29	628.29	5.24
baguette	41	9,185	6.4	11.535	573.922	5.188
bwoy	30	231	5.477	16.398	626.086	5.139

Table18: Top ten collocates for ‘rude’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
awakening	2,639	76,345	51.368	14.488	48,079.77	9.699
A	5,488	329,105,715	67.4	3.471	17,401.39	-0.872
shock	746	521,565	27.284	9.893	8,762.49	5.476
curveball	368	4,264	19.182	15.808	7,366.45	8.561
comment	840	3,116,644	28.821	7.485	7,071.99	3.129
mechs	247	647	15.716	17.953	5,765.11	8.159
people	1,119	22,451,783	32.442	5.05	5,707.60	0.705
behavior	607	1,418,738	24.55	8.152	5,663.86	3.781
boy	594	1,780,121	24.262	7.793	5,248.08	3.428
word	571	5,082,778	23.575	6.223	3,810.94	1.872

Table 19: Top ten collocates for ‘rude’ sorted by Log Likelihood

‘Awakening’ is the first noun appearing in both lists sorted by Log Likelihood and Log Dice. ‘Awakening’ is thus the most exclusive and significant noun collocate for ‘rude’. The adjective “rude” in English is often used to describe an impolite person’s behaviour. The corresponding false friend in French, however, does not have the same sense. It is often used to refer to hard and rough tasks or weather but is not used to describe behaviour.

Grand prize

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
prix	54,236	74,398	232.869	13.745	983,923.51	10.724
rapids	50,471	90,521	224.636	13.358	872,034.07	10.598
jury	43,512	357,111	208.504	11.164	593,975.20	10.062
canyon	32,089	197,006	179.075	11.583	457,736.86	9.808
slam	21,718	160,816	147.312	11.313	300,754.48	9.290
opening	37,873	1,187,063	194.286	9.231	411,770.81	9.164
theft	17,894	223,166	133.679	10.561	228,021.52	8.933
finale	11,106	85,087	105.342	11.264	152,864.06	8.422
prise	17,810	743,563	133.158	8.818	182,928.51	8.406
forks	9,504	18,557	97.478	13.236	161,491.64	8.29

Table 20: Top ten collocates for ‘grand’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
prix	54,236	74,398	232.869	13.745	983,923.51	10.724
rapids	50,471	90,521	224.636	13.358	872,034.07	10.598
jury	43,512	357,111	208.504	11.164	593,975.20	10.062
the	270,702	936,353,281	424.779	2.445	533,310.13	3.242
canyon	32,089	197,006	179.075	11.583	457,736.86	9.808
opening	37,873	1,187,063	194.286	9.231	411,770.81	9.164
slam	21,718	160,816	147.312	11.313	300,754.48	9.290
theft	17,894	223,166	133.679	10.561	228,021.52	8.933
prise	17,810	743,563	133.158	8.818	182,928.51	8.406
A	94,811	329,105,715	251.19	2.44	172,514.33	3.234

Table 21**:** Top ten collocates for ‘rude’ sorted by Log Dice

The nouns ‘jury’ and ‘prix’, ‘opening’, ‘theft’, ‘slam’, occurred in the two lists. Therefore, they make good noun candidates for the adjective ‘grand’. All these nouns are both exclusive and significant collocates for the adjective ‘grand’. A ‘grand jury’

is a group of people who decide whether there is enough evidence to bring someone to the trial. The adjective “grand” in English means ‘magnificent’ but not big, which is the sense of the “grand “in French. Therefore, it is highly likely that we say ‘grand city’ to mean ‘magnificent city’ with impressive buildings but not to mean *grande ville* in French, which simply means a big city.

Inhabited Island

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
sparsely	239	18,663	15.457	13.111	3,871.50	7.434
densely	169	41,276	12.995	11.466	2,350.32	6.35
continuously	558	268,340	23.605	10.489	7,011.60	5.954
locality	274	123,288	16.542	10.585	3,476.48	5.904
dwelling	276	155,203	16.599	10.263	3,378.78	5.636
svalbard	62	15,241	7.871	11.457	861.194	5.601
continent	547	377,527	23.364	9.967	6,477.87	5.47
naga	43	18,009	6.553	10.688	551.376	4.981
atoll	40	20,396	6.319	10.404	497.153	4.801
island	1,769	2,120,857	41.986	9.171	19,080.61	4.754

Table 22: Top ten collocates for ‘inhabited’ sorted by Log Dice

The two lists show that ‘island’ is a common noun collocate whether the results have been sorted by Log Dice or Log Likelihood. This noun, therefore, fulfils the criterion set for choosing the best noun collocates.

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
By	9,004	80,783,447	93.657	6.267	63,844.81	1.868
area	2,522	12,023,675	49.873	7.179	20,336.94	2.777
island	1,769	2,120,857	41.986	9.171	19,080.61	4.754
world	1,650	14,871,139	40.09	6.26	11,166.30	1.859
continuously	558	268,340	23.605	10.489	7,011.60	5.954
continent	547	377,527	23.364	9.967	6,477.87	5.47
land	737	4,433,577	26.911	6.844	5,551.92	2.436
city	861	10,529,007	28.823	5.82	5,283.34	1.418
planet	513	911,221	22.591	8.604	5,105.79	4.163
village	547	1,978,311	23.265	7.578	4,669.71	3.16

Table 23: Top ten collocates for ‘inhabited’ sorted by Log Likelihood

The adjective ‘inhabited’ in English and its corresponding form in French ‘inhabité’ are opposite in terms of meaning. While ‘inhabited’ in English is used to describe a

populated place, the adjective *inhabité* refers to an 'uninhabited place'. The noun 'island' seems to be one of the most frequent, significant, exclusive and strong noun collocate for the adjective 'inhabited' according to the results given by T-Score, Log-likelihood, Log Dice, and MI respectively.

Jolly Lady

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
phonics	767	23,960	27.693	14.622	14,064.80	9.053
rancher	1,226	53,635	35.012	14.137	21,669.30	9.027
ranchers	245	2,626	15.652	16.166	5,026.92	8.272
roger	1,072	226,355	32.732	11.866	15,545.95	7.136
postman	111	12,013	10.534	12.83	1,753.90	6.685
hangman	85	5,824	9.218	13.489	1,421.15	6.577
holly	372	107,246	19.28	11.416	5,151.20	6.544
jumper	158	52,345	12.564	11.216	2,142.43	6.096
vijay	95	24,600	9.743	11.571	1,334.79	6.021
jumps	50	2,962	7.07	13.699	850.556	5.961

Table 24: Top ten collocates for 'jolly' sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
rancher	1,226	53,635	35.012	14.137	21,669.30	9.027
roger	1,072	226,355	32.732	11.866	15,545.95	7.136
phonics	767	23,960	27.693	14.622	14,064.80	9.053
holly	372	107,246	19.28	11.416	5,151.20	6.544
ranchers	245	2,626	15.652	16.166	5,026.92	8.272
A	2,419	329,105,715	40.694	2.534	4,676.38	-2.053
the	3,839	936,353,281	42.786	1.692	4,035.01	-2.896
green	486	3,314,817	21.854	6.852	3,663.33	2.254
Or	937	61,023,201	28.081	3.597	2,985.74	-0.991
david	350	2,147,460	18.562	7.005	2,709.67	2.401

Table 25: Top ten collocates for 'jolly' sorted by Log Likelihood

The three common nouns shared between the three lists are 'rancher', 'phonics', and 'roger'. All of these nouns when combined with the adjective 'jolly' are names for a specific brand, approach, or flag. The lists, therefore, fail to give any common noun collocate that does not produce a specific compound noun.

Rentable space

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
negocio	18	711	4.242	18.330	422.104	8.118
nicho	11	519	3.316	18.074	253.942	7.544
empresarial	9	813	2.999	17.137	195.975	7.050
impresión	4	97	2	19.034	97.729	6.440
conversiones	3	262	1.732	17.186	65.516	5.875
self-storage	3	2,163	1.731	14.14	52.82	4.781
sqm	12	22,550	3.463	12.758	188.36	4.036
mas	9	36,195	2.999	11.661	127.555	2.971
airbnb	3	18,759	1.731	11.024	39.858	2.284
sauna	9	60,059	2.998	10.93	118.443	2.262

Table 25: Top ten collocates for ‘rentable’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
space	275	4,919,989	16.56	9.507	3,132.66	0.872
area	189	12,023,675	13.68	7.677	1,661.68	-0.957
office	72	6,051,526	8.43	7.275	586.825	-1.359
Of	254	486,716,028	13.592	2.764	576.365	-5.869
negocio	18	711	4.242	18.33	422.104	8.118
net	41	1,490,521	6.385	8.484	401.683	-0.151
total	45	3,011,684	6.673	7.604	386.28	-1.03
nicho	11	519	3.316	18.074	253.942	7.544
pavilion	19	159,679	4.356	10.597	241.422	1.950
A	133	329,105,715	9.341	2.395	234.886	-6.238

Table 26: Top ten collocates for ‘rentable’ sorted by Log Likelihood

Apart from ‘square’ which is a measurement unit for space, there is no other noun found in both lists. Although ‘space’ might be a good noun collocate for ‘rentable’, the fact that it does not appear in both lists excludes it from the choice.

‘Rentable’ in English is always associated with space that can be rented. However ‘rentable’ in French has got a different denotation. It refers to anything lucrative. *Entreprise rentable* (‘*rentable* enterprise’) does not mean a “rentable company”. Rather, it refers to a company with a good source of income.

Secular humanism

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
humanism	2,069	27,126	45.481	13.153	33,780.16	8.548
humanist	1,821	42,858	42.664	12.309	27,533.54	8.244
stagnation	1,048	33,963	32.364	11.848	15,157.66	7.513
franciscan	541	44,367	23.243	10.508	6,808.55	6.482
jew	1,368	347,872	36.907	8.875	14,120.16	6.481
clergy	625	140,499	24.952	9.053	6,602.20	6.12
sermon	574	173,348	23.897	8.627	5,724.43	5.844
nationalism	428	106,392	20.645	8.908	4,434.55	5.751
priest	1,060	491,809	32.431	8.008	9,665.91	5.749
worldview	360	84,100	18.936	8.998	3,774.55	5.631

Table 27: Top ten collocates for ‘secular’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
society	7,032	5,018,251	83.356	7.387	58,366.46	5.477
A	21,667	329,105,715	128.482	2.975	54,023.89	1.108
state	7,330	15,309,911	84.118	5.837	45,264.81	3.957
humanism	2,069	27,126	45.481	13.153	33,780.16	8.548
humanist	1,821	42,858	42.664	12.309	27,533.54	8.244
The	25,113	936,353,281	109.014	1.679	26,064.64	-0.186
world	4,069	14,871,139	61.837	5.03	20,591.68	3.149
And	15,083	524,893,821	87.039	1.779	16,612.51	-0.087
stagnation	1,048	33,963	32.364	11.848	15,157.66	7.513
Jew	1,368	347,872	36.907	8.875	14,120.16	6.481

Table 28: Top ten collocates for ‘secular’ sorted by Log Likelihood

The two lists demonstrates three shared noun collocates in the two lists sorted by Log Likelihood and Log Dice: ‘humanism’, ‘humanist’, and ‘Jew’. However, ‘humanism’ comes first in both lists. Therefore, ‘secular humanism’ will be the chosen collocation.

The adjective *séculaire* in French is associated with events that took place centuries ago and has nothing to do with ‘secular’ in English. The word *séculaire* is related to the word *siècle* in French which means ‘century’. The adjectives ‘secular’ and *séculaire* do not fall under the same semantic field. For this reason, no shared collocate in both languages exists in this bilingual list.

