



Mohamed Kheider University of Biskra  
Faculty of Letters and Languages  
Department of Foreign Languages

## MASTER THESIS

Letters and Foreign Languages  
English Language  
Sciences of the language

---

Submitted and Defended by:  
**HAMIDANE Zineb**

On: Monday, 28 June 2021

**at Mohamed Kheider University of Biskra**

Title:

**Exploring the Use of a Web Corpus for Learning Collocations:**  
The Case of First Year Master Students of English at Biskra University

---

### Board of Examiners:

Dr. Ahmed Chaouki <b>HOADJLI</b>	Biskra	Supervisor
Dr. Tarek <b>ASSASSI</b>	Biskra	Examiner
Ms. Kenza <b>MERGHMI</b>	Biskra	Chairperson

**Academic Year: 2020-2021**

**Declaration**

I, Zineb **HAMIDANE**, do hereby declare that the present work is the product of my own efforts. It is my original work that has not been submitted nor published beforehand by any individual, university, or institution for any degree. All the previously published works used in this research are rightfully acknowledged through citations and references. This research was conducted and completed at Mohamed Kheider University, Biskra, Algeria.

Author's name: Zineb HAMIDANE

Signature:

Date:

**Dedication**

To my parents, thank you for your love, care, and support. I love you beyond words. May

God protect you.

To my siblings, those of you who helped and supported me through this journey, thank

you, I love you.

To my soul mates Dina, Maissa, Nour, Marwa, you have been ever so kind, nurturing, caring, and loving. I simply could not imagine my life without you all. You have made this moment possible, and to that, I am incredibly grateful. You have had a major impact on my life

and I will cherish you all as long as I live.

To my best friends Ala, Wiem, Rania, Ikram, you have showed me nothing but kindness. You have always been there whenever I needed help and reassurance. Thank you, sincerely, I am forever grateful to have you in my life.

To everyone that has ever wished me good omen

In loving memory of my aunts, uncle, and grandparents, I love and miss you very much.

May Allah grant you paradise.

### Acknowledgements

In the name of Allah, the most merciful and forgiving, peace be upon his prophet and messenger Mohammed. Praise be to Allah, Lord of all creation, it was his mercy and power that allowed me to muster the courage and patience to complete this work.

I would like to express sincere gratitude for every individual that helped in the completion of this work.

First and foremost, profound gratitude goes to the supervisor of this work, Dr. Ahmed Chaouki **HOADJLI**. His knowledge, guidance, patience, and continuous assistance in the research process were outstanding and are greatly appreciated.

Second, I would like to express the deepest gratitude to the Board of Examiners: Dr. Tarek **ASSASSI** and Ms. Kenza **MERGHMI** for their efforts in providing thoughtful comments for the betterment of this work, and for taking the time to correct and evaluate it.

I would also like to give thanks to the master one students who participated in this study, for taking the time to sit for the test and to answer the questionnaire. I am deeply appreciative of their help and efforts.

Sincere thanks and deepest gratitude goes to the teachers who humbly agreed to sit for the interviews despite their busy schedules. I am incredibly humbled and appreciative of the assistance provided by: Dr. **ASSASSI**, Ms. **MERGHMI**, Mr. **SENOUCI**, Dr. **MEDDOUR**.

Lastly, I am indebted to Dr. Ramdane Mehiri, Prof. Saliha Chelli, Mr. Abdelhak Chenini, Dr. Tarek Assassi, Ms. Kenza Merghmi, Dr. Mostafa Meddour, and Mr. Zakaria Senouci for constantly responding to my emails and for their kindness, assistance, and guidance throughout the research process.

### **Abstract**

Vocabulary knowledge is a pillar in language learning. This knowledge encompasses more profound understandings. Hence, an essential element of vocabulary knowledge is collocational knowledge, specifically, how certain word combinations naturally occur in language.

Nonetheless, it was observed that the majority of master one students at the Section of English at Biskra University struggle in terms of diverse vocabulary and using accurate word collocations.

It was also anticipated that the traditional methods and materials used might contribute to the inadequate learning of collocations. Moving from that point, this study attempted to assess the level of collocational competence of 37 students who were conveniently sampled. Moreover, the present study aimed at exploring the possibility of integrating data-driven learning (DDL) as a teaching and learning approach and the potential obstacles that might prevent its application in our context. In order to reach those objectives, a Mixed-methods approach following the Case Study design was applied. Thus, a test and a questionnaire were designed and developed for the students and interviews were conducted with 4 teachers of the Section. Ultimately, the findings revealed that the majority of the students in the sample had below average collocational competence. Furthermore, several factors might have led to such results including the inadequate teaching of collocations and lack of diversity in teaching methods. The results also unveiled that the participants were enthusiastic about the integration of the DDL approach. Lastly, the obstacles that may face the implementation of this approach vary and include management and administrative issues. Nevertheless, the application of the DDL in our context was dubbed as possible and recommended under the conditions of well planning and thoughtful organisation.

*Keywords:* Collocational knowledge, corpus linguistics, corpus of contemporary American English, data-driven learning, lexical units

**List of Abbreviations and Acronyms**

**AD:** After Digitalisation

**BC:** Before Computers

**BNC:** British National Corpus

**CALL:** Computer-Assisted Language Learning

**CBA:** Competency-Based Approach

**CCEC:** Collins COBUILD English Course

**CL:** Corpus Linguistics

**CLT:** Communicative Language Teaching

**COCA:** Corpus of Contemporary American English

**DDL:** Data-Driven Learning

**EFL:** English as a Foreign Language

**ELT:** English Language Teaching

**ESL:** English as a Second Language

**ESP:** English for Specific Purposes

**FL:** Foreign Language

**FLA:** Foreign Language Acquisition

**ICT:** Information and Communication Technology

**KWIC:** Key Word In Context

**L1:** First Language

**L2:** Second Language

**LT:** Language Teaching

**M1:** Master One

**SL:** Second Language

**SLA:** Second Language Acquisition

**vs:** Versus

**List of Tables**

<b>Table 1.1</b> <i>Lewis's typology of lexical chunks</i> .....	30
<b>Table 1.2</b> <i>Types of Lexical Collocations</i> .....	43
<b>Table 1.3</b> <i>Types of Grammatical Collocations</i> .....	44
<b>Table 2.1</b> <i>Different types of corpora</i> .....	55
<b>Table 2.2</b> <i>Key differences between Chomskyan linguistics and Corpus linguistics</i> .....	57
<b>Table 3.1</b> <i>The content and aim of teachers' interview</i> .....	89
<b>Table 3.2</b> <i>Results of task one</i> .....	93
<b>Table 3.3</b> <i>Results of task two</i> .....	94
<b>Table 3.4</b> <i>Results of task three</i> .....	95
<b>Table 3.5</b> <i>Results of task four</i> .....	96
<b>Table 3.6</b> <i>Results of task four (simplified)</i> .....	97
<b>Table 3.7</b> <i>Results of task five</i> .....	98
<b>Table 3.8</b> <i>The overall results of the test</i> .....	99
<b>Table 3.9</b> <i>The respondents' gender</i> .....	99
<b>Table 3.10</b> <i>The respondents' age</i> .....	101
<b>Table 3.11</b> <i>Respondents' years studying English</i> .....	102
<b>Table 3.12</b> <i>Respondents' familiarity with collocations</i> .....	103
<b>Table 3.13</b> <i>Respondents' exposure to collocations</i> .....	104
<b>Table 3.14</b> <i>Respondents' views on collocation learning</i> .....	104
<b>Table 3.15</b> <i>Difficulties concerning collocations</i> .....	105
<b>Table 3.16</b> <i>The level of difficulty of collocations</i> .....	106
<b>Table 3.17</b> <i>The agreement of respondents towards the statements</i> .....	108



<b>Table 3.18</b> <i>The materials used by teachers</i> .....	110
<b>Table 3.19</b> <i>Respondents' choice when meeting a new word</i> .....	112
<b>Table 3.20</b> <i>The frequency of consulting the internet</i> .....	114
<b>Table 3.21</b> <i>Teacher-centred or learner-centred classroom</i> .....	115
<b>Table 3.22</b> <i>Respondents' preference between authentic and adapted materials</i> .....	116
<b>Table 3.23</b> <i>How often do respondents' lose interest?</i> .....	117
<b>Table 3.24</b> <i>Reasons behind loss of interest</i> .....	118
<b>Table 3.25</b> <i>Respondents' views regarding implementing internet-based methods</i> .....	120
<b>Table 3.26</b> <i>The respondents' willingness to learn through internet-based methods</i> .....	121
<b>Table 3.27</b> <i>Respondents' beliefs towards learning via internet</i> .....	121
<b>Table 3.28</b> <i>Respondents' familiarity with corpus linguistics</i> .....	122
<b>Table 3.29</b> <i>Did teachers introduce the concept of corpora?</i> .....	123
<b>Table 3.30</b> <i>Instances where respondents learned via corpus</i> .....	124
<b>Table 3.31</b> <i>The reasons behind not using corpus programmes</i> .....	125
<b>Table 3.32</b> <i>Teachers' period of teaching</i> .....	126
<b>Table 3.33</b> <i>The courses and levels teachers are in charge of</i> .....	127
<b>Table 3.34</b> <i>Teachers' definitions of collocations</i> .....	129
<b>Table 3.35</b> <i>Teachers' sample responses toward the learners' collocational level</i> .....	130
<b>Table 3.36</b> <i>Interview sample responses about common lexical units</i> .....	134
<b>Table 3.37</b> <i>Teachers' responses regarding potentially impeding factors of learning</i> ...	136
<b>Table 3.38</b> <i>Potential obstacles of integrating DDL</i> .....	146
<b>Table 3.39</b> <i>Teachers' responses to using authentic materials</i> .....	150
<b>Table 3.40</b> <i>Reasons behind using traditional materials</i> .....	151