Sympathetic ear

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
dystrophy	603	27,112	24.553	13.329	9,959.02	8.207
reflex	625	49,639	24.995	12.508	9,605.49	7.839
ganglion	268	21,988	16.367	12.461	4,098.89	7.152
nerve	1,367	286,920	36.956	11.106	18,368.15	7.099
innervation	169	3,548	12.999	14.427	3,051.10	6.999
portrayal	344	69,464	18.539	11.161	4,639.95	6.689
denervation	118	1,968	10.862	14.759	2,186.02	6.534
ophthalmia	96	446	9.797	16.603	2,040.26	6.29
Ear	1,400	568,950	37.383	10.152	16,960.61	6.235
neuron	384	160,851	19.578	10.109	4,618.85	5.971

Table 29**:** Top ten collocates for ‘sympathetic’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
A	7,774	329,105,715	80.102	3.45	24,419.60	-0.369
nerve	1,367	286,920	36.956	11.106	18,368.15	7.099
Ear	1,400	568,950	37.383	10.152	16,960.61	6.235
character	1,154	2,350,034	33.821	7.827	10,259.08	3.983
dystrophy	603	27,112	24.553	13.329	9,959.02	8.207
reflex	625	49,639	24.995	12.508	9,605.49	7.839
observer	491	351,727	22.124	9.335	5,380.42	5.361
portrayal	344	69,464	18.539	11.161	4,639.95	6.689
neuron	384	160,851	19.578	10.109	4,618.85	5.971
more	1,204	40,350,816	32.185	3.787	4,119.36	-0.033

Table 30: Top ten collocates for ‘sympathetic’ sorted by Log Likelihood

‘Sympathetic’ is an adjective used to describe someone who is compassionate and feels for others. In French, however, the adjective *sympathique* is used to describe somebody who is friendly and nice but not necessarily compassionate and understanding of others’ problems. Although ‘sympathetic’ falls under the same general semantic field, which is human character and feeling, it has got different usages to French *sympathique*. Therefore, ‘sympathetic character’ and *personnage sympathique* are not the translation of each other.

Ulterior motive

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
motive	5,755	165,349	75.86	16.398	126,068.72	10.092
motives	56	1,767	7.483	16.263	1,152.77	7.643
hispania	36	1,289	5.999	16.08	731.726	7.083
vistas	18	4,703	4.242	13.213	293.835	5.605
tytannia	7	102	2.645	17.377	155.134	4.932
not-so-secret	6	1,152	2.449	13.657	101.64	4.521
epicure	5	1,793	2.235	12.756	78.44	4.154
baetica	4	188	1.999	15.688	79.082	4.108
ululation	4	573	1.999	14.08	70.109	4.037
publicată	3	14	1.732	19.02	73.799	3.726

Table 31: Top ten collocates for ‘ulterior’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
motive	5,755	165,349	75.86	16.398	126,068.72	10.092
An	1,524	59,415,245	38.425	5.992	9,987.36	-0.25
purpose	274	2,773,151	16.485	7.937	2,479.57	1.691
have	737	147,205,635	24.962	3.635	2,423.14	-2.607
No	358	17,910,918	18.539	5.632	2,110.63	-0.611
with	472	119,656,805	19.506	3.291	1,330.84	-2.951
some	248	21,650,064	15.194	4.829	1,189.55	-1.414
motives	56	1,767	7.483	16.263	1,152.77	7.643
Any	215	16,049,912	14.221	5.055	1,095.53	-1.188
without	136	6,515,489	11.436	5.694	809.362	-0.549

Table 32: Top ten collocates for ‘ulterior’ sorted by Log Likelihood

The adjective ‘motive’ is one of the most frequent and exclusive and strongest collocates for the adjective ‘ulterior’ according to the scores given by Log Likelihood, Log Dice, and Mutual Information. The corresponding false friend *ultérieur* means ‘subsequent’ or ‘postponed’, but not ‘secret’ or ‘hidden’.

Petulant child

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
pipsqueak	11	669	3.316	16.504	229.901	6.522
wino	8	1,072	2.828	15.364	154.482	5.921
man-child	8	1,865	2.828	14.566	145.597	5.678
plutocrat	7	2,945	2.645	13.714	119.117	5.21
poetess	6	2,573	2.449	13.686	101.869	5.076
wavertons	3	15	1.732	20.109	78.278	4.911
childishly	4	1,237	1.999	14.158	70.528	4.867
pissant	3	244	1.732	16.085	60.936	4.813
jitteriness	3	262	1.732	15.982	60.506	4.806
pout	11	10,149	3.316	12.581	169.909	4.749

Table 33: Top ten collocates for ‘petulant’ sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
Child	362	13,420,276	18.901	7.252	2,961.92	-0.178
A	760	329,105,715	25.456	3.706	2,668.87	-3.724
And	507	524,893,821	18.394	2.449	950.061	-4.981
,	551	790,253,324	17.519	1.979	746.208	-5.451
Little	54	5,674,923	7.211	5.749	325.3	-1.682
His	90	35,760,988	8.82	3.83	312.803	-3.6
teenager	26	332,527	5.087	8.788	265.091	1.343
pipsqueak	11	669	3.316	16.504	229.901	6.522
Like	55	19,003,482	6.963	4.032	205.018	-3.398
outburst	14	32,992	3.74	11.228	190.002	3.661

Table 34: Top ten collocates for ‘petulant’ sorted by Log Likelihood

The only shared noun collocate in the two lists is the word ‘pipsqueak’. This term is an informal word. For this reason, it will not be chosen. Another reason for not choosing this word is that Algerian students won’t understand its meaning, and therefore the likelihood of not translating this collocation successfully will be high.

“Petulant” and *pétulant* are adjectives used to describe human behaviour in English and French respectively. They do not, however, convey the same sense. ‘Petulant’ is always used to refer to a bad-tempered person who behaves like an angry child. It is

for this reason that ‘petulant’ collocates often with “child” as shown by the Log Likelihood scores above. By contrast, “pétulant” in French refers to someone who is vivaceous and lively and does not have the same disapproving connotation that its English corresponding false friend does. Therefore, *princesse pétulante* cannot be rendered as ‘petulant princess’.

The automatic extraction of the Arabic collocations by Log Dice and Log Likelihood agreement

مزاج متعكر / *mizāj mutaʕakkir*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
جيد	457	405,381	21.367	11.069	6,149.10	5.191
خاص	376	846,281	19.367	9.725	4,351.80	3.856
حار	186	36,692	13.636	13.237	3,050.41	7.213
سوء	247	518,180	15.698	9.827	2,885.83	3.953
شخصي	195	92,237	13.96	11.976	2,856.83	6.047
متعكر	83	430	9.11	18.488	1,980.11	9.147
بارد	124	57,604	11.132	12.002	1,819.06	6.034
حاد	116	89,626	10.765	11.268	1,583.40	5.337
ب	751	254,980,043	22.521	2.488	1,451.35	-3.373
معتدل	86	27,653	9.272	12.532	1,324.20	6.459

Table 35: The top ten collocates for the noun مزاج / *mizāj* sorted by Log Likelihood

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
متعكر	83	430	9.11	18.488	1,980.11	9.147
حار	186	36,692	13.636	13.237	3,050.41	7.213
احتفالي	27	5,214	5.195	13.268	442.955	6.528
معتدل	86	27,653	9.272	12.532	1,324.20	6.459
شخصي	195	92,237	13.96	11.976	2,856.83	6.047
بارد	124	57,604	11.132	12.002	1,819.06	6.034
متوتر	16	5,657	3.999	12.396	243.063	5.708
منحرف	21	10,519	4.581	11.893	304.399	5.53
ازرق	40	24,193	6.322	11.621	564.891	5.52
متشائم	13	5,155	3.604	12.23	194.493	5.483

Table 36: The top ten collocates for the noun مزاج / *mizāj* sorted by Log Dice

The adjective متعكر / *mutaʕakkir* occurs in both lists and is ranked first in the list sorted by Log Dice. However, it is the sixth in the list in which the results are sorted by Log likelihood. The adjective شخصي / *šaxṣī* comes before it. A Google search shows that the collocation مزاج شخصي / *mizāj šaxṣī* has 10,900 results (11/01/2019), while مزاج متعكر / *mizāj mutaʕakkir* has 16,200 results (11/01/2019). This latter is, therefore, both stronger and more frequent in Arabic.

توجهات سياسية / *tawajjuhāt siyāsiyya*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	log likelihood	log Dice
يسارية	145	17,563	12.039	12.182	2,161.57	7.198
فكرية	542	115,818	23.272	11.363	7,477.41	7.087
اصلاحية	125	27,041	11.176	11.345	1,717.84	6.613
ايديولوجية	174	53,879	13.183	10.828	2,266.82	6.375
مستقبلية	145	55,832	12.033	10.514	1,825.48	6.072
سياسية	1,756	871,722	41.867	10.147	21,414.35	6.02
علمانية	133	52,395	11.524	10.481	1,668.21	6.019
حزبية	133	61,118	11.523	10.259	1,627.23	5.843
معادية	92	44,482	9.583	10.185	1,115.99	5.668
اسلامية	487	312,145	22.042	9.778	5,646.08	5.609

Table 37: Top ten collocates for the noun توجهات Tawajuhāt sorted by Log Dice

collocates	Co-occurrence count	Candidate count	T-Score	MI	log likelihood	logDice
سياسية	1,756	871,722	41.867	10.147	21,414.35	6.02
ذات	1,041	2,100,387	32.148	8.124	9,725.39	4.011
مختلفة	809	890,191	28.387	8.999	8,523.64	4.872
جديدة	905	1,813,758	29.976	8.134	8,458.30	4.019
فكرية	542	115,818	23.272	11.363	7,477.41	7.087
اسلامية	487	312,145	22.042	9.778	5,646.08	5.609
معينة	508	555,817	22.495	9.007	5,347.43	4.866
دينية	342	254,690	18.468	9.562	3,859.02	5.378
هناك	504	6,574,867	21.929	5.431	2,827.14	1.325
عامة	282	753,153	16.713	7.72	2,462.17	3.588

Table 38: Top ten collocates for the noun توجهات Tawajuhāt sorted by Log Likelihood

The two lists demonstrate that the adjective سياسية / *siyāsiyya* collocates exclusively and frequently with the noun توجهات / *tawajjuhāt*. This adjective is ranked second in the list sorted by Log Dice scores and third in that sorted by Log Likelihood. The

collocation توجهات سياسية / *tawajjuhāt siyāsiyya* can be rendered as ‘political orientations’.

عجز مالي / *ʕajz mālī*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
الموازنة	10,713	149,557	103.457	11.124	145,016.49	9.718
الميزانية	4,046	135,079	63.54	9.866	47,440.71	8.365
الميزان	1,605	126,695	39.96	8.624	16,018.65	7.061
مالي	1,831	198,914	42.641	8.164	17,103.40	7.008
سد	1,460	224,349	38.021	7.663	12,623.57	6.605
خفض	1,181	197,234	34.181	7.543	10,013.75	6.381
ميزانية	819	179,564	28.416	7.151	6,498.98	5.909
ميزان	720	150,025	26.653	7.224	5,786.09	5.821
تغطية	910	284,465	29.863	6.639	6,580.00	5.756
صام	523	52,053	22.796	8.29	4,974.54	5.747

Table 39: Top ten collocates for the noun عجز ʕajz sorted by Log Dice

collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
عن	45,899	43,570,665	207.714	5.036	239,451.77	5.1
الموازنة	10,713	149,557	103.457	11.124	145,016.49	9.718
ه	37,377	203,672,611	159.527	2.515	72,483.86	2.586
الميزانية	4,046	135,079	63.54	9.866	47,440.71	8.365
هم	14,437	57,089,763	104.907	2.978	35,010.29	3.044
و	44,007	582,826,773	120.628	1.234	27,378.28	1.306
اذا	6,546	11,556,033	76.324	4.141	25,383.19	4.181
ان	18,267	151,659,145	99.149	1.908	22,221.92	1.978
سبب	3,312	2,652,057	56.071	5.282	17,843.04	5.216
مالي	1,831	198,914	42.641	8.164	17,103.40	7.008

Table 40: Top ten collocates for the noun عجز ʕajz sorted by Log Likelihood

The adjective which appears in the two lists and collocates with عجز / *ʕajz* is مالي / *mālī*. This adjective comes fourth in the list in which the results are sorted by Log Dice and last where the results are sorted by Log Likelihood. As usual, lists sorted by Log Likelihood tend to assign function words high scores and thus they are ranked

first. It is for this reason that Log Dice can outperform Log Likelihood in the ranking of adjectives. The collocation عجز مالي / *ʕajz mālī* is typically rendered in English as 'financial deficit'.

عقبة كؤود / *ʕaqaba ka'ūd*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
امام	7,485	3,947,544	85.797	6.911	57,311.03	5.91
عن	11,032	43,570,665	98.498	4.006	41,429.04	3.048
ميت	2,021	90,253	44.923	10.473	25,380.31	8.224
يشكل	2,480	518,613	49.635	8.245	23,464.25	6.966
تجاوز	1,793	495,639	42.159	7.843	15,955.79	6.55
تشكل	1,756	578,237	41.687	7.59	15,013.34	6.341
تقف	1,484	323,156	38.39	8.187	13,909.32	6.742
حديث	2,093	2,035,647	45.048	6.028	13,403.98	4.984
كبيرة	1,909	1,782,522	43.049	6.087	12,376.40	5.03
كؤود	596	947	24.412	15.285	11,935.70	7.208

Table 41: Top ten collocates for the noun عقبة ʕaqaba sorted by Log Likelihood

collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
تخطي	2,225	48,868	47.153	11.496	31,154.79	8.662
ميت	2,021	90,253	44.923	10.473	25,380.31	8.224
جبر	949	70,977	30.769	9.729	10,923.48	7.265
كؤود	596	947	24.412	15.285	11,935.70	7.208
يشكل	2,480	518,613	49.635	8.245	23,464.25	6.966
تخطى	563	23,199	23.712	10.589	7,155.61	6.901
رئيسية	845	112,429	29.007	8.898	8,748.77	6.829
تقف	1,484	323,156	38.39	8.187	13,909.32	6.742
اجتياز	511	45,440	22.573	9.479	5,702.41	6.567

Table 42: Top ten collocates for the noun عقبة ʕaqaba sorted by Log Dice

كؤود / *ka'ūd* is one of the adjectives we first think of in relation to the noun عقبة / *ʕaqaba*. This adjective occurs in both lists. It is ranked third in the list sorted by Log Dice, and last in the list sorted by Log Likelihood. The noun عقبة / *ʕaqaba* together with the adjective كؤود / *ka'ūd* makes a very strong and exclusive collocation. This

can be demonstrated through the highest score assigned to this adjective by MI. Out of 947 which is the total number of occurrence of the adjective كؤود / *ka'ūd* the latter occurred 596 times together with the noun عقبة / *ʕaqaba*. For this reason the collocation عقبة كؤود / *ʕaqaba ka'ūd* is considered the strongest collocation in the two lists.

احترافية عالية / *ʻiḥtirafiyya ʻāliya*

collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
تجربة	1,628	591,028	40.306	9.92	19,256.24	6.439
عروضا	124	25,052	11.129	10.766	1,604.13	6.382
كاميرا	187	51,405	13.664	10.321	2,304.29	6.351
عالية	1,168	477,340	34.136	9.749	13,514.97	6.255
عقودا	80	21,717	8.937	10.34	987.439	5.852
مهنية	177	79,161	13.287	9.619	2,008.62	5.817
عروض	292	152,889	17.062	9.392	3,223.05	5.76
عقلية	210	134,573	14.464	9.1	2,232.43	5.442
طريقة	1,771	1,460,382	41.984	8.737	18,054.62	5.289
ناجحة	110	94,704	10.462	8.674	1,104.03	4.928

Table 43: Top ten collocates for the noun احترافية ʻḥtirafiyya sorted by Log Dice

collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
تجربة	1,628	591,028	40.306	9.92	19,256.24	6.439
طريقة	1,771	1,460,382	41.984	8.737	18,054.62	5.289
عالية	1,168	477,340	34.136	9.749	13,514.97	6.255
ب	4,693	254,980,043	57.927	2.695	10,328.16	-0.729
و	5,891	582,826,773	55.172	1.83	7,353.09	-1.594
اكثر	1,035	6,505,049	31.596	5.806	6,343.43	2.377
بطولة	370	505,058	19.16	8.009	3,377.46	4.519
عروض	292	152,889	17.062	9.392	3,223.05	5.76
كبيرة	387	1,782,522	19.414	6.255	2,598.41	2.811
كل	541	7,390,478	22.356	4.686	2,486.70	1.257

Table 44: Top ten collocates for the noun احترافية ʻḥtirafiyya sorted by Log Likelihood

The adjective عالٍ / *ʕālin* collocates often with the noun احترافية / *ʻiḥtirafiyya*. The adjective is ranked fourth in the list in which the results are sorted by Log Dice, and third in the list sorted by Log Likelihood. The adjective is both frequent and exclusive.

The collocation احترافية عالية / *ʿiḥtirafiyya ʕāliya* can be rendered as 'high professionalism' or 'high degree of professionalism'.

مهارة فائقة / *mahāra fā'iqa*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
ب	17,366	254,980,043	112.704	2.788	40,387.66	1.157
فائقة	2,085	59,450	45.648	11.796	30,055.65	8.915
عالية	2,505	477,340	49.955	9.056	26,538.25	7.197
و	20,473	582,826,773	102.925	1.833	25,617.60	0.202
ه	9,206	203,672,611	75.019	2.196	14,308.20	0.566
تنمية	1,179	479,675	34.198	7.961	10,684.30	6.103
اكتساب	690	97,855	26.231	9.482	7,702.76	6.973
القراءة	766	520,700	27.491	7.221	6,154.71	5.38
ارنا	398	16,748	19.941	11.235	5,414.68	7.044
هم	2,958	57,089,763	44.038	2.393	5,097.11	0.761

Table 45**:** Top ten collocates for the noun مهارة mahāra sorted by Log Likelihood

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
فائقة	2,085	59,450	45.648	11.796	30,055.65	8.915
عالية	2,505	477,340	49.955	9.056	26,538.25	7.197
ارنا	398	16,748	19.941	11.235	5,414.68	7.044
اكتساب	690	97,855	26.231	9.482	7,702.76	6.973
اتقان	269	41,756	16.376	9.351	2,952.88	6.153
تنمية	1,179	479,675	34.198	7.961	10,684.30	6.103
الاستماع	463	140,459	21.453	8.385	4,462.98	6.091
تتطلب	507	192,988	22.432	8.057	4,657.70	5.916
صقل	191	36,058	13.794	9.069	2,021.62	5.727
تكتسب	176	38,907	13.237	8.841	1,807.24	5.575

Table 46: Top ten collocates for the noun مهارة mahāra sorted by Log Dice

The word فائق / *fā'iq* is one of the adjectives often associated with the word مهارة *mahāra* in Arabic. The adjective فائق / *fā'iq* comes first in the list in which the results are sorted by Log Dice, and second in the list in which the results are sorted by the Log Likelihood. This adjective is assigned the highest scores by MI, Log Likelihood, and Log Dice. Thus the adjective فائق / *fā'iq* can be considered a strong, frequent, and exclusive collocate of مهارة / *mahāra*.

حجة دامغة / *ḥujja dāmiġa*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
ذي	59,990	556,851	244.85	11.614	866,216.18	11.187
ذو	14,865	413,709	121.805	10.03	178,317.85	9.443
ذى	3,864	42,017	62.137	11.386	53,684.57	8.592
بالغة	2,978	115,369	54.498	9.553	33,598.96	7.925
1432	2,013	57,944	44.822	9.981	23,917.47	7.583
اقامة	7,968	1,162,597	88.816	7.639	68,810.25	7.493
1430	1,624	36,746	40.267	10.328	20,090.27	7.365
مقارعة	1,388	21,939	37.235	10.846	18,192.26	7.206
دامغة	1,232	10,735	35.089	11.705	17,681.21	7.088

Table 47**:** Top ten collocates for the noun حجة / *ḥujja* sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
ذي	59,990	556,851	244.85	11.614	866,216.18	11.187
ذو	14,865	413,709	121.805	10.03	178,317.85	9.443
اقامة	7,968	1,162,597	88.816	7.639	68,810.25	7.493
الامام	7,641	2,079,648	86.595	6.74	56,490.16	6.725
بالغة	2,978	115,369	54.498	9.553	33,598.96	7.925
هذه	10,063	24,271,839	92	3.592	31,968.27	3.747
على	16,711	83,389,002	107.104	2.543	31,934.25	2.71
1432	2,013	57,944	44.822	9.981	23,917.47	7.583
مقارعة	1,388	21,939	37.235	10.846	18,192.26	7.206
دامغة	1,232	10,735	35.089	11.705	17,681.21	7.088

Table 48: Top ten collocates for the noun حجة ḥujja sorted by Log Likelihood

The two lists demonstrate that the two adjectives بالغ / *bāliġ* and دامغ / *dāmiġ* both appear in the lists within the first top ten collocates. The tables demonstrate that the adjectiveبالغ / *bāliġ* is more frequent in Arabic (73,900 results on Google: 16/01/2019) than دامغ / *dāmiġ* (20,500 results on Google: 16/01/2019). However, the adjective دامغ / *dāmiġ* is stronger and more exclusive according to the scores given by MI and Log Dice. In this case, the adjective which is ranked first by both lists is the one chosen.

جرح نازف / *jurḥ nāzif*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
نازف	1,437	5,588	37.905	14.038	25,514.64	8.47
عمدي	848	6,251	29.117	13.115	13,849.13	7.702
غائر	763	2,885	27.62	14.078	13,592.95	7.586
ائمة	1,615	198,639	40.111	9.055	17,083.32	7.343
التئام	594	9,571	24.366	11.987	8,724.03	7.153
مفسر	646	22,108	25.403	10.9	8,493.25	7.147
يلتئم	420	8,417	20.487	11.672	5,979.44	6.665
العميق	657	105,681	25.569	8.668	6,591.62	6.53
علماء	1,818	642,085	42.407	7.533	15,401.13	6.275
ينزف	340	18,331	18.423	10.245	4,156.76	6.258

Table 49**:** Top ten collocates for the noun جرح / *jurḥ* sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
و	58,719	582,826,773	205.563	2.72	146,624.77	1.722
على	12,552	83,389,002	100.66	3.3	35,911.48	2.3
نازف	1,437	5,588	37.905	14.038	25,514.64	8.47
ائمة	1,615	198,639	40.111	9.055	17,083.32	7.343
علم	2,286	1,698,188	47.269	6.46	15,998.11	5.358
علماء	1,818	642,085	42.407	7.533	15,401.13	6.275
عمدي	848	6,251	29.117	13.115	13,849.13	7.702
غائر	763	2,885	27.62	14.078	13,592.95	7.586
الذي	3,533	22,854,008	53.563	3.338	10,064.85	2.332
ي	8,619	166,193,283	65.481	1.762	9,206.01	0.763
التئام	594	9,571	24.366	11.987	8,724.03	7.153

Table 50: Top ten collocates for the noun جرح / *jurḥ* sorted by Log Likelihood

The two lists share three adjectives within the top ten collocates غائر / *ġā'ir*, نازف / *nāzif* and عمدي / *ʕamdī*. The rank in both lists will be the factor based on which the best adjective will be chosen. The first adjective نازف / *nāzif* appears first in the list in which the results are sorted by Log Dice and third in the list sorted by Log Likelihood scores. The second adjective غائر / *ġā'ir* is ranked third in the Log Dice list and eighth in Log Likelihood list. The third adjective عمدي / *ʕamdī* comes second in the list in which the results are sorted by Log Dice and seventh in the list in which the results are sorted by Log Likelihood. The adjective نازف / *nāzif* is ranked before the other

two adjectives in both lists. The collocation جرح النازف / *jurḥ nāzif* is a both frequent and exclusive expression and an equivalent collocation in English can be found (‘bleeding wound’).