**Table 3.41** *Teachers' opinions about DDL* ..... 152

### List of Figures

<b>Figure 1.1</b> <i>What is involved in knowing a word (Nation 1990)</i> .....	28
<b>Figure 1.2</b> <i>Cowies' (1988, 2001) classification of word combinations.</i> .....	36
<b>Figure 1.3</b> <i>Distributional categories</i> .....	37
<b>Figure 2.1</b> <i>Example document for each section of the Brown Corpus</i> .....	59
<b>Figure 2.2</b> <i>Sample concordance lines for way from Limerick Corpus of Irish English (LCIE), sorted to the left of the screen</i> .....	62
<b>Figure 2.3</b> <i>Results of the entry word "absolutely" in COCA</i> .....	65
<b>Figure 2.4</b> <i>Collocates of the entry word "absolutely" in COCA</i> .....	66
<b>Figure 2.5</b> <i>The relationship between corpus linguistics (CL) and language teaching</i> ....	68
<b>Figure 2.6</b> <i>Applications of corpora in language teaching</i> .....	69
<b>Figure 3.1</b> <i>The gender distribution of the sample</i> .....	100
<b>Figure 3.2</b> <i>The age range of the sample</i> .....	101
<b>Figure 3.3</b> <i>The level of difficulty of collocations</i> .....	106
<b>Figure 3.4</b> <i>Respondents' choices when meeting a new word</i> .....	112
<b>Figure 3.5</b> <i>How often the respondents consult the internet</i> .....	114
<b>Figure 3.6</b> <i>Reasons behind lack of interest in learning</i> .....	118
<b>Figure 3.7</b> <i>Possible obstacles facing the integration of DDL</i> .....	125

**List of Appendices**

<b>Appendix A: Consent Form for Teachers .....</b>	<b>173</b>
<b>Appendix B: The Test.....</b>	<b>175</b>
<b>Appendix C: The Opinionnaire for the Test .....</b>	<b>178</b>
<b>Appendix D: The Students' Questionnaire .....</b>	<b>180</b>
<b>Appendix E: The Online Questionnaire .....</b>	<b>186</b>
<b>Appendix F: The Opinionnaire for the Questionnaire.....</b>	<b>187</b>
<b>Appendix G: Teachers' Interview Questions .....</b>	<b>189</b>
<b>Appendix H: The Opinionnaire for the Interview Questions.....</b>	<b>190</b>

## Contents

<b>Declaration</b> .....	<b>Erreur ! Signet non défini.</b>
<b>Dedication</b> .....	<b>3</b>
<b>Acknowledgements</b> .....	<b>4</b>
<b>Abstract</b> .....	<b>5</b>
<b>List of Abbreviations and Acronyms</b> .....	<b>6</b>
<b>List of Tables</b> .....	<b>8</b>
<b>List of Figures</b> .....	<b>11</b>
<b>List of Appendices</b> .....	<b>12</b>
<b>Contents</b> .....	<b>13</b>

## General Introduction

1. Background of the Study.....	18
2. Statement of the Problem .....	19
3. The Research Questions .....	20
4. The Research Hypotheses .....	21
5. The Research Aims .....	21
6. The Research Methodology for this Study.....	22
7. Population and Sampling Technique.....	22
8. Significance of the Study .....	23
9. The Referencing Style for this Dissertation .....	23

10. Structure of the Dissertation..... 23

11. Operational Definitions ..... 24

**Chapter One: Collocations: An Overview**

Introduction..... 26

1.1 Vocabulary Knowledge ..... 26

1.2 The Lexical Approach to Language Teaching..... 28

1.3 Formulaic Language ..... 31

    1.3.1 Formulaic Language in Foreign Language Teaching ..... 31

1.4 Collocations: Definitions and Origin ..... 31

1.5 Approaches to Defining Collocations..... 33

    1.5.1 The Frequency-based Approach ..... 33

    1.5.2 The Phraseological Approach ..... 34

    1.5.3 The Lexical Approach..... 37

    1.5.4 The Semantic Approach..... 39

    1.5.5 The Structural Approach..... 40

1.6 Types of Collocations ..... 42

    1.6.1 Lexical Collocations ..... 42

    1.6.2 Grammatical Collocations ..... 43

1.7 The Importance of Teaching Collocations..... 44

1.8 Collocations and Language Fluency..... 45

1.9 Collocational Competence and Communicative Competence.....	47
Conclusion .....	48

## **Chapter Two: Corpus Linguistics and Data-Driven Learning**

Introduction.....	50
2.1 Corpus Linguistics .....	50
2.1.1 What is a Corpus? .....	51
2.1.2 Criteria for Defining a Corpus .....	51
2.1.3 Corpus vs. Database.....	52
2.1.4 Corpus vs. Web.....	52
2.2 The Corpus-based vs. the Corpus-driven Approach.....	53
2.3 Types of Corpora .....	54
2.4 Corpus Linguistics: an Overview .....	56
2.5 Why Is Corpora Important? .....	60
2.6 Corpus and Concordancing.....	61
2.7 Corpus and Collocation .....	63
2.8 Corpus of Contemporary American English .....	64
2.9 The Areas of Applications of Corpus Linguistics .....	66
2.10 Corpus in Language Teaching and Learning.....	67
2.10.1 Syllabus and Material Design .....	69
2.10.2 Data-Driven Learning .....	70

2.10.2.1 Advantages of Data-Driven Learning ..... 72

2.10.2.2 Limitations of Data-Driven Learning ..... 73

Conclusion ..... 75

**Chapter Three: The Research Methodology, Results, and Data Analysis**

Introduction..... 79

3.1 Research Methodology for this Study: Choices and Rationale ..... 79

3.1.1 Research Paradigms ..... 79

3.1.2 Research Approaches ..... 81

3.1.3 Research Design..... 82

3.1.4 Data Collection Methods ..... 83

3.1.4.1 The Test ..... 84

3.1.4.1.1 Structure and Aim ..... 84

3.1.4.1.2 Piloting and Validation ..... 85

3.1.4.2 Students’ Questionnaire ..... 86

3.1.4.2.1 Structure and Aim ..... 86

3.1.4.2.2 Piloting and Validation ..... 87

3.1.4.3 Teachers’ Interview ..... 88

3.1.4.3.1 Structure and Aim ..... 88

3.1.4.3.2 Piloting and Validation ..... 90

3.1.5 Data Collection Procedures..... 90



3.1.6 Data Analysis Procedures .....	91
3.1.7 Population and Sampling .....	91
3.2 Results of the Study .....	92
3.2.1 Results of the Test.....	93
3.2.2 Results of the Students' Questionnaire .....	99
3.2.3 Results of the Teachers' Interviews .....	126
3.3 Discussion and Synthesis of the Results .....	153
Conclusion .....	157
<b>General Conclusion</b> .....	158
Implications and Recommendations.....	160
Limitations and Suggestions for Future Research .....	162
<b>References</b> .....	163
<b>Appendices</b> .....	173
ملخص الدراسة .....	192

## General introduction

### 1. Background of the Study

One of the ultimate goals of language learners is to reach academic proficiency. Language learners are expected to form an all rounded knowledge of the different elements needed in learning a language by the end of their academic journey. It is highly important for English as a foreign language (henceforth EFL) learners to develop a set of adequate vocabulary items that will aid them comprehend and use language in an accurate context. Vocabulary depth and knowledge have been linked to language proficiency and fluency for the past few decades. In fact, developing sufficient vocabulary is necessary for the success of language use because with a limited vocabulary, it will be difficult to achieve accurate communication. There is no doubt that learning grammar rules and sentence structure is important in foreign language settings; however, insufficient vocabulary knowledge can hinder the entirety of the learning experience.

Knowledge of vocabulary encompasses more than just knowing the words and their meanings; it also includes understanding how words occur and function in natural settings. Accordingly, the complex nature of the English language whether the spoken or the written form makes the learning process quite challenging. Learners often find themselves having to learn an abundance of words in order to utilise the target language properly. In addition, The English vocabulary is rich in clusters and word combinations that are formulaic (conventional) and natural to the native speaker but may be difficult for a foreign language learner to comprehend.

Collocations are words that are more often than not found together in a language. Expressions such as: “make the bed” or “do homework” are very prominent in the English language. While these sequences are built in the native speakers’ lexicon, the non-native learner must be aware of the existence of such patterns as well as actively learn them. That is not to

imply that similar patterns do not exist in their mother tongue, but that they are different and often less rigid than those found in the English language. Furthermore, collocational knowledge is necessary for developing the learners' communicative competence, their fluency in speech, and their overall academic achievement. In this regard, many scholars have advocated for the importance of implementing collocation teaching in English as a second language (ESL) and English as a foreign language (EFL) teaching.

One approach to teaching collocations is the corpus-based approach. The integrating of corpus findings and methods made significant impact in foreign language teaching, since corpus linguistics presents real examples of genuine language use. Furthermore, collocations have been a subject of interest in corpus studies throughout the years. The use of online corpus software to teach collocations is one prime example. This is commonly referred to as data-driven learning (DDL). Despite its numerous benefits to the learning experience, this kind of method is still not popularised in language teaching.

## **2. Statement of the Problem**

It is evident that foreign language learners mostly struggle with the aspect of vocabulary learning as it can take multiple years for them to develop a sufficient repertoire. Furthermore, vocabulary knowledge does not only include word recognition or sentence level knowledge. It also includes the awareness of more complicated word combinations and knowing how they function in a native context. The focus on teaching grammar rules and syntax has been given more importance in the English language classroom than lexis, a problem that caused significant deficit in the vocabulary awareness of learners.

Collocations are common lexical units found in natural language use. Consequently, collocations are highly important for building vocabulary and developing language fluency.

Being aware of such existing sequences in the English language can be classified as an essential part of being a proficient English speaker.

However, it was observed last year that a multitude of EFL learners at the level of Master one Applied Linguistics find difficulty in recognizing and developing these formulaic structures. In addition, students often used the incorrect word combinations i.e. using words that do not collocate correctly in the English language. A number of reasons can cause this. One major factor is the lack of instructions and the amount of time dedicated to teaching collocations. Another reason is the limited materials used to help students grasp such a concept. The materials used are usually in form of hand-outs and worksheets which are the traditional mediums used by instructors. Students' involvement in the process also plays a very important role in the comprehension and retention of collocational structures.

One possible solution for this problem is the use of a corpus-based programme which is a search engine that shows forms of texts (words, word combinations, and sentences) extracted from a corpus of a language. In other words, texts that are compiled, organised, and stored in corpora. A concordance line in a concordance programme comprises of the searched target word that is found in the middle of the line surrounded by other words, which regularly occur with it (Haywood, n.d.). This means that corpus software can promote studying words and word combinations in their natural occurring context. Therefore, it is a reliable tool for learning collocations as it gathers data from a sum of authentic and reliable sources as well as displays common word combinations in context.

### **3. The Research Questions**

This research seeks to answer the following research questions:

**RQ1:** What is the current level of collocational competence of first year master students?

**RQ2:** What are the factors impeding students' learning of collocations?

**RQ3:** How do teachers and learners perceive data-driven Learning?

**RQ4:** To what degree is the application of data-driven Learning and a corpus-based approach achievable?

#### **4. The Research Hypotheses**

Based on the abovementioned research questions, we propose the following research hypotheses:

**RH1:** Students will have deficiencies in collocational knowledge.

**RH2:** There are different factors affecting the learning of collocations.

**RH3:** Both teachers and learners will perceive the Data-driven Learning approach positively.

**RH4:** Data-driven Learning and a corpus-based approach are applicable to some degree.

#### **5. The Research Aims**

The general aim of this study is to gain a better understanding of the trouble areas of first year master EFL learners at Mohamed Kheider University regarding their comprehension and use of collocations as an aspect of vocabulary learning.