أعذار واهية / *ʿaʕḏār wāhiya*

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
واهية	828	13,501	28.773	13.864	14,330.54	9.151
تماس	1,159	44,689	34.038	12.622	18,033.65	8.91
اختلاق	761	22,330	27.582	13.016	12,253.84	8.784
مبيحة	358	1,545	18.92	15.782	7,210.60	8.357
يختلق	283	4,725	16.821	13.83	4,879.62	7.895
تختلق	190	3,246	13.783	13.797	3,266.55	7.376
نلتمس	199	9,994	14.103	12.241	2,984.34	7.204
مخففة	189	8,048	13.745	12.479	2,897.36	7.195
يختلقون	152	1,995	12.328	14.177	2,696.02	7.103
يلتمس	208	17,828	14.417	11.47	2,895.32	7.032

Table 51: Top ten collocates for the noun أعذار / *ʿaʕḏār* sorted by Log Dice

Collocates	Co-occurrence count	Candidate count	T-Score	MI	Log Likelihood	Log Dice
واهية	828	13,501	28.773	13.864	14,330.54	9.151
التماس	1,159	44,689	34.038	12.622	18,033.65	8.91
اختلاق	761	22,330	27.582	13.016	12,253.84	8.784
مبيحة	358	1,545	18.92	15.782	7,210.60	8.357
يختلق	283	4,725	16.821	13.83	4,879.62	7.895
تختلق	190	3,246	13.783	13.797	3,266.55	7.376
نلتمس	199	9,994	14.103	12.241	2,984.34	7.204
مخففة	189	8,048	13.745	12.479	2,897.36	7.195
يختلقون	152	1,995	12.328	14.177	2,696.02	7.103
يلتمس	208	17,828	14.417	11.47	2,895.32	7.032

Table 52: Top ten collocates for the noun اعذار / *ʿaʕḏār* sorted by Log Likelihood

The adjective واهية’ / *wāhiya* is ranked first in both lists of top ten collocates. Therefore, it make a perfect choice to be the adjective collocate which is associated with the noun ‘اعذار / *ʿaʕḏār* to give the collocation أعذار واهية / *ʿaʕḏār wāhiya* , which can be rendered into English as ‘flimsy excuses’.

Appendix E: Applying Dickins' 2020 compositionality model to the collocations used in this thesis

N	Constituents' types / Collocations	Type 1	Type 2	Type 3	Type 4
01	Heinous crime				++
02	Sensible person				++
03	Actual Cost				++
04	Concurrent enrolment		+		+
05	Candid camera		+		+
06	Comprehensive approach				++
07	Consistent manner				++
08	Eventual winner				++
09	Excited anticipation				++
10	Fastidious person				++
11	Formidable opponent				++
12	Rude awakening	++			
13	Grand prize		+		+
14	Inhabited Island				++
15	Jolly woman				++
16	Rentable space				++
17	Secular humanism		+		+
18	Sympathetic ear				++
19	Ulterior motive				++
20	Petulant child				++
21	مزاج متعكر				++
22	خداع بصري		+		+
23	توجهات سياسية				++
24	عجز مالي				++
25	عقبة كؤود		+		+
26	احترافية عالية				++
27	مهارة فائقة				++
28	حجة دامغة				++
29	جرح نازف				++
30	أعذار واهية				++

Table: 1 Applying Dickins' 2020 compositionality model to the collocations used in this research

When applied, Dickins'2020 model of compositionality on the extracted collocations used in this research showed that most of the collocations were fully free compositional. This means that both constituents were of type 4 (section 2.3.2). Only

7 collocations, however, were not fully free compositional having at least one of the constituents with an independent sense in one context as explained below.

Concurrent enrollment

This refers to a programme where students are enrolled in two schools at the same time, also called ‘dual enrollment’. Dual enrollment is a programme which allows high school students to enroll in college classes while they are still in school, often at a local community college. In other cases, students may take classes to earn both high school and college credits in the same class. While ‘enrolment’ in this expression has an independent sense in unlimited contexts, the adjective ‘concurrent’ here has an independent sense in just this one limited context. For more about 'concurrent enrollment', see: https://www.edmit.me/blog/how-to-save-on-college-tuition-with-concurrent-enrollment

Candid Camera

"Candid camera in American English (noun): a small, handy camera, esp. one having a fast lens for informal pictures" (Collins English Dictionary). While the noun ‘camera’ here has an independent sense in unlimited contexts (it is a type 4 constituent), the adjective ‘candid’ has a different sense to the one that is usually used. ‘Candid’, in this case, is a type 2 constituent (section 2.3.2).

Rude awakening

“If you have a rude awakening, you are suddenly made aware of an unpleasant fact” (Collins English Dictionary). In this expression, none of the constituents has an independent sense, so both of them are type1 constituents.

Grand prize

"The highest prize awarded in a lottery or competition, especially one for products in some particular line at an exhibition. *Origin* Mid-18th century. After French *grand prix*" (Collins English Dictionary). Unlike the word ‘prize’, which is a type 4 constituent having an independent sense in unlimited contexts, the adjective ‘grand’ has an independent sense just in this particular context and therefore it is a type 2 constituent.

Secular humanism

The "humanism that systematically opposes the introduction of religious ideas or standards into the functions of the state, esp. into public education" (Collins English Dictionary). Also, *"secular humanism promotes the belief that, while religion is to be respected, we are fast growing out of it."* (Lexico.com). While 'humanism' is a type 4 constituent retaining its basic sense in potentially unlimited context, the word 'secular' has this independent sense just in this context, i.e. it is a type 2 constituent.

خداع بصري / *xidāʕ baṣarī*

Meaning optical illusion (Almaany dictionary). While the word خداع/ xidāʕ has its basic sense which is found in potentially unlimited contexts, the adjective بصري / *baṣarī* has an independent sense different from the one used on standard contexts, i.e. it is a type 2 constituent.

عقبة كؤود / *ʕaqaba ka'ūd*

Meaning unsurmountable obstacle (Almaany dictionary). While the noun عقبة / *ʕaqaba* has a sense which is found in potentially unlimited contexts meaning, i.e. it is a type 4 constituent, the adjective كؤود / *ka'ūd* has an independent sense only in this context, i.e. it is a type 2 constituent.

Appendix F: Normality tests

	Kolmogorov-Smirnov	Shapiro-Wilk	Skewness	Kurtosis
Total scores for Translation	.87/ sig .09	.98/ sig .90	.29-/sig.25	-.73/sig .51
Scores for English Collocations	.81/ sig .20	.97/sig .05	.26/sig.25	-.74/sig .51
Scores for Arabic Collocations	.13/sig .00	.95/sig.00	.15 /sig.25	-.99/sig .51
Scores for Males	.15/sig.20	.97/sig.84	.22/sig.51	-.56/sig-.99
Scores for Females	.81/sig.20	.98/sig.20	.23/sig.30	-.75/sig 0.57
Age	.32/sig.00	.53/sig.00	4.04 /sig.25	19.97/sig.51
pilot	.18/sig.13	.94/sig.23	.67/sig.51	1.05/sig.99
The High Arab Institute	.19/sig.20	.95/sig.60	-.65/sig.64	.11/sig.23
University of Algiers	.10/sig.20	.97/sig.32	.19/sig.35	-.94/sig.73
University of Oran	.13/sig.20	.97/sig.74	.48/sig.55	-.12/sig1.06

Table 01: Tests of normality related to the raised sub-questions (1-4)

Appendix G: Distribution of translations for each collocation across the four groups

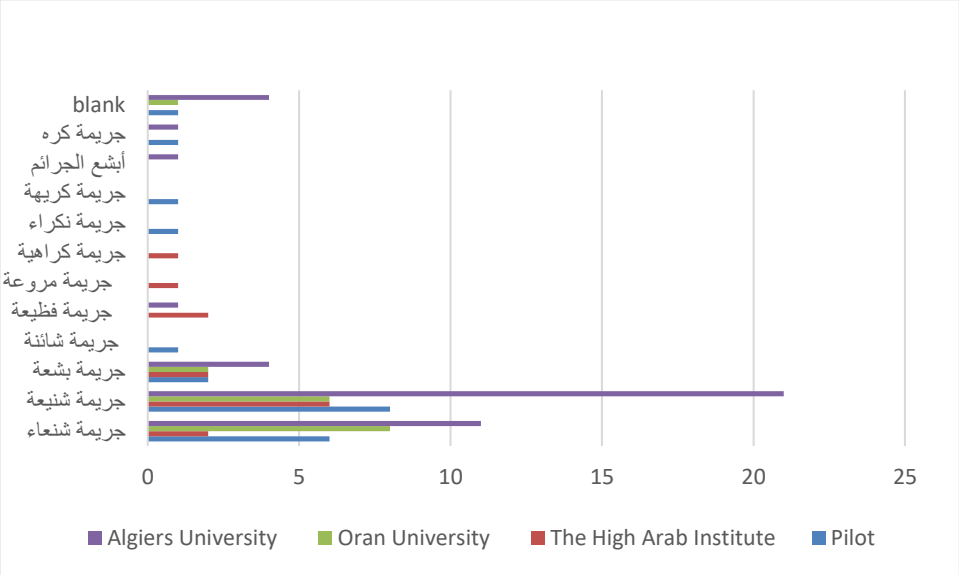

Figure 01: Distribution of translations of "heinous crime" across the four groups

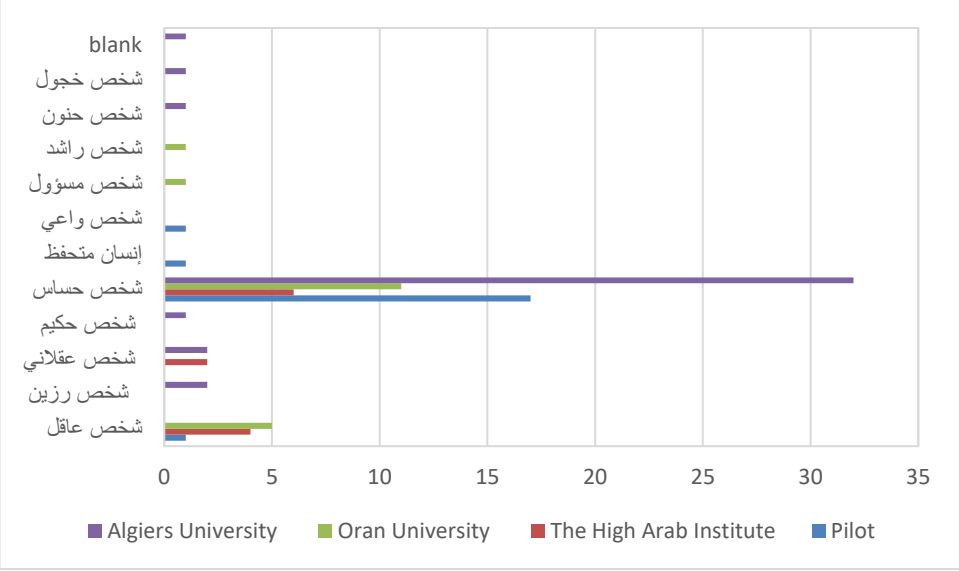

Figure 2: Distribution of translations of "sensible person" across the four groups

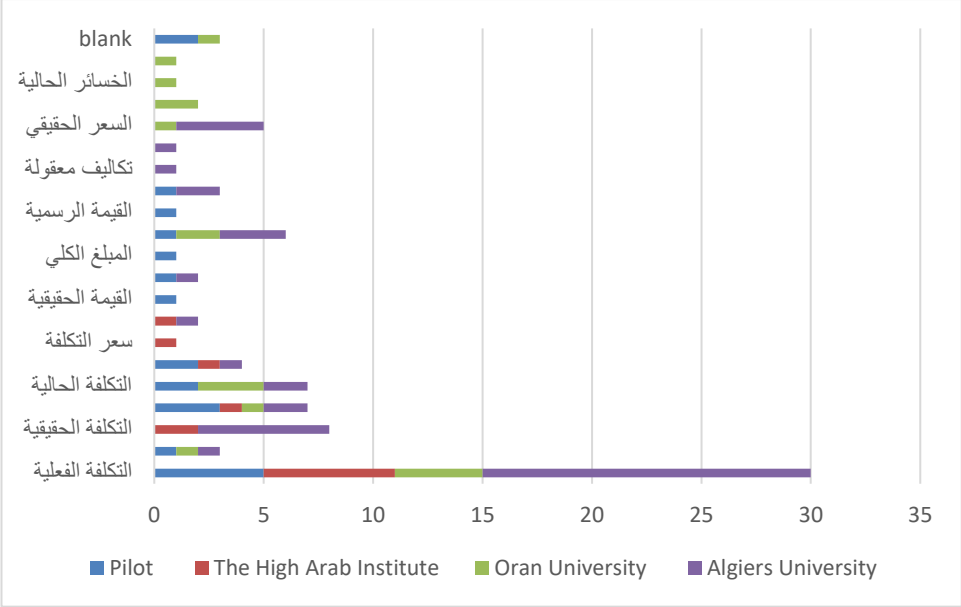

Figure 3: Distribution of translations of "actual cost" across the four groups

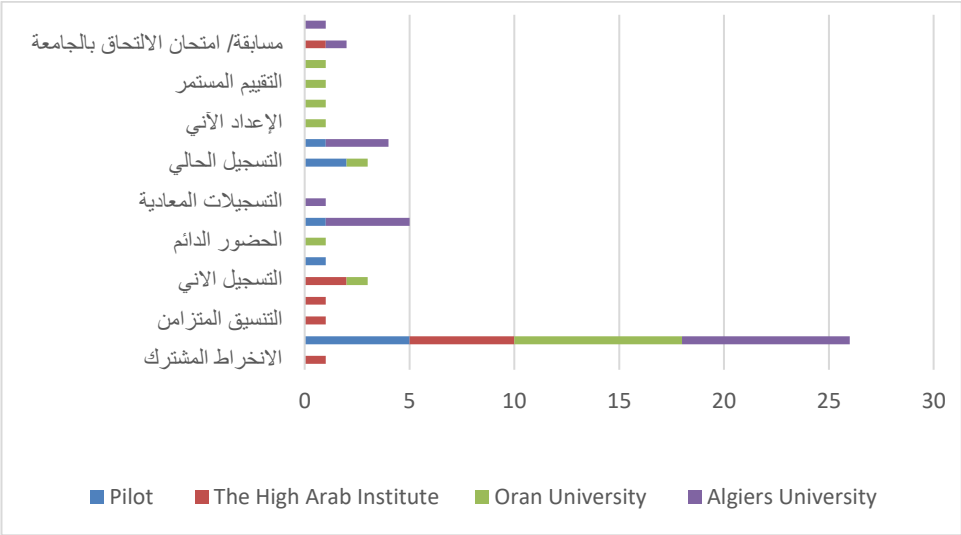

Figure 4: Distribution of translations of "concurrent enrolment" across the four groups

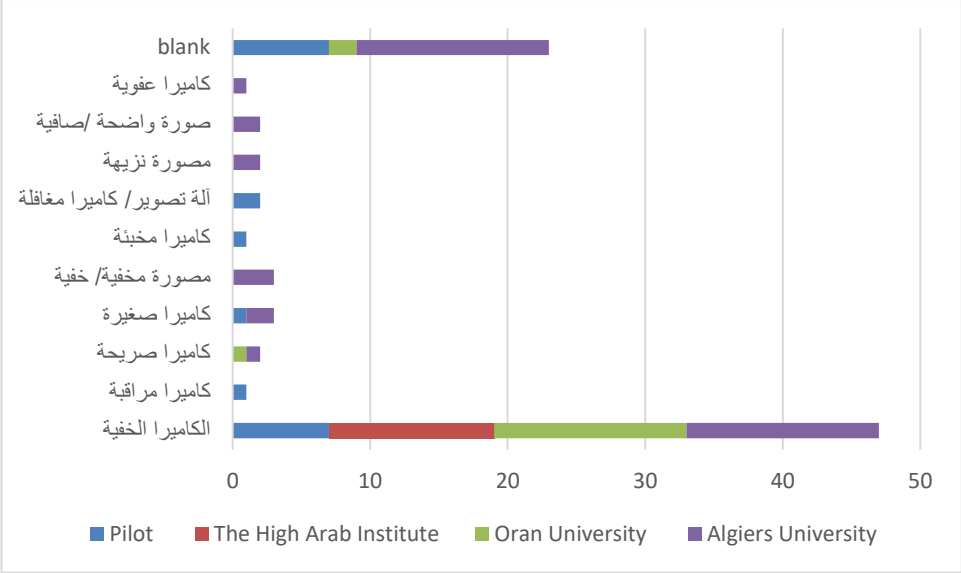

Figure 5: Distribution of translations of "candid camera" across the four groups

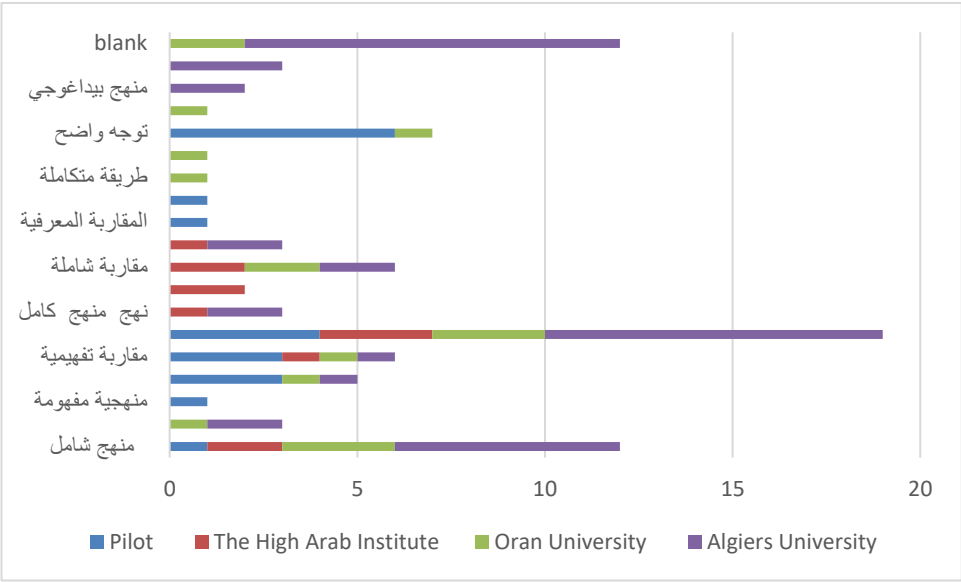

Figure 6: Distribution of translations of "comprehensive approach" across the four groups

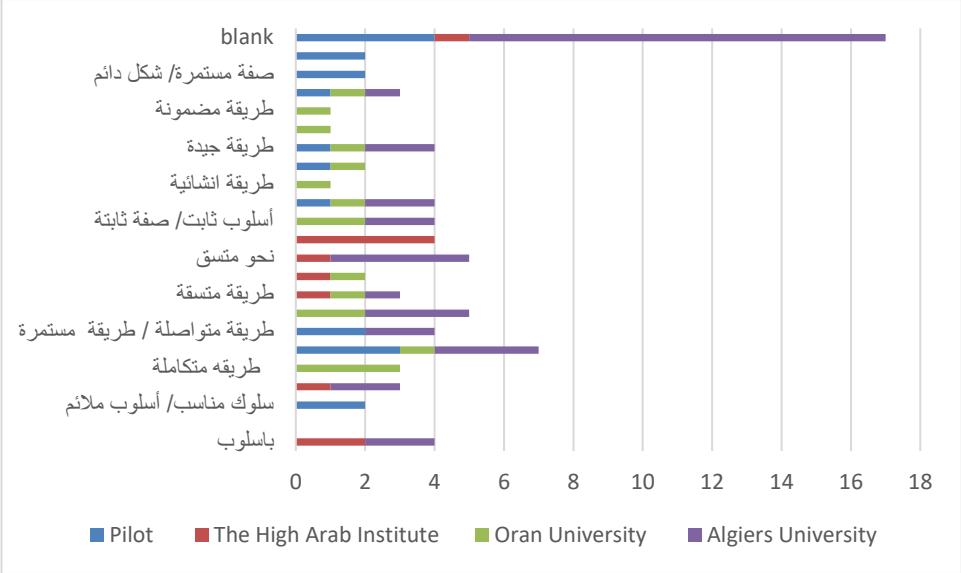

Figure 7: Distribution of Translations for “consistent manner” across the four groups

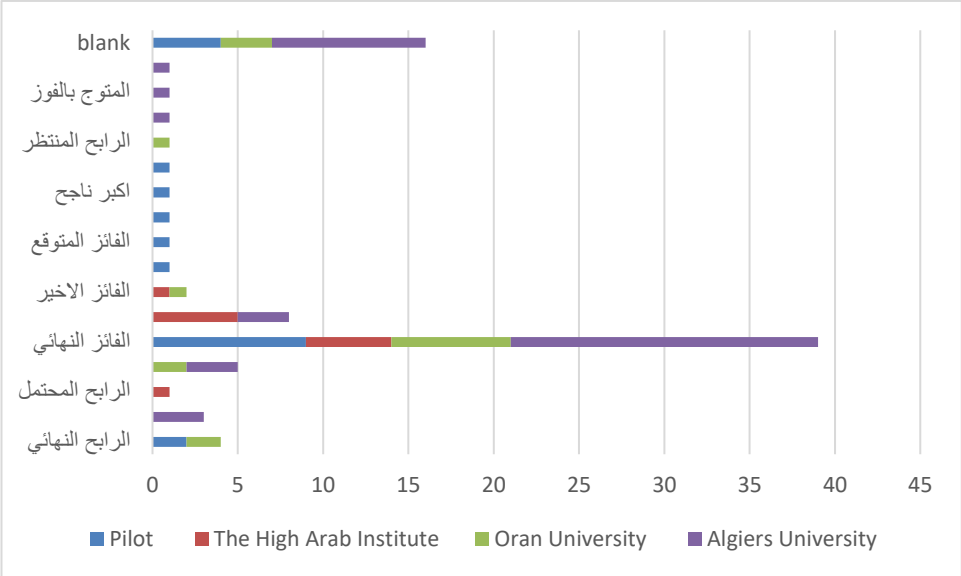

Figure 8: Distribution of translations of “eventual winner” across the four groups

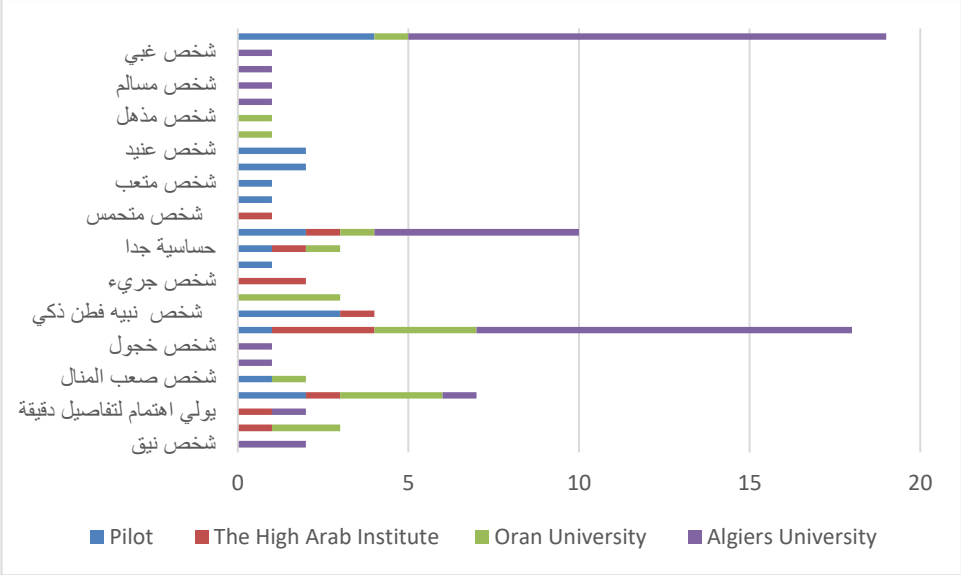

Figure 09: Distribution of translations of “fastidious person” across the four groups

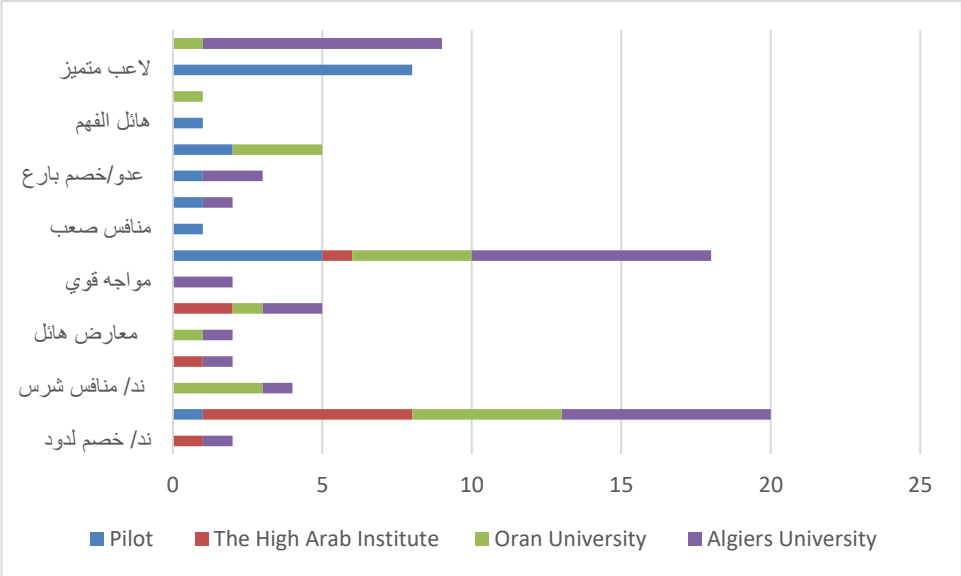

Figure 10: Distribution of translations of “formidable person” across the four groups