More specifically, it aims to:

- assess the prior knowledge of master one students on collocations.
- explore the teachers' views on collocational competence.
- discover teachers and learners' perceptions regarding the implementation of the data-driven learning approach.
- understand the possible problems affecting the application of a data-driven approach; and
- highlight the role of web corpora in acquiring collocations.

## **6. The Research Methodology for this Study**

In this particular study, the researcher will opt for a Mixed-methods approach due to the diversity of the research questions and the exploratory nature of the research itself. The latter is an attempt at eliciting the level of collocational knowledge of master one students, exploring the possible reasons behind their inability to appropriately acquire collocations, suggesting and reviewing the probability of integrating the DDL as a potential teaching method, and listing the pitfalls that may prevent its integration. Hence, the use of both quantitative and qualitative instruments was viewed as appropriate.

Concerning the research design, the study will follow a Case Study research design. This choice is in line with the nature of the study as it is meant to gain deeper understanding of the research problem at hand in its natural context as well as obtain the different opinions of the participants. Furthermore, the ultimate goal of this study is simply to answer the research questions and not to generalise the findings.

## **7. Population and Sampling Technique**

The targeted population of this research work is the first year master students as well as the teachers of applied linguistics at Biskra University. Following the criteria of accessibility and availability, the convenience sampling technique will be used. It is the non-probability sampling strategy where the participants do not have equal chances of being selected for the sample. Accordingly, 37 master one students are conveniently chosen to partake in a test on collocational knowledge and to respond to the questionnaire. Moreover, four applied linguistics teachers are interviewed for further insights.

## **8. Significance of the Study**

This present work seeks to gain better understanding at the level of collocational competence that the first year master students at Mohamed Kheider University possess. Admittedly, There is still a negligence of the introduction of formulaic sequences namely collocations in our EFL context in Algeria. For this reason, this study seeks to shed light on the importance of teaching English collocations. Furthermore, the approach of data-driven learning is a rather new and unexplored one in this context as well. It is therefore the aim of the researcher to attempt and understand the issues that prevent implementing such an approach in the EFL classroom of our context. The findings of the study will serve as a contribution to the English language teaching situation and can further help with future considerations for syllabus designers and EFL teachers.

## **9. The Referencing Style for this Dissertation**

Owing to the fact that this work belongs to the social science and humanities discipline, it follows the American Psychological Association (APA) citing and referencing style, 7<sup>th</sup> edition to be precise. Nevertheless, some minor details and options such as layout and cover page are selected following the recommendations of the supervisor.

## **10. Structure of the Dissertation**

This dissertation is organised according to the following structure:

**Chapter One** serves as an overview of the concept of collocation in relation to vocabulary learning and the lexical approach. It highlights the various approaches to research on collocations, the different types of collocations, and the importance of collocations in language teaching, as well as it defines collocational competence.

**Chapter Two** emphasises the importance of corpus linguistics and corpus evidence, especially in language teaching. It accounts for corpus in relation to collocations and concordance software. It further explains the intricacies of the Corpus of Contemporary American English and how it can be used in order to learn collocations. Finally, it explains data-driven Learning, its advantages, and its limitations.

**Chapter Three** involves the fieldwork and data analysis section of the dissertation. It seeks to analyse, display, describe and classify the obtained data in order to accurately report the findings. Consequentially, through the obtained findings, results can be discussed and summarised in this section.

## 11. Operational Definitions

A number of terms require some elucidation to determine how and in which sense the researcher uses them.

**Corpus.** The plural form of this word is corpora and this term is used to refer to the collection of different text in the English language whether spoken or written often found in a computer data-base (McCarthy, 2004).

**Collocations.** Words that are commonly found together in a language.

**Concordance.** This use of the term is in corpus linguistics and according to O’Keeffe et al. (2007) it is a tool “to find every occurrence of a particular word or phrase” (p. 8).

**Concordance line.** It is defined as “Collections of example sentences extracted from a corpus” (Lee et al., 2017).



## **Chapter One: Collocations: An Overview**

Introduction

**1.1** Vocabulary Knowledge

**1.2** The Lexical Approach to Language Teaching

**1.3** Formulaic Language

**1.3.1** Formulaic Language in Foreign Language Teaching

**1.4** Collocations: Definitions and Origin

**1.5** Approaches to Defining Collocations

**1.5.1** The Frequency-based Approach

**1.5.2** The Phraseological Approach

**1.5.3** The Lexical Approach

**1.5.4** The Semantic Approach

**1.5.5** The Structural Approach

**1.6** Types of Collocations

**1.6.1** Lexical Collocations

**1.6.2** Grammatical Collocations

**1.7** The Importance of Teaching Collocations

**1.8** Collocations and Language Fluency

**1.9** Collocational Competence and Communicative Competence

Conclusion

## **Introduction**

This chapter intends to provide a closer look into the concept of collocations as a lexical phenomenon. It explores the aspect of collocations in relation to vocabulary knowledge, the main principles of the lexical approach to language learning, and formulaic language. It also presents an overview of the history and origin of this lexical phenomenon, its different approaches and types. Additionally, this chapter highlights the importance of collocations and its correlation to language fluency and competence.

### **1.1 Vocabulary Knowledge**

Although different researchers have approached the identification of vocabulary knowledge from different perspectives, the term vocabulary knowledge generally means the knowledge of words and what they mean. It encompasses different categories of knowledge which includes morphological, semantic, grammatical, collocational, connotative, associational, and social knowledge (Laufer & Goldstein, 2004).

According to Moghadam et al. (2012) vocabulary knowledge is “a systematic procedure in which various types of knowledge are learned until all aspects of knowledge are known for an item.” (p. 557). It is therefore observed that multiple aspects need to be taken into account when learning a new word. In other words, the knowledge of words is embedded in the ability to accurately recognise, understand, and produce words and word combinations in any given context. This ability for the native speakers of the language is natural and effortless since they acquire the correct forms of words innately as their vocabulary expands while they grow older. It is however, an on-going task for the non-native language learner. The foreign language learner must actively learn words at some point in their learning process at an attempt to gain a better understanding of how the language works.

In the recent decades, the importance of vocabulary has been highlighted in English Language Teaching (ELT). It has been linked to the four skills of language, especially the role it plays in reading comprehension. In their work on vocabulary knowledge, Anderson and Freebody (1981) linked between the linguistic ability and vocabulary knowledge in a theory they called the instrumentalist hypothesis. They hypothesised that vocabulary knowledge explicitly influences the quality of text comprehension (p. 80). In addition, it is claimed that vocabulary facilitates the decoding process during reading which serves as background knowledge (Moghadam et al., 2012).

Meara (1996) (as cited in Hasan & Shabdin, 2016) argued that foreign language (FL) and second language (SL) learners who are equipped with adequate vocabulary knowledge are more competent in using the language than those who have limited vocabulary knowledge. It is agreed that unlocking this aspect of language learning can give the learners an advantage over their peers whether in the productive or the receptive aspect of language.

Moreover, Nation (1990 as cited in Hodne, 2009) suggested the concept of what constitutes learning a word. He presented the various components in learning vocabulary words displayed in figure 1.1 below. It is worth mentioning that R stands for receptive knowledge while P stands for productive knowledge.

**Figure 1.1**

*What is involved in knowing a word (Nation 1990)*

<b>Form</b>	Spoken form	R	What does the word sound like?
		P	How is the word pronounced?
	Written form	R	What does the word look like?
		P	How is the word written and spelled?
<b>Position</b>	Grammatical Patterns	R	In what patterns does the word occur?
		P	In what patterns must we use the word?
	Collocations	R	What words or types of words can be expected before or after the word?
		P	What words or types of words must we use with this word?
<b>Function</b>	Frequency	R	How common is the word?
		P	How often should the word be used?
	Appropriateness	R	Where would we expect to meet this word?
		P	Where can this word be used?
<b>Meaning</b>	Concept	R	What does the word mean?
		P	What word should be used to express this meaning?
	Associations	R	What other words does this word make us think of?
		P	What other words could we use instead of this one?

*Note.* Retrieved from *Investigating word combinations in two English textbooks for Norwegian upper secondary school students*, by S. Hodne, 2009, p. 3.

To sum up, Vocabulary learning is a complex procedure with many aspects to consider. The more vocabulary a FL learner has, the better chances he/she has at mastering their target language. Besides, insufficient vocabulary can hinder the quality of the receptive and the productive skills of the individual.

### **1.2 The Lexical Approach to Language Teaching**

Developed by Michael Lewis in 1993, the lexical approach is an approach to English language teaching where the main premise is that language consists of lexical items rather than grammatical structures, for example: fixed expressions and language chunks formed by collocations (*Lexical approach*, 2019). It also supports the idea that language learners should be taught common phrases found in the language rather than singular vocabulary items or lists

(Nordquist, 2019). Lewis emphasised the role of successful communication rather than grammatical memorisation and the concept of correctness (Nemoianu, n.d.).

Lewis's view on language teaching favours the teaching of lexis he argued that "language consists of grammaticalized lexis, not lexicalized grammar." (as cited in Mudraia, 2001, p. 2). It goes against former beliefs that grammar is a prerequisite for language learning and prioritises the learning of lexical items rather than grammatical rules. The lexical approach supporters often rationalise that language is formed by meaningful sets of words that can be combined and formed into continual comprehensible texts or content (Mudraia, 2001). Lewis (1997) as cited in Mudraia (2001) suggested the implementation of those lexical items: multiple word units or polywords (by the way, upside down), collocations (commit suicide, absolutely appalled), conventional utterances (we'll see, if I were you), text phrases (firstly, in this paper). (p. 3).

Many scholars sought to classify the different lexical units into varied categories through the years. This area is still blurred, because different researchers use different approaches regarding the classification of such items. Lewis's (1993) categorisation of lexical chunks is worth mentioning in this regard (as cited in Zhao, 2009).