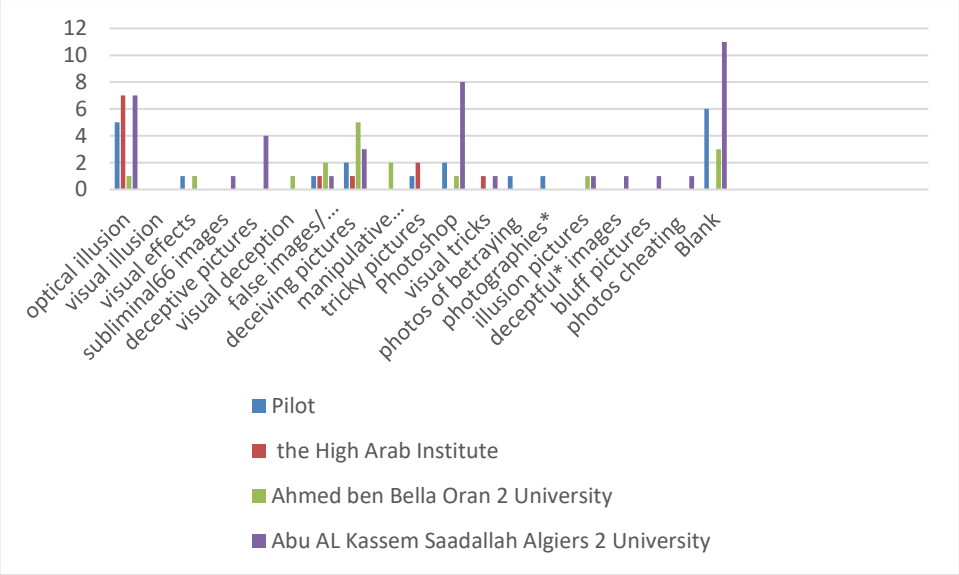

Figure 11: Distribution of translations of “rude awakening” across the four groups

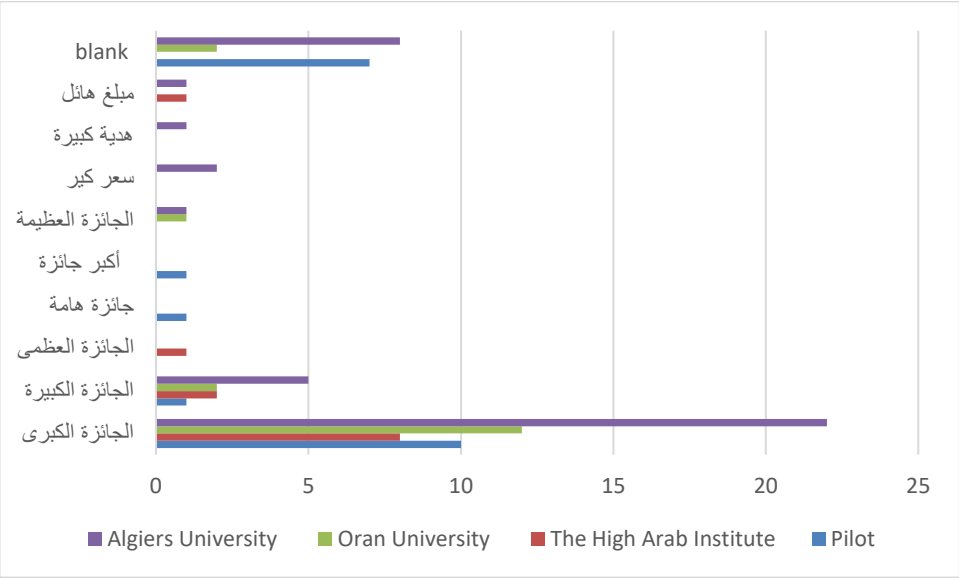

Figure 12: Distribution of translations of “grand prize” across the four groups

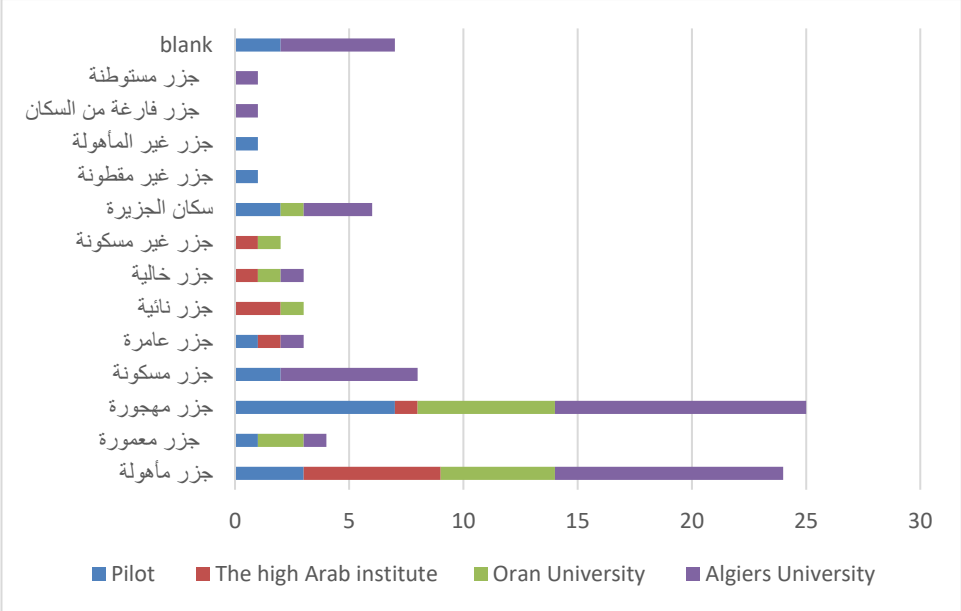

Figure 13: Distribution of translations of “inhabited islands” across the four groups

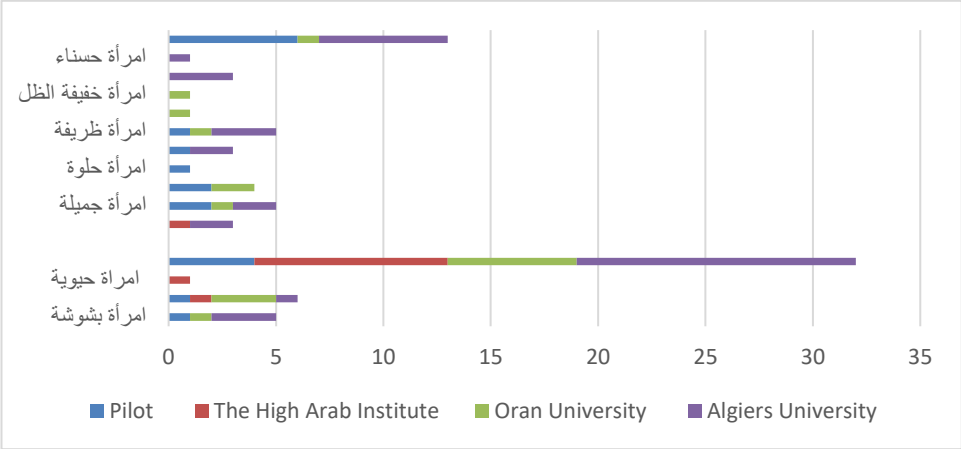

Figure 14: Distribution of translations for “jolly woman” across the four groups

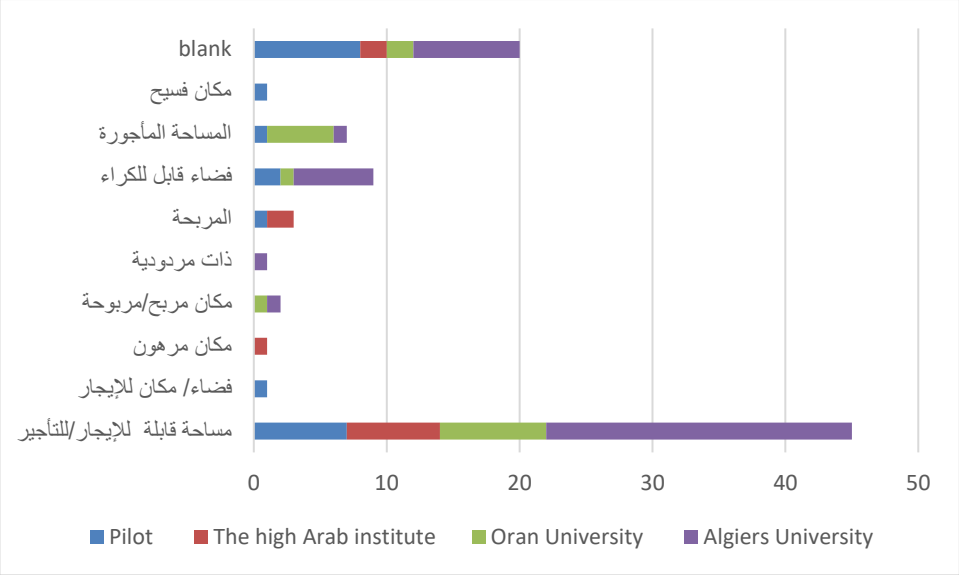

Figure 15: Distribution of translations of "rentable space" across the four groups

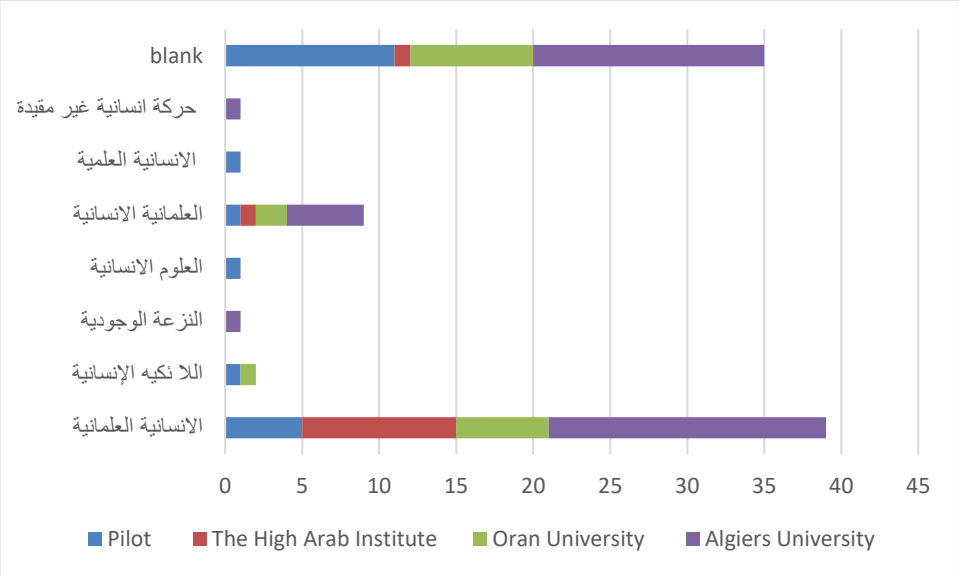

Figure 16: Distribution of translations of "secular humanism" across the four groups

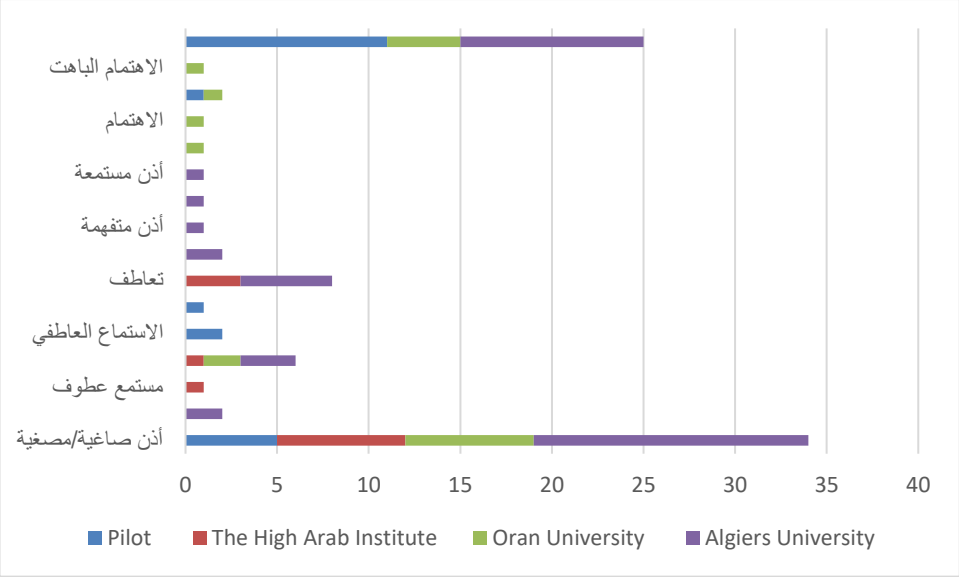

Figure 17: Distribution of translations of “sympathetic ear” across the four groups

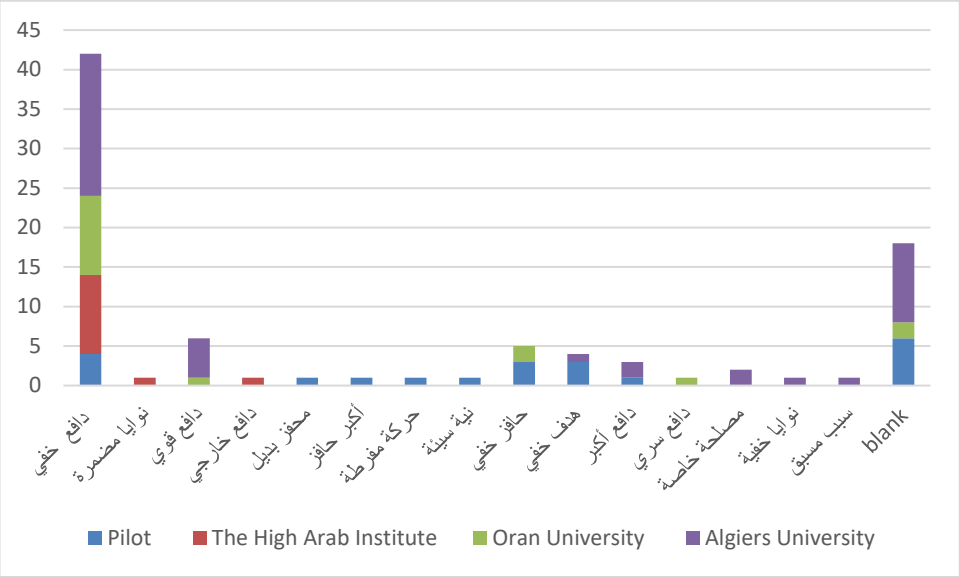

Figure 18: Distribution of translations of “ulterior motive” across the four groups

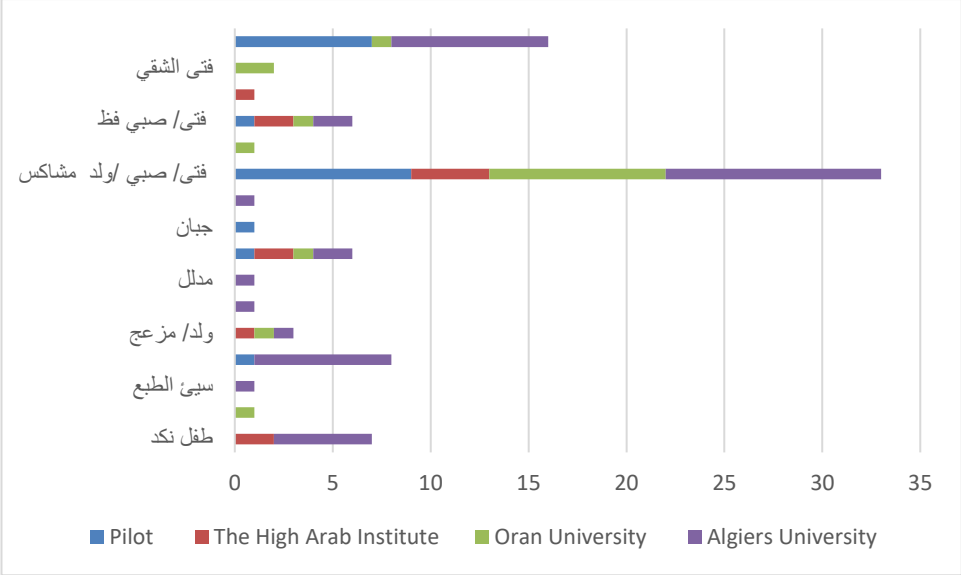

Figure 19: Distribution of translations of "petulant boy" across the four groups

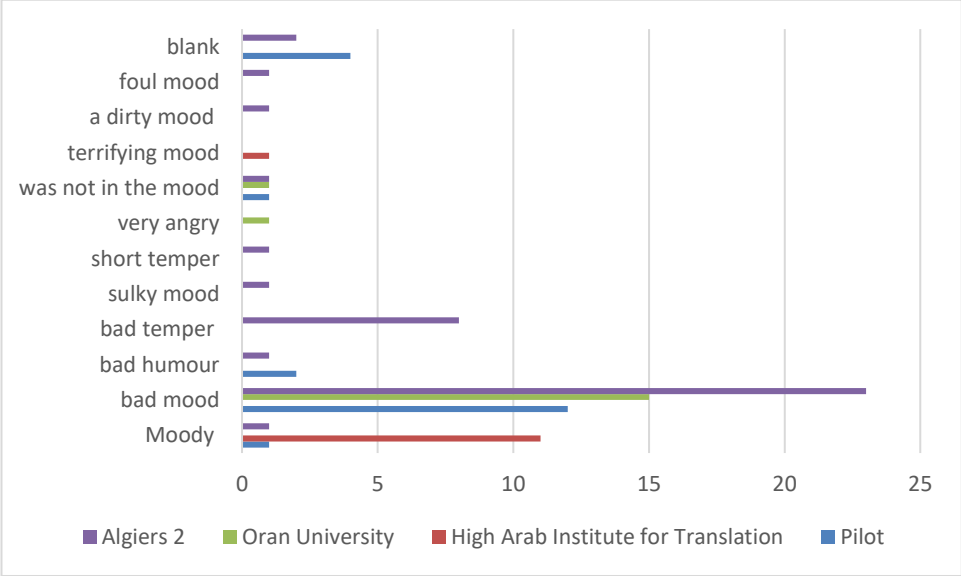

Figure 20 Distribution of translations of مزاج متعكر / *mizāj mutaʕakkir* across the four groups

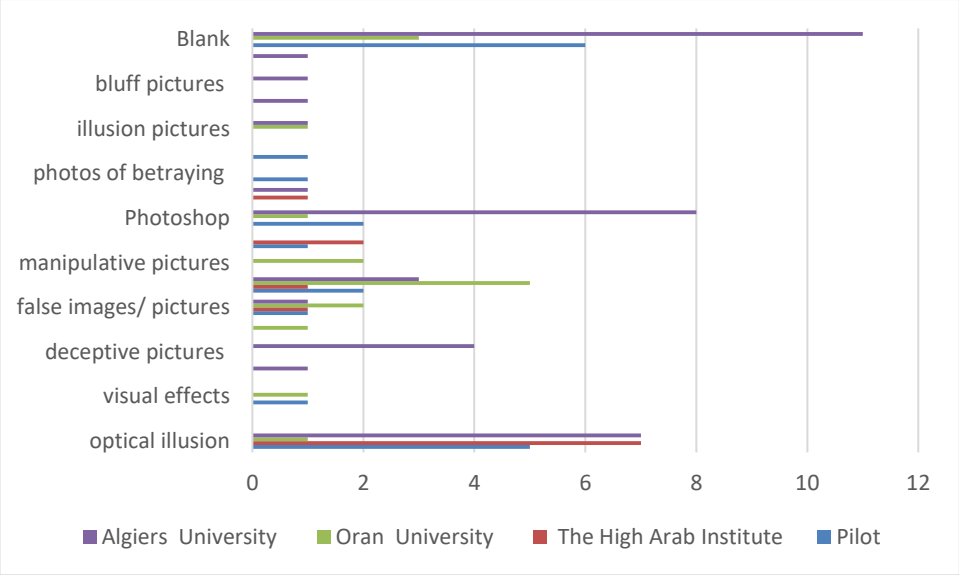

Figure 21: Distribution of translations of خداع بصري / *xidāʕ baṣarī* across the four groups

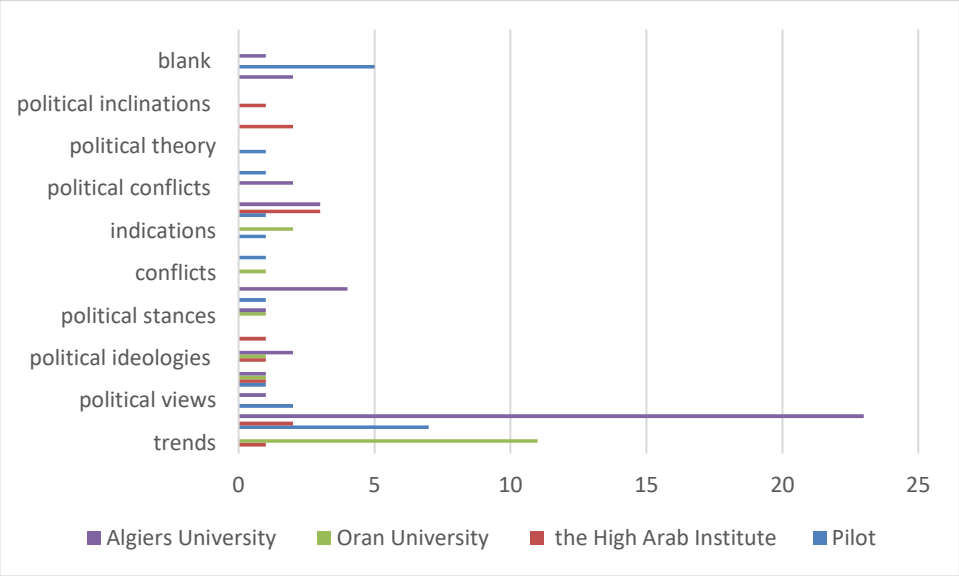

Figure 22: Distribution of translations of توجهات سياسية / *tawajjuhāt siyāsiyya* across the four groups

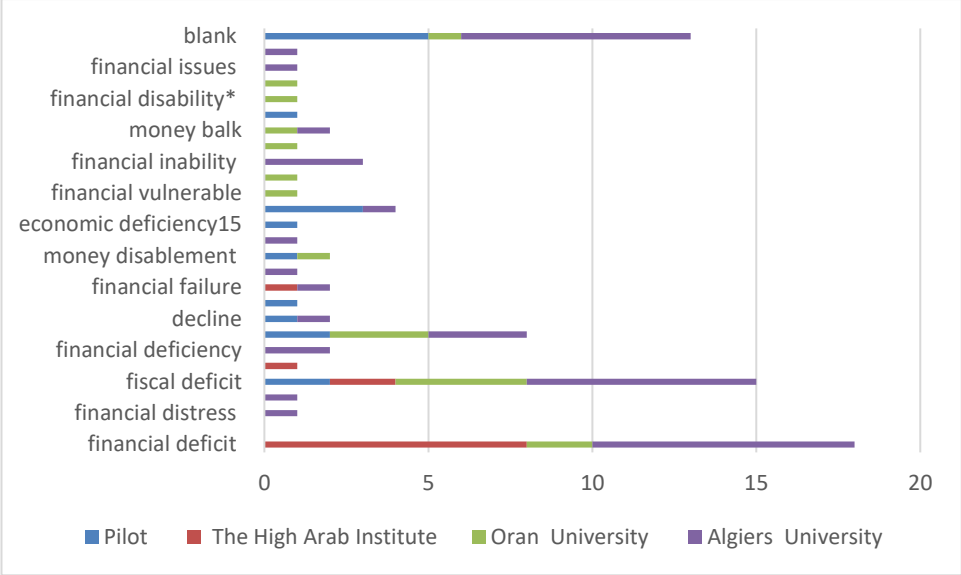

Figure 23: Distribution of translations of عجز مالي/ *ʕajz mālī* across the four groups