**Table 1.1***Lewis's typology of lexical chunks*

Types of lexical chunks	Examples
<p><b>Polywords:</b> extension of words, which is composed of more than one word. And it is often considered to be the essential vocabulary for learners to acquire.</p>	<p>as soon as, on the one hand, talk about, after all, grow up and so on</p>
<p><b>Collocation:</b> refers to pairs of words that frequently co-occur with each other. These frequent associations merge into habitual connection and sometimes they are in a fixed order.</p>	<p>fixed order: knife and fork, bread and butter verb+noun: play the basketball, shake hands, catch a cold adjective+noun: bright red, splendid future</p>
<p><b>Institutionalized utterances:</b> Chunks that are called whole units and conventionalized in the language. They tend to express pragmatic rather than referential meaning. The chunks may be full sentences, usable with no variation but always with instantly identifiable pragmatic meaning.</p>	<p>accepting: I'd be delighted to offering: can I give you a hand supposing: If I were you...</p>
<p><b>Sentence frames and heads:</b> serve as the framework builder of the whole sentences.</p>	<p>frames and heads: It is suggested that..., The fact is..., My point is that... composition frames: This paper concentrates on..., firstly, secondly..., finally...</p>

*Note.* Retrieved from *An empirical study of L2 learners' use of lexical chunks and language production*, by Zhao, Y., 2009. p. 9.

### **1.3 Formulaic Language**

Formulaic language or formulaic sequences are defined as those fixed combinations of words that have different functions in speech production, and are frequently stored and recalled in the brain as a whole lexical item (Wood, 2006). This also comprises multiword units that are less rigid and can be modified according to context. These lexical items include: collocations, idioms, phrasal verbs, lexical bundles...etc.

#### **1.3.1 Formulaic Language in Foreign Language Teaching**

The attention given to formulaic patterns is still underwhelming regardless of the significant part that it takes in language. Studies in foreign language acquisition (FLA) often focus on the explicit instructions of vocabulary items as single words while less emphasis is given to learning fixed combinations in language (Sirkel, 2017). Implementing the teaching of such sequences can improve the learners' awareness of this aspect of the language, as well as increase their communicative competence. It is also worth mentioning that the benefits of learning those fixed expressions can manifest in more fluent and native-like speech. Furthermore, Guz (2014) emphasised the influence of formulaic sequences on speaker's fluency by arguing that they reduce "the cognitive pressures involved in the construction of fluent speech" (p. 115). This essentially means that storing and retrieving word combinations as one item puts less constraints on the brain during the process of speaking.

### **1.4 Collocations: Definitions and Origin**

The origin of the word collocation is derived from the Latin verb *Collocare* which means "to place together" or "to arrange" (Martyńska, 2004, p. 2). Simply put, collocations are the group of two or more words that are often found together in language. The word part Co-meaning together and -location meaning to place (Beare, 2018).

According to Wood (2019) collocations are words that co-occur together which subscribe to specific semantic or syntactic criteria. Lewis (1997) gave another definition to collocations: “collocations are those combinations of words which occur naturally with greater than random frequency.” (as cited in El-Dakhs, 2017, p. 69).

The works of Harold E. Palmer (1933), in his book, “Second Interim Report on English Collocations, A New Classification of English Tones”, pioneered the introduction of the concept of collocations. In modern linguistics, the British linguist John Firth is also considered as a founding father of the study of collocations due to his contributions to the field in 1950. When it comes to crediting the founder of such a concept, it can be controversial. The works of Palmer are often discredited while research credits Firth as the pioneer of the study on collocations. It can be due to the lack of shared information in past times that left people uninformed with Palmer’s works (Williams & Millon, 2012). Even though only a minority had access to his works, it still heavily influenced works on Phraseology (the study of fixed expressions or word combinations). Moreover, Nation (2000) observed that collocation is often referred to as a Firthian term, yet he argued that Palmer used it many years before and that he produced valuable reports on English collocations.

While palmer had published his innovative works in Japan for ELT purposes, Firth was conducting his own framework on collocations in parallel without having heard of Palmer or his contributions (Williams & Millon, 2012). An explanation to the different reactions to their works is the fact that these two scholars have advocated for such different frameworks regarding the English language (Williams & Millon, 2012).

Phraseologists and lexicographers attempt to categorise language in order to draw clear lines within their frameworks and to produce better and more precise publications. (Williams &



Millon, 2012). Phraseologists supported the fixed view of collocations as explicitly linguistically bound units with occasionally pragmatic considerations. While the NeoFirthian approach which was developed by John Sinclair perceived collocations from a wider angle in which a collocation is a “dynamic process in which meanings are created and exploited within textual contexts” (Williams & Millon, 2012, Introduction section, para. 1).

### **1.5 Approaches to Defining Collocations**

Different scholars have used different approaches to clearly define collocations over the years, some following a more rigid phraseological approach while others following a more flexible one. Classifying collocations often depends on syntactic, semantic, or lexical restrictions. Some researchers favour one aspect over the others. This idea raises the question of what exactly is considered a collocation and what is not.

The main two approaches that are prominent in any body of literature discussing collocations are the frequency-based approach and the phraseological approach.

#### **1.5.1 The Frequency-based Approach**

The frequency-based approach is rooted in the works of John Firth (1951, 1957) and “it deals with statistical probabilities of words appearing together” (Wood, 2019, p. 31). This approach defines a collocation as the frequent occurrence of a word in lexical combinations compared to its appearance in normal language use (Wood, 2019). Collocates are defined as words that occur near a certain “core” or “node” word i.e. the central word of a collocation. For instance, the expressions: good night, good bye, good morning are all collocations with the node word “good” while the following words: night, bye, morning are considered collocates. It is also noticed that those words were situated exactly one position to the right of the core word (Webb & Kagimoto, 2011). The most known advocates for this approach are namely: Halliday, Sinclair,

Kjellmer, Mitchell and Greenbaum. These scholars involved syntactic and semantic aspects for describing collocations. In fact, Johns and Sinclair (1974) as cited in Wood (2019) discovered that the ideal frequency of a collocation is four words to the left or the right of a node word. Kjellmer established a theory about collocational continuum in which word combinations range from free unrestricted collocations to fixed and invariable collocations (Martyńska, 2004). He also concluded that collocations are sequences that recur in language use with well-established grammar.

This notion helped establish the new study of collocations which is frequency initiated and computer-based (Wood, 2019). The frequency-based approach adapts an inductive view to identifying phraseological units instead of the traditional deductive view. Furthermore, instead of defining those units on linguistic basis it applies corpus-based approaches to highlight lexical co-occurrences in language (Granger & Paquot, 2008). Dissimilar to the traditional approach, supporters of this approach view phraseology as integral i.e. phraseological sequences have priority over single items; however, this view faced waves of criticism over the years (Granger & Paquot, 2008). For example, Gaatone (as cited in Granger & Paquot, 2008) accounted for the role of multi-unit expressions while still disclaiming the idea that everything in language is phraseological.

### **1.5.2 The Phraseological Approach**

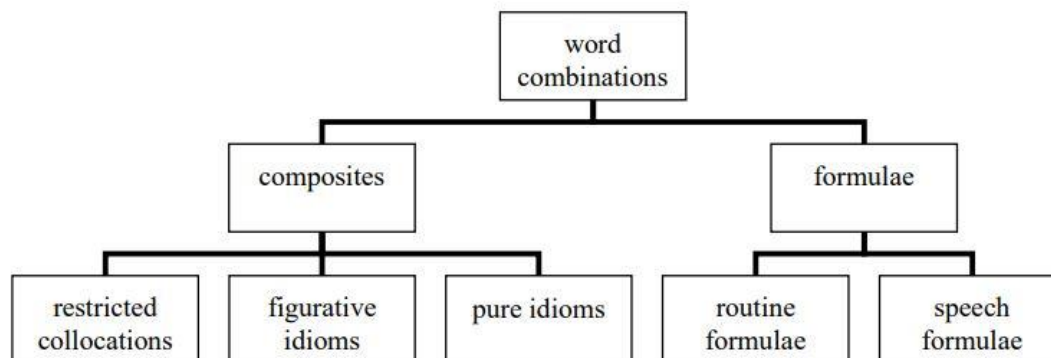
This particular approach emphasises on “establishing the semantic relationship between two (or more) words and the degree of noncompositionality of their meaning” (Gablasova et al., 2017, p. 158). In other words, this framework studies and describes word combinations according to their level of fixedness and flexibility. The phraseological views come from the old Soviet and other Eastern European countries’ phraseology (Granger & Paquot, 2008; Wood,

2019). Collocations in this approach are no different from other phraseological items as idioms. Thus, collocations are simply understood as linguistic units which meaning is dependent on the phrase in which its located (Kunin, 1996; Canli & Canli, 2013; Kalayci, 2012, as cited in Varlamova et al., 2016).

Moreover, this approach perceives collocations as more rigid multiple word combinations and its classification of what constitutes a collocation is less flexible (Wood, 2019). This framework to phraseology views word combinations (including collocations) as a spectrum or continuum moving from “free”, unrestricted combinations on one end to static and fixed combinations on the other (Granger & Paquot, 2008). Phraseologists view collocations as multi-word units which vary and have clear meaning to some level (Wood, 2019). The insights of Cowie (as cited in Granger & Paquot, 2008; Wood, 2019) are the most note-worthy in this approach. Hence, he provided a continuum of word combinations ranging from composites, which are units with syntactic and lexical descriptions that can often be fully fixed and invariable (idiomatic expressions); as well as formulae which consists of pragmatic conventions (e.g. how are you?). In addition, Cowie mentioned restricted collocations in which one part is figurative and another is literal (Wood, 2019, p. 4). This is illustrated in the following figure.

**Figure 1.2**

*Cowies' (1988, 2001) classification of word combinations.*

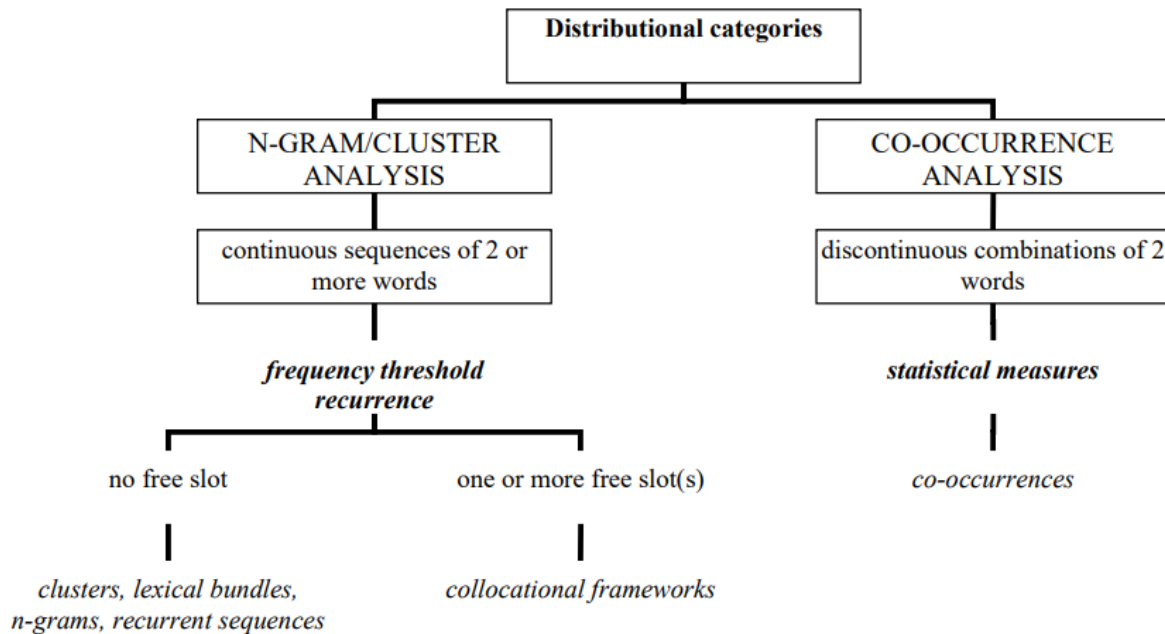


*Note.* Retrieved from “Disentangling the phraseological web” by Granger, S. and Paquot, M. (2008). *Phraseology: An Interdisciplinary Perspective*, September 2014, 27–49.

<https://doi.org/10.1075/z.139.07gra>

This approach is commonly known as the traditional approach among scholars, as it does not attempt at assessing the tendencies of word co-occurrences in language use. It rather primarily focuses on linguistic formations of those units.

Furthermore, Granger and Paquot (2008) put forward a typology of the various types of phraseological units according to their extraction criteria and methods.

**Figure 1.3***Distributional categories*

*Note.* Retrieved from “Disentangling the phraseological web” by Granger, S. and Paquot, M. (2008). *Phraseology: An Interdisciplinary Perspective*, September 2014, 27–49.

<https://doi.org/10.1075/z.139.07gra>

There exist other classifications and frameworks to research on collocations, which further proves the challenging nature of collocations as an aspect of language learning. Some of the other approaches that became known are:

### **1.5.3 The Lexical Approach**

The lexical approach advocates support the idea that the meaning of certain words is dependent on its co-occurring words. First initiated and developed by John Firth, this approach considers lexis to be an independent and detached element from grammar (Martyńska, 2004; Danilevičienė & Vaznonienė, 2018). In addition, according to Martyńska (2004) a part of determining a word’s meaning lies in the fact that it collocates with certain words; nevertheless,

these combinations are often limited. Firth believed that the meaning of words is not static and that it is inherently correlated to the context in which the words are located. In fact, his famous phrase “you shall know a word by the company it keeps” serves as an emphasis to the importance of collocations. He also insisted that collocations are not merely the juxtaposition of words; but also a matter of mutual expectancy (Manca, 2012). As a result, he introduced the term “meaning by collocation”. Firth (1957) explained:

Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with dark, and of dark of course, collocation with night. (as cited in Manca, 2012, p. 40)

M. K. Halliday and John Sinclair were students and followers of Firth’s framework, Sinclair being the developer of the “Neo-Firthian” approach.

Halliday emphasised the crucial role of collocations in lexis. He believed that collocations transcend grammatical boundaries, for example he argued that the sentences “he argued strongly” and “the strength of his argument” are grammatical derivations of the collocation “strong argument” (Martyńska, 2004, p. 2). In addition, he introduced the term “lexical set”, which is a group of lexemes that frequently collocate (Danilevičienė & Vazonienė, 2018).

As mentioned before, Sinclair was one of Firth’s students. He developed the Firthian framework into what is referred to as the Neo-Firthian approach under the umbrella of corpus studies. The central element of this approach is collocations and collocation analysis from textual evidence. Sinclair was the pioneer of the terminology “node” “collocates”, and “span” in collocation studies. He later dismissed the previous belief that lexis is strictly independent from

grammar. Consequently, he adapted a new approach where both grammatical and lexical factors are accounted for in the study of collocations (Martyńska, 2004).

Moving from the belief that full fixedness and full freedom of word choice are rare, Sinclair elaborated on two principles to explain how words are derived and understood from texts (Manca, 2012). The open-choice principle is the idea that by the end of each word, phrase or clause a range of possible choices is presented or “open” with regards to grammatical restrictions nonetheless (Sinclair, 1991, as cited in, Manca, 2012). Meanwhile, the idiom principle considers the idea that language users can choose from a wide range of semi-preconstructed expressions that form a single choice, even though they may seem dividable to different parts (Sinclair, 1991, as cited in Erman & Warren, 2000). Followers of this approach view collocations as “independent lexemes” and prefer their analysis to be at the lexical level (Danilevičienė & Vazonienė, 2018).

#### **1.5.4 The Semantic Approach**

This approach exceeds the mere observations of collocations and attempts at examining them from the semantic view point separately from grammar as well (Martyńska, 2004; Danilevičienė & Vazonienė, 2018). The main goal for the advocates of this approach is to investigate the reason words collocate with certain words. For instance, why does the word “blonde” collocate with the word “hair” but not with the word “car” (Martyńska, 2004, p. 3). John Sinclair highlighted the semantic aspect in collocations in his later works. He further explained that meaning is not explicitly attached to words: “It is anticipated that meanings also arise from the loose and varying co-occurrences of several words, not necessarily next to each other” (Sinclair, 1996, as cited in, Danilevičienė & Vazonienė, 2018, p. 22). According to him,

lexical semantics cannot be separated from grammar, because without meaning of combined lexemes our lexicon becomes empty (Danilevičienė & Vazonienė, 2018).

### **1.5.5 The Structural Approach**

According to the structural approach, collocation is decided according to structure, and patterns are identifiable. The views of this approach oppose the before mentioned approaches in terms of grammar. The structural approach supporters believe that the study of collocations should encompass grammar and that grammar and lexis are inseparable (Martyńska, 2004). For this reason, two categories of collocations became known: lexical collocations and grammatical collocations. These categories represent different aspects of the same phenomenon (Martyńska, 2004; Danilevičienė & Vazonienė, 2018). Grammatical collocations often consist of a noun, an adjective, a verb plus a preposition or a grammatical unit, such as “to+ infinitive” or “that+ clause” (Martyńska, 2004, p. 3). Lexical collocations involve nouns, adjectives, verbs as well as adverbs, but do not involve propositions, infinitives or clauses (Martyńska, 2004).

Among the well-known supporters of this framework R. Carter, G. Kjellmer, M. Lewis, and S. Hunston are mentioned.

Carter (1998, as cited in Danilevičienė & Vazonienė, 2018) defined collocations as a set of recurring words in language. He pointed out grammatical collocations that are not limited to grammatical relations, but can also immerge from co-occurrence of lexical items in specific contexts (Danilevičienė & Vazonienė, 2018). Additionally, he noted that the overall number of words that collocate with a certain word is called a “cluster” of that specific word.

G. Kjellmer attempted at distinguishing collocational and non-collocational word classes. He then concluded that articles, prepositions, singular nouns, and base forms of verbs were collocational in nature while adjectives, proper nouns, and adverbs were non-collocational



(Martyńska, 2004). Furthermore, he claimed that English words are dispersed across a spectrum ranging from completely contextually predictable items to completely unpredictable ones (Martyńska, 2004). His results indicated that most items appear at the beginning of the spectrum. He also described a “scale of fixedness” in which collocations go from totally free, unconstrained combinations to totally fixed and invariable ones (Martyńska, 2004; Danilevičienė & Vaznonienė, 2018).

Lewis identified collocations as “a subcategory of multi-word items, made up of individual words which habitually co-occur and can be found within the free-fixed collocational continuum” (Martyńska, 2004, p. 4). He argued that most collocations are located in the middle of the continuum, which indicates that strong collocations are rare. Consequently, he distinguished between strong, medium strong, and common collocations (Danilevičienė & Vaznonienė, 2018). He further explained that common collocations make up various numbers of word combinations and that medium strong collocations are considered the largest part of the lexis needed for the language learner e.g. “significantly different” (Martyńska, 2004, p. 4).

According to Lewis, collocations differ from other multi-word items because collocations provide more insights about what the language user is trying to express rather than what the language user is doing e.g. apologizing. He believed that collocation is not decided based on frequency or logic and that it is rather arbitrary and conventionally determined (Martyńska, 2004).

Hunston concluded that grammatical sequences and lexical meaning are correlated (Martyńska, 2004). All words can be found in certain patterns and those words and their meanings share common traits; that is to say that a word displays a specific meaning when it co-occurs with another specific word in a set (Danilevičienė & Vaznonienė, 2018).

## 1.6 Types of Collocations

Several researchers attempted to classify the different variations of collocational patterns, namely Benson et al. (2010) who collected and divided collocations into two main groups: grammatical collocations and lexical collocations. Grammatical collocations consist of a “dominant” word (noun, adjective, and verb) and a preposition or grammatical structure. While lexical collocations do not include a dominant word. They suggest the following patterns for lexical collocations: verb + noun, adjective + noun, noun + verb, noun + noun, adverb + adjective, adverb + verb (p. xiii).

### 1.6.1 Lexical Collocations

A lexical collocation is defined by Haji (2015) as a phrase that involves syntagmatic relations between content words (noun, adjective, verb, and adverb) such as: “draw a conclusion” (p. 2). There are commonly eight distinguished patterns of lexical collocations. Lexical collocations can be found in these patterns: verb+ noun, adjective+ noun, noun+ noun, verb+ adverb, adverb+ adjective, verb+ adjective, noun+ verb, noun+ adjective.

Furthermore, Yan (2010), Hamed Mahvelati and Mukundan (2012), and Haji (2015) highlight these types of lexical collocations as well as grammatical ones.

**Table 1.2***Types of Lexical Collocations*

Type	Example
Verb+ noun	Make mistakes
Adjective+ noun	Strong wind
Noun+ noun	Human resources
Verb+ adverb	Appear suddenly
Adverb+ adjective	Absolutely right
Verb+ adjective	Turn grey
Noun+ verb	Dog barks
Noun+ adjective	Crystal clear

*Note.* Adapted from *Collocation: Theoretical Considerations, Methods and Techniques for Teaching it*, by Haji, S. 2015. p. 2.

**1.6.2 Grammatical Collocations**

As mentioned above, grammatical collocations are represented in form of combinations of content words and function words (prepositions, particle) or content words and grammatical structures (to+ infinitive or “that” clause). For example: “by chance”, “interested in”.