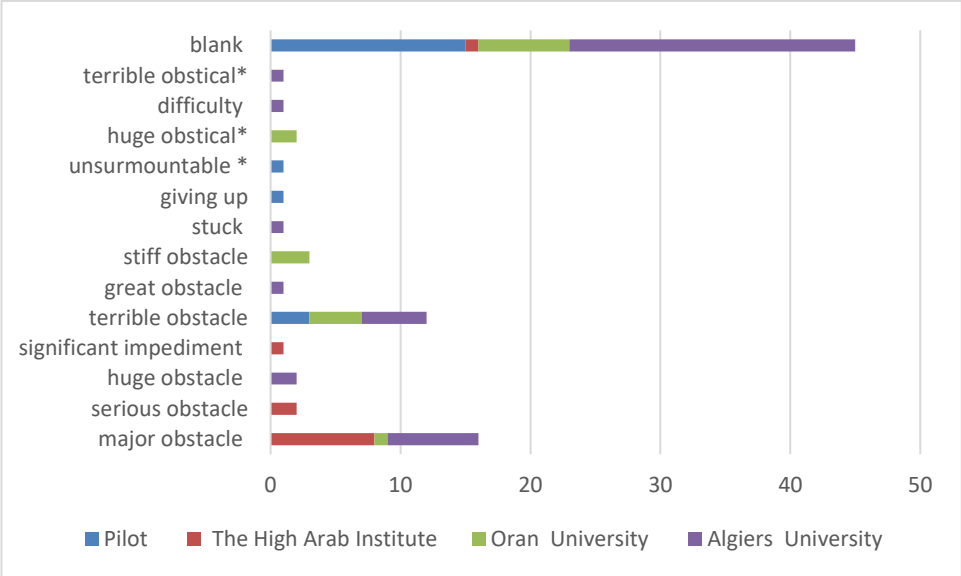

Figure 24: Distribution of translations of عقبة كؤود / *ʕaqaba ka'ūd* across the four groups

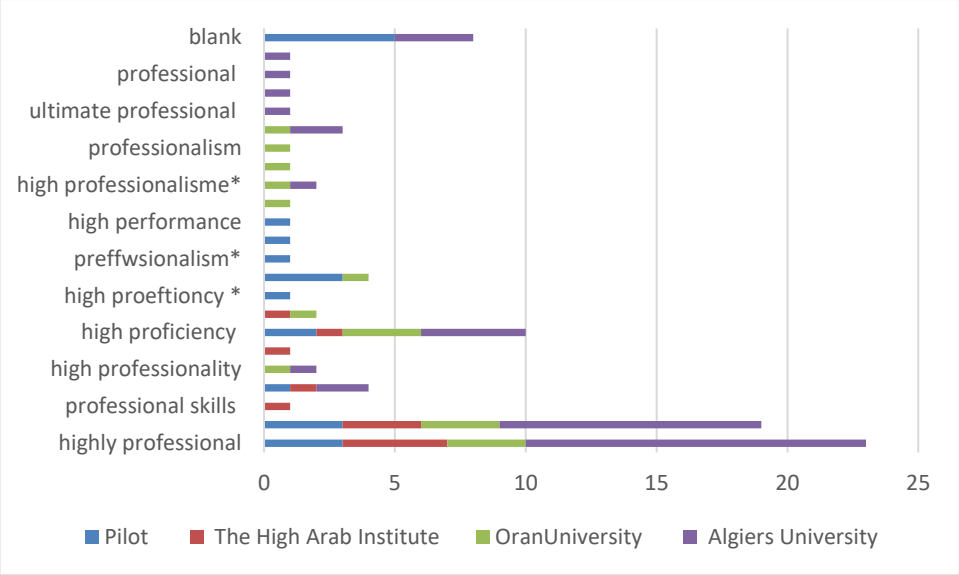

Figure 25: Distribution of translations of احترافية عالية /*'iḥtirafiyya 'āliya* across the four groups

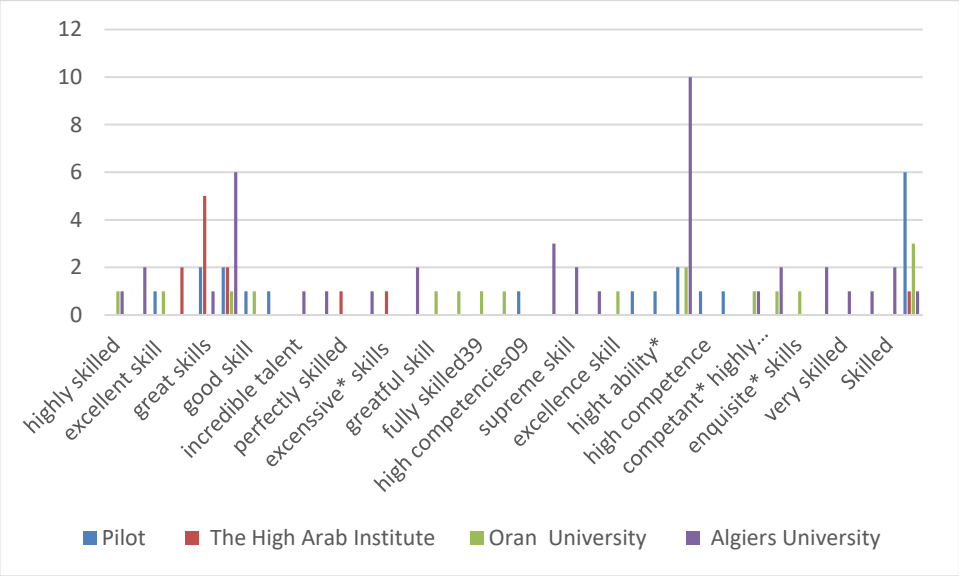

Figure 26: Distribution of translations of مهارة فائقة / *mahāra fā'iqa* across the four groups

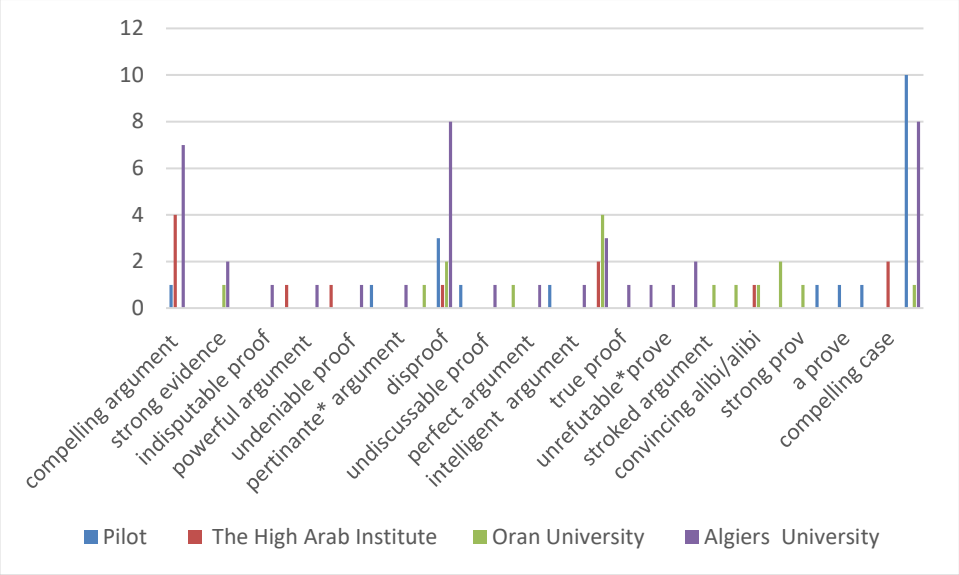

Figure 27: Distribution of translations of حجة دامغة / *ḥujja dāmiġa* across the four groups

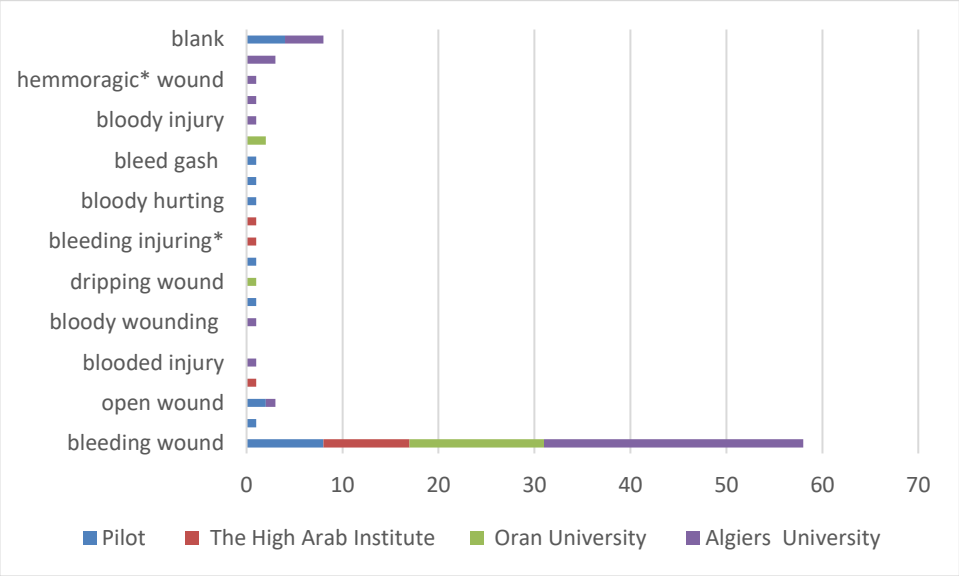

Figure 28: Distribution of translations of جرح نازف / *jurḥ nāzif* across the four groups

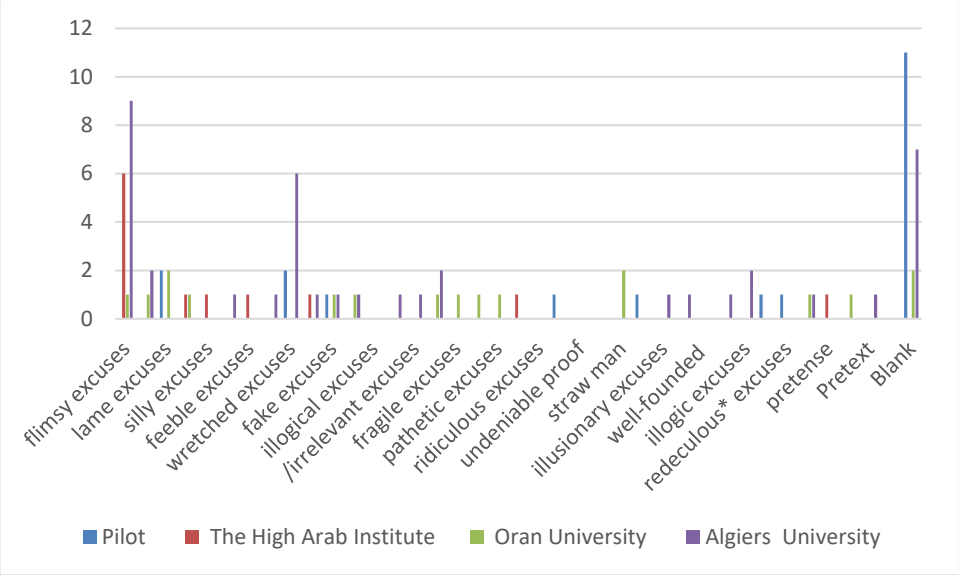

Figure 29: Distribution of translations of أعذارًا واهية / *aʕḏār wāhiya* across the four groups