**Table 1.3***Types of Grammatical Collocations*

Type	Example
Preposition+ noun	In agony
Noun+ preposition	Increase in
Verb+ preposition	Rely on
Adjective+ preposition	Familiar with
Quantifier+ noun	A drop of water
Preposition+ preposition	Apart from

*Note.* Adapted from *Collocation: Theoretical Considerations, Methods and Techniques for Teaching it*, by Haji, S. 2015. (pp. 2-3).

There is an increase in interest of investigating both types of collocations as they both serve different functions in language. Generally, it is observed that lexical collocations are more common in language use than grammatical ones; nevertheless, both are essential for the sound use of language whether spoken or written.

### **1.7 The Importance of Teaching Collocations**

The increasing number of research works on the language phenomenon that is collocations is standing proof of its importance in language learning. Scores of researchers have attempted to draw the attention of language teachers to these multi-word units as they emphasised their impact on language fluency in speech and writing. However, the most prominent impact remains on the acquisition of vocabulary, which is crucial to language learning.

Additionally, there has been a shift from language studies focusing on grammar to being more focused on lexis and vocabulary acquisition in the last decades. For this reason, the surge of focus on collocations in relation to vocabulary acquisition has developed, because collocational knowledge is part of vocabulary knowledge. In fact, Nation (2000) devoted a section in his book “Learning Vocabulary in Another Language” to collocational knowledge called “Language knowledge is collocational knowledge”. He explained, “Language knowledge and use is based on associations between sequentially observed language items. This viewpoint sees collocational knowledge as the essence of language knowledge.” (2000, p. 518). Moving from the idea that collocational knowledge is at the base of language learning, both regular and idiomatic collocations are given importance, with emphasis on the most frequent forms (Nation, 2000). The major contributor of the quality of learning by word association is the frequency of exposure to these language forms in use i.e. the law of practice (Nation, 2000). Although Nation specified that explicit study of collocations has a major role in learning, he suggested that most of the learning will take place through meaning focused receptive and productive language use (2000, p. 523).

### **1.8 Collocations and Language Fluency**

Most of the studies regarding collocations have linked its knowledge to language proficiency and “native-like” fluency. James Carl (as cited in Rao, 2018) viewed that the accurate use of collocations greatly contributes to the speaker’s idiomaticity and native likeness (p. 3). In addition, Lewis (1997) noted, “fluency is based on the acquisition of a large store of fixed and semi-fixed prefabricated items” (p. 15). Sonaiya (1988, as cited in Rao, 2018) further explained that successful communication relies on word choice; thus, lexical errors need to be considered more seriously.

Experimental studies also supported the claim that knowledge of collocations helps with the development of receptive and productive skills, more specifically, speaking and writing. In her research conducted in 2016, Demiray Akbulut attempted to answer the research question “Does collocational teaching have a positive effect on academic writing?” (2018, p. 69). The researcher investigated the effect of teaching collocational expressions on students’ writing skill through a quasi-experimental design that included one control and one experimental group. The experimental group was taught collocational expressions. After the treatment period both groups were tested on their rewriting skills of passages, they were previously exposed to. The results of the data analysis showed that the experimental group outperformed the control group and it was noticed that “there is a strong relationship between collocational vocabulary size and effective writing” (p. 78).

In another study, Attar and Allami (2013) explored the impact of collocations on ESL learners’ speaking skill. Two groups of learners were assigned to an experimental and a control group. A pre-test and an interview on collocations were conducted with both groups. Afterwards, the experimental group undertook the treatment regarding collocations in use. A post-test was later conducted on both the experimental and the control group in forms of a test and an interview as well. The results of the t-test proved that the experimental group displayed superior development in their speaking skills in the post-test. Furthermore, data analyses showed that the participants from the experimental group increased their performance during the partaken interview. In addition, Attar and Allami stated that the students showed positive attitudes regarding the explicit instruction of collocations. They concluded their study by implying that teaching collocations in the EFL/ESL context is a better method for training students to become more “native-like speakers” (p. 1070).

### **1.9 Collocational Competence and Communicative Competence**

The study of collocations received a great deal of interest through the years; however, researchers in ESL and EFL studies still believe it is an under investigated area in language teaching. Thus, the existing literature on collocations often stresses the importance of implementing collocations in the teaching syllabuses. That is because learning collocational patterns can benefit language learners significantly and in many aspects. As Hodne (2009) demonstrated after refereeing to multiple research works: “including collocations in course material is both attainable and highly recommendable.” (p. 2)

It was Hill (1999 as cited in Miščin & Pavičić Takač, 2013) who introduced the term “collocational competence” as an emphasis on the major role of collocations in language learning. She stressed that the knowledge of lexis is not limited to the overall meaning of words but also their collocational span (Miščin & Pavičić Takač, 2013). According to Martyńska (2004) learners are less likely to develop collocational competence when they are not able to collocate words properly. Consequentially, if learners lack lexical combinations in their mental lexicon, they will divert to making word combinations based on grammar rules which may lead to erroneous language use (Martyńska, 2004). Besides, studies have also indicated that collocational errors are the most common errors committed by non-native speakers (James, 1998 as cited in Miščin & Pavičić Takač, 2013).

Furthermore, research on lexical chunks, particularly collocation, displayed that learning and memorising word combinations as full units is more efficient than learning individual word units. Hence, it is more helpful for the processes of vocabulary retention and production.

Martyńska (2004) illustrated that “A wide range of meaningful chunks and collocations in the learner’s mental lexicon makes it possible to quickly find the right word. It also facilitates and

accelerates the communication process.” (p. 11). Thus, collocational knowledge leads to language competence and more proficient performance. As collocational knowledge develops, it enables more accurate use of word combinations, which can be an indicator of effective language use. This accurate use of language can be linked to the learners’ overall communicative competence that is conveniently defined as “a learner's ability to use language to communicate successfully.” (*Communicative competence*, n.d.).

### **Conclusion**

This chapter aimed at gaining insights on the concept of collocations from multiple perspectives. The focus was on collocations as an integral part of vocabulary acquisition and its importance. Moving from the lexical approach to language teaching, formulaic language was mentioned. Moreover; history, origin, and approaches of collocational patterns were introduced. This chapter also made distinctions between types of collocations as well as it stressed the impact of learning collocations on language proficiency and performance. In the forthcoming chapter, the researcher will provide an account for the field of corpus linguistics as well as the method of data-driven learning.



## **Chapter Two: Corpus Linguistics and Data-Driven Learning**

Introduction

**2.1** Corpus Linguistics

**2.1.1** What is a Corpus?

**2.1.2** Criteria for Defining a Corpus

**2.1.3** Corpus vs. Database

**2.1.4** Corpus vs. Web

**2.2** The Corpus-based vs. the Corpus-driven Approach

**2.3** Types of Corpora

**2.4** Corpus Linguistics: an Overview

**2.5** Why Is Corpora Important?

**2.6** Corpus and Concordancing

**2.7** Corpus and Collocation

**2.8** Corpus of Contemporary American English

**2.9** The Areas of Applications of Corpus Linguistics

**2.10** Corpus in Language Teaching and Learning

**2.10.1** Syllabus and Material Design

**2.10.2** Data-Driven Learning

**2.10.2.1** Advantage of data-driven learning

**2.10.2.2** Limitations of data-driven learning

Conclusion

## **Introduction**

This chapter provides an account on the history of corpus linguistics. It also highlights the meaning of corpus, the corpus-based and the corpus-driven debate, and a selection of corpora types. In addition, it accounts for the importance of corpus studies and corpus linguistics (CL). It discusses corpus and concordances, corpus and collocations, and supplies information regarding the Corpus of Contemporary English. Moving to areas of application, it lists the disciplines that CL influences namely the area of language teaching and learning which includes syllabus design and data-driven learning (DDL). By the end of the chapter, the intricacies of DDL are explained, as well as its advantages and limitations as a language teaching/learning approach.

### **2.1 Corpus Linguistics**

The study of language has developed and progressed through the years. The emergence of modern day technology and computer innovations has also helped revolutionise the way language is analysed, understood, and taught.

Corpus linguistics (CL) is the field of study that deals with linguistic phenomena as found in a language's corpus or corpora. In other words, CL deals with language as data extracted from large computer-analysed language entries. McEnery and Hardie (2012) defined it as, "An area which focuses upon a set of procedures, or methods, for studying language." (p. 1). Although, CL is not limited to methodology, it can be far more complex and broad in its scope. They explained:

While some generalisations can be made that characterise much of what is called 'corpus linguistics', it is very important to realise that corpus linguistics is a heterogeneous field. Differences exist within corpus linguistics which separate out and subcategorise varying approaches to the use of corpus data. (p. 1)

### **2.1.1 What is a Corpus?**

In order to fully grasp corpus linguistics, it is evident that the word corpus needs to be defined. Numerous prominent researchers in CL agree that corpus is a selection of spoken or written authentic materials that has been compiled for particular purposes (Flowerdew, 2012). These purposes are often linguistic in nature but can also be socio-pragmatic (Flowerdew, 2012). Sinclair (1991) defined corpus as “a collection of naturally-occurring language text, chosen to characterize a state or variety of a language.” (p. 171). Thus, the existence of corpus largely has to do with the diversity of natural language use in different contexts. Another distinctive feature of corpus is that it provides machine-readable data that can be accessed and analysed (McEnery & Hardie, 2012; Flowerdew, 2012).

### **2.1.2 Criteria for Defining a Corpus**

Flowerdew (2012) identifies four principals for defining a corpus:

- A corpus consists of authentic, naturally occurring data;
- A corpus is assembled according to explicit design criteria;
- A corpus is representative of a particular language or genre;
- A corpus is designed for a specific linguistic or socio- pragmatic purpose. (p. 3).

It is observed that developing or defining a corpus is not an arbitrary procedure but rather quite a complex and time-consuming one. In fact, one of the central issues in defining a corpus is its size. The size of corpus is dependent on its purpose. Hence, according to Flowerdew (2012) large scale corpora with general purposes are usually in the range of 100 million to 500 million words; whereas, more specialised, genre-based corpora can be from around 50,000 to 250,000 words.

However, in general scholars namely Sinclair, argue that the larger the corpus the better. It is in favour of a corpus to contain more texts and data for a better and more thorough representation of language. In addition, Sampson (2001 as cited in Flowerdew, 2012, p. 4) argued that “a sizable sample of real-life usage” is needed for formulating and testing hypothesis about the language, which can be seen as another purpose of corpus-based studies by itself.

### **2.1.3 Corpus vs. Database**

A corpus and a database (text archive) are different and not to be mixed with each other. Flowerdew (2012) depicted the differences between a corpus and a database and stated:

A corpus is a collection of naturally occurring language, which has been systematically planned and collected in accordance with principled external design criteria with an a priori purpose in mind, which, in turn, determines the design parameters. A database, or text archive, on the other hand, is a large repository of text which is unstructured and often compiled according to what is easily obtainable rather than based on systematic sampling techniques. There is also a difference in the ‘reading’ of a corpus vs a database: a corpus is read non- linearly whereas it is usually a whole text which is accessed in a database. (p.7)

It is clear from the provided distinction that corpus and database differ in a number of ways. The main distinction is that, unlike most corpora, a database does not have to adhere to a specific structure or serve a specific purpose.

### **2.1.4 Corpus vs. Web**

Yet another interesting point of discussion has been whether the web is a large corpus or whether it contains corpus properties. Sinclair (2005) (as cited in Flowerdew, 2012) has denied the fact that the web is a corpus he argued, “The World Wide Web is not a corpus, because its

dimensions are unknown and constantly changing, and because it has not been designed from a linguistic perspective.” (p. 7).

However, this statement has been debated by scholars who highlight the idea that web is indeed a corpus, if a corpus is seen as a compilation of data which is considered as an object of language (Flowerdew, 2012). In other words, they stress the fact that if a corpus is appropriate for a certain task or language study, the web in this case can identify as a corpus. Nevertheless, the main argument against this stance is the question of representativeness. Representativeness has to do with the extent to which a corpus can accurately represent language aspects.

Flowerdew (2012) further explained, “If we wish to develop a corpus of general English, we may think it should be representative of general English, so we then need to define the population of ‘general English- language events’ of which the corpus will be a sample.” (p. 8).

## **2.2 The Corpus-based vs. the Corpus-driven Approach**

There is a debate or more accurately, a distinction between corpus linguists regarding the areas that corpus linguistics covers. The main question is whether corpus linguistics is considered as a methodology or that it can expand to include theoretical formation. The corpus-based approach supports the idea that CL is a methodology. In other words, corpus-based studies use corpus data as means to explore, validate, refine, confirm, or disconfirm an existing hypothesis or theory regarding language (McEnery & Hardie, 2012). On the other hand, the corpus-driven view argues that CL has strong theoretical properties and it rejects the idea of CL as a methodology (McEnery & Hardie, 2012). That is to say, the corpus-driven approach underpins the idea that theories and hypotheses can be formed solely based on corpus evidence. “It is thus claimed that the corpus itself embodies its own theory of language” (McEnery & Hardie, 2012, p. 6).

Some linguists support the corpus-based approach while others support the corpus-driven one. Most advocates for the corpus-driven view are followers of John Sinclair (Ngula, 2018). Meanwhile, other linguists argue that all corpus studies should be classified under the corpus-based framework, namely McEnery, Wilson, and Hardie (Ngula, 2018). Despite this debate, it is important to note that corpus linguists on both ends have supported and worked together even co-writing journals (Ngula, 2018). “Indeed, recent corpus studies tend to apply key ideas from the two camps” (Ngula, 2018, p. 208).

### **2.3 Types of Corpora**

There is a surprisingly big number of different corpora which vary according to their aims. Wynne and Berglund Prytz, (2012) distinguish between seven different types of corpora. These are shown in Table 2.1.

**Table 2.1***Different types of corpora*

Type of corpora	Brief definition	Example
Balanced, representative	Texts selected in pre-defined proportions to mirror a particular language or language variety	- Brown family - BNC: British National Corpus
Monitor	New texts added by and to 'monitor' language change.	- BoE: Bank of English - COCA: Corpus of Contemporary American English
Parallel (translation)	Same texts in two (or more) languages	- OPUS: open source parallel corpus - ENPC English-Norwegian Parallel Corpus
Comparable	Similar texts in two, or more, languages or language varieties	- ICLE: International Corpus of Learner English
Diachronic	Include texts from different (consecutive) periods, preferable comparable ones	- Helsinki Corpus of English Texts - COCA
Specialized	Include a specific type of text	- Air Traffic Control Speech corpus - Lampeter Corpus of Early Modern English Tracts
Multi-media	Include multi-media material (video recordings, transcriptions)	- SACODEY

*Note.* Adapted from "Types of corpora and some famous (English) examples" by M., Wynne & Y., Berglund Prytz, 2012. In *Corpus Linguistics course*. OUCS Hilary.

<http://tinyurl.com/669o4zt>

Other types of corpora include learner corpora, raw, tagged, and error-annotated corpora, to name a few. A learner corpus is a corpus that contains language produced by language learners whether speech or writing (*Corpus types*, 2016). It can be extremely beneficial in terms of error analysis and measuring learners' progress. A raw corpus is a corpus with no annotation, while a tagged corpus has annotation e.g. parts of speech annotation (*a glossary of corpus types*, 2017).

Leech (2005) defined corpus annotation as the act of adding interpretive and linguistic information to a corpus. He provided the example of adding labels such as word class. Error-annotated corpora contain texts produced by language learners and translators, which happen to display errors and error annotations (*Corpus types*, 2016). They can be used for investigating different errors committed by different groups of learners (*Corpus types*, 2016).

#### **2.4 Corpus Linguistics: an Overview**

Corpus linguistic has developed throughout the years. Two main phases are known in the history of its development: the corpora BC and the corpora AD. The BC stands for before computers and AD refers to after the advent of Digitalisation (information converted into digital forms which can be processed by computers) (Flowerdew, 2012).

The history of collecting texts and language data traces back to works of lexicographers and pre-Chomskyan structural linguists who essentially collected data as part of their works. First, the collections were in paper form that posed a great deal of challenges to the scholars and particularly dictionary makers at the time. After multiple years of working with a paper corpus, Dr. Samuel Johnson pioneered the first exhaustive English dictionary in 1755 (McCarthy & O'Keeffe, 2010). It was "endless slips of paper logging samples of usage from the period 1560 to 1660." (McCarthy & O'Keeffe, 2010, p. 4). The works of Johnson at the time were seen as a new



way of studying language. It was a way of learning and testing knowledge through empirical evidence and investigation that preceded the scientific development (Flowerdew, 2012).

American structuralists in the 1950s, namely Fries, Harris, and Hill carried what is known as corpus linguistics. To those linguists corpus was an essential component of language descriptions and as a result, the idea of collecting real life data flourished at the time (McCarthy & O’Keeffe, 2010). Nevertheless, in the following 20 years the evolution witnessed a decrease especially with the rise of the Chomskyan linguistic view.

**Table 2.2**

*Key differences between Chomskyan linguistics and Corpus linguistics*

Chomskyan linguistics	Corpus linguistics
Competence (internalised knowledge of a language, i.e. what can be said or written)	Performance (external evidence of language competence, i.e. what is actually said or written)
Linguistic ‘facts’ accessed through introspection or intuitive means. Data are not ‘objective’, i.e. not verifiable	Linguistic ‘facts’ based on attested instances of authentic language. Data are observable and therefore verifiable
Structure: what is ‘grammatical’ or ‘ungrammatical’? Judgements based on usually artificial, i.e. invented examples	Use: what are the ‘degrees of grammaticality’, and what is ‘acceptable’? Judgements based on naturally occurring data
Potentially infinite number of examples, accommodating the concept of creativity	Finite number of examples which focus on formulaicity

*Note.* Retrieved from “Corpora and language education” by L. Flowerdew, 2012. *Research and Practice in Applied Linguistics* (pp. 39-40). Hong Kong University of Science and Technology, Hong Kong SAR, China.

From the table above clear distinctions can be observed regarding corpus and Chomskyan linguistics. Furthermore, the stance of Noam Chomsky on CL is apparent. Chomsky as a well-established mentalist was against the rise of corpus evidence at the time. He favoured a more

intuitive and introspective view in dealing with language. In other words, Chomsky advocated for native speaker intuition over empirical or corpus data. However, the concept of native speaker intuition has been open to debate over the years due to its problematic nature in the linguistic scene. This point posed significant conflict because Chomskyan linguists followed an intuitive method of describing language and argued against observable data based on language use.

The idea that the native speaker by nature can decide what is correct or not has been challenged to some extent, especially in regards to word combinations and their frequency in use. Francis (1993, as cited in McGee, 2006) argued: “Intuition may be useful to linguists in a number of ways, but for the purposes of saying exactly how language is used, it is notoriously unreliable” (p. 3). In addition, Biber et al. (1996) observed that, “Intuitions regarding lexical associations are often unreliable and inaccurate” (p. 120). However, scholars mainly corpus linguists often agree that both approaches need to be taken into account while dealing with language. Flowerdew (2012) explained: “rather than viewing these two camps as diametrically opposed to each other, it may be more fruitful to focus on their respective insights into language and adopt a more accommodationist approach where the advantages and drawbacks of each are acknowledged.” (p. 40).

“It was not the linguistic climate but the technological one that stimulated the development of corpora.” (Bonelli, 2010, p. 15). In spite of the support the Chomskyan linguistic had during the 1950s and 1960s, the rise of corpora and corpus linguistics did not stop. It especially grew stronger with the development of technology and the invention of machine-readable corpora (Flowerdew, 2012).

The digital era being on the rise as well as the advancement of the internet provided a well-established environment for the evolution of CL. With these innovations, the first electronic corpus was introduced to the scene that was the Brown corpus. It was the initial computer-readable corpus of written language, established in the 1960s at Brown University by Nelson Francis and Henry Kuc̆era (Bonelli, 2010). The corpus contains text from 500 sources that have been categorised by genre, such as news, editorial, and more (Bird et al., 2009).

### Figure 2.1

*Example document for each section of the Brown Corpus*

ID	File	Genre	Description
A16	ca16	news	Chicago Tribune: <i>Society Reportage</i>
B02	cb02	editorial	Christian Science Monitor: <i>Editorials</i>
C17	cc17	reviews	Time Magazine: <i>Reviews</i>
D12	cd12	religion	Underwood: <i>Probing the Ethics of Realtors</i>
E36	ce36	hobbies	Norling: <i>Renting a Car in Europe</i>
F25	cf25	lore	Boroff: <i>Jewish Teenage Culture</i>
G22	cg22	belles_lettres	Reiner: <i>Coping with Runaway Technology</i>
H15	ch15	government	US Office of Civil and Defence Mobilization: <i>The Family Fallout Shelter</i>
J17	cj19	learned	Mosteller: <i>Probability with Statistical Applications</i>
K04	ck04	fiction	W.E.B. Du Bois: <i>Worlds of Color</i>
L13	cl13	mystery	Hitchens: <i>Footsteps in the Night</i>
M01	cm01	science_fiction	Heinlein: <i>Stranger in a Strange Land</i>
N14	cn15	adventure	Field: <i>Rattlesnake Ridge</i>
P12	cp12	romance	Callaghan: <i>A Passion in Rome</i>
R06	cr06	humor	Thurber: <i>The Future, If Any, of Comedy</i>

*Note.* Retrieved from “Accessing Text Corpora and Lexical Resources” by Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O’reilly.

The Brown corpus allowed access to lists of words and sentences within specific categories to choose from as shown in figure 2.1. It also provided opportunities for studying systematic differences and similarities between linguistic genres (Bird et al., 2009). It notably contains a million words of American English from documents that had been published in the year 1961 (Bonelli, 2010). Although it may seem to be small compared to new and much larger electronic corpora, the Brown corpus is still in use to this day and it remains as the standard corpus.

### **2.5 Why Is Corpora Important?**

The importance of corpus investigation and corpus linguistics manifested in the interest that it has been given in recent years. Furthermore, because corpora provide evidence that can be used in various disciplines, it has been studied for different reasons. Corpus studies yield important data that can answer language related questions. McEnery and Xiao (2011) emphasised the importance of corpus analysis in providing empirical and objective data. They argued: “One of the strengths of corpus data lies in its empirical nature, which pools together the intuitions of a great number of speakers and makes linguistic analysis more objective” (p. 364). Consequently, such evidence led to advancements in the different areas of linguistics namely dictionary making and language pedagogy.

Investigating corpus information can be beneficial in providing language teachers with insights on: word frequency counts, keyword analysis, cluster analysis, lexico-grammatical profiles, lexicography, grammar, stylistics, translation and concordance (Yılmaz & Soruç, 2015, p. 2627). Similarly, Krieger (2003) agreed that corpus-based analysis can be used on multiple linguistic patterns: lexical, structural, lexico-grammatical, discourse, phonological,

morphological. The investigation can have specific objectives, for example distinguishing between male and female usage of tag questions (Krieger, 2003).

Another interesting aspect of language that corpus studies cover is language nuances. According to Bennett (2010) nuances refer to questions about language that students may ask yet language teachers may not have the answers to. “Often, the questions specifically relate to areas of collocation and frequency”, Bennett stated. This has to do with the arbitrary nature of language. Some language aspects exist with no apparent rule or explanation; this is where corpus evidence plays a significant role. Because corpus evidence yields objective data unlike that of native-speaker intuition John Sinclair (1998 as cited in Krieger, 2003) pointed out: “this is because speakers do not have access to the subliminal patterns which run through a language” (The Advantages of Doing Corpus-Based Analyses, para. 1).

Corpus linguistics provides important insights regarding language and how language works. It helps point out language nuances that are hard to detect without saved evidence of use and attempts to generate theories on why it may happen.

## **2.6 Corpus and Concordancing**

A special feature that corpus linguistics has provided to the linguistic scene is that of concordancers. A concordancer or concordance software is a tool that can be used to search for words, word combinations, or phrases displayed in Key Word In Context (KWIC) format (Meyer, 2004). Essentially, that means that words are shown in surrounding text in form of a line (the context) with the searched or key word also known as the “node” highlighted. That is referred to as the concordance line.

O’Keeffe et al. (2007) define concordancing as “using corpus software to find every occurrence of a particular word or phrase.” (p. 8). Indeed, concordance programmes present rich

data about how words occur in natural language. Not only does it present those words in context, but also it displays them based on their frequency of occurrence. This can be extremely useful for foreign language learners especially for noticing lexical and grammatical patterns.

Nevertheless, it takes some practice to be able to read a concordance line format. O’Keeffe et al. (2007) explain: “Concordance lines challenge us to read in an entirely new way, vertically, or even from the centre outwards in both directions.” (p. 8). The researcher using the corpus can choose the number words seen in the concordance line in order to view more of the context surrounding the node (O’Keeffe et al., 2007). Figure 2.2 further demonstrates how it works.

### Figure 2.2

*Sample concordance lines for way from Limerick Corpus of Irish English (LCIE), sorted to the left of the screen*

```

ether in northern Ireland is no different in a way then em what they were desperately
you see it? Some of you anyhow? Now in a way 'What Dreams may come' it's not
subject to study in college in fact it's a way of life and you find this right
and how could he present things in such a way that he would persuade people.
ul and the purpose of life is to live in such a way that when you die your soul is
t he was obviously he obviously lived a certain way of live and they wanted to know
lem that they had to deal with in a different way they couldn't deal with it by
asically in football stadium that's the easiest way to describe it. There is a large
sking for you ok I find this the most effective way. Ok now today em you have as well
speculative because there is no evidence either way. You can't have evidence about
e theologian starts from the top and works his way down. The theologian will have
rts from the ground so it speaks and works its way up. The theologian starts from

```

*Note.* Retrieved from *From Corpus to Classroom: language use and language teaching*, by O’Keeffe, A., McCarthy, M., & Carter, R. (2007). Cambridge University Press.

Concordance programmes developed through the years to include general large corpora as well as new features to facilitate the research process. For example, some software display lemmas of words, which allows the search for every form of the verb “to be” for example (Meyer, 2004). Other programmes include clusters i.e. word combinations found in language.

Some famous concordance programmes are, to name a few: WordSmith, British National Corpus (BNC), and Corpus of Contemporary American English (COCA).

### **2.7 Corpus and Collocation**

A Central idea to corpus studies is that language highly consists of patterns and that it consists of a high degree of re-current sequences with lexis and grammar being inseparable (Römer, 2008). This new perspective led to the development of the Collins COBUILD English Course (CCEC) (Römer, 2008). It followed the pioneering COBUILD project by Sinclair and was pedagogically oriented in lexicography. This new lexical syllabus focused on teaching “the commonest words and phrases in English and their meanings” (Willis, 1990 as cited in Römer, 2008, p. 114).

The concept of collocations has long been linked to studies of CL and corpus analysis, since collocations are known for their frequency of occurrence in a formulaic manner. Corpus studies allow better understanding of how words re-occur in speech and writing thanks to statistical analysis of corpus evidence. Corpus studies on collocations also demonstrate which words are frequently used together. O’Keeffe et al. (2007) attested that “Corpus descriptions have also enhanced our understandings of units of fixed phrasing, collocation, and more extended language patterns” (p. 23). Likewise, Flowerdew (2012) believed that a corpus is crucial for detecting collocational sequences that are “visible through the vertical display of the node word” (p. 18).

Furthermore, collocations can have certain pragmatic meanings especially in specialised language; corpus analysis can detect such meanings (Koester, 2010). For example, Flowerdew (2008, as cited in Koester, 2010) observed that the collocation “associated with” was very frequent in an environmental reports corpus.

## **2.8 Corpus of Contemporary American English**

The Corpus of Contemporary American English (COCA) is a free online corpus of the American English language. It is the most commonly used English corpora on the web. It contains more than one billion words used from the year 1990 to 2019 (English Corpora: Most Widely Used Online Corpora. Billions of Words of Data: Free Online Access, n.d.). It comprises data derived from different genres including academic journals, written fiction, spoken entries from movies or television shows, newspapers or magazines, and website blogs. Not only is it balanced in genre, but it is also constantly updated which is useful for instructional purposes (Gee, 2015). The COCA website contains a search engine that allows detailed and complex searching. It can be very useful for language specialists yet; it is simple enough for learners to use. Although, after a number of entries it requires the user to sign up for an account in order to continue using the platform. The process of signing up is cost-free and basic; similar to other platforms it requires an email address and a password.



**Figure 2.3**

Results of the entry word “absolutely” in COCA

The screenshot displays the COCA interface for the word "absolutely". The top navigation bar includes "SEARCH", "WORD", "CONTEXT", and "ACCOUNT". The "WORD" tab is active, showing the word "absolutely" (ADV) with a frequency of #980. A bar chart shows the distribution of "absolutely" across different sources: BLOG, WEB, TV/M, SPOK, FIC, MAG, NEWS, and ACAD. The definition is provided: "1. completely and without qualification 2. totally and definitely". Below the definition are links for "YouGlish", "PlayPhrase", and "Yarn", along with a "Translate: choose language" option. A "SYNONYMS" section lists words like "definitely", "totally", "completely", "enormously", "entirely", "extremely", "utterly", "very", "unconditionally", "categorically", "undeniably", "unequivocally", and "unquestionably". On the right side, there are sections for "TOPICS" (e.g., "is, fabulous, unidentified, ok, dress, sort, stunning, beautiful, wear, wonderful, fashion, inaudible, magazine, unintelligible, mm, oscar, basically, incredible, pleasure, dr") and "COLLOCATES" (e.g., "obama, john, larry, us, america, al, george, clinton" for NOUN; "adore, terrify, thrill, devastate, corrupt, forbid, stun, astound" for VERB; "right, true, necessary, correct, certain, essential, amazing, wonderful" for ADJ; "positively, unequivocally, spot, categorically, unconditionally, stunningly, irrevocably, fantastically" for ADV).

As figure 2.3 shows, while using COCA, The learner will be able to search for target words and will have access not only to the definition and to synonyms of that word, but also the context in which it is used. In addition, a section of the words that commonly collocate with the target word is displayed. Thus, COCA can be beneficial for learning new vocabulary items and collocations.

**Figure 2.4**

*Collocates of the entry word “absolutely” in COCA*



In figure 2.4 collocates of the word “absolutely” are displayed according to their word class (noun, adjective, verb, adverb) and are ordered in frequency of occurrence. Also, once the user clicks on a collocate it provides the context in which it appears, in form of a concordance line.

### 2.9 The Areas of Applications of Corpus Linguistics

Corpus linguistics has long been used for different purposes. It provides the grounds for the analysis of language in order to achieve a variety of objectives. In the general sense, it seeks to understand what features are associated with lexical and grammatical patterns (Bennett, 2010). Thus, arguably, the advancements of CL has led to the refinement of lexical descriptions and consequently led to the enhancement of dictionary making. Additionally, it allowed for empirical and detailed studies of aspects of grammar (McCarthy & O’Keeffe, 2010). Nevertheless, CL is being applied to much broader disciplines. It can be applied in areas such as: language teaching

and learning, discourse analysis, pragmatics, English for Specific Purposes (ESP), literary stylistics, forensic linguistics, speech technology, sociolinguistics and many more (Bennett, 2010; McCarthy & O’Keeffe, 2010).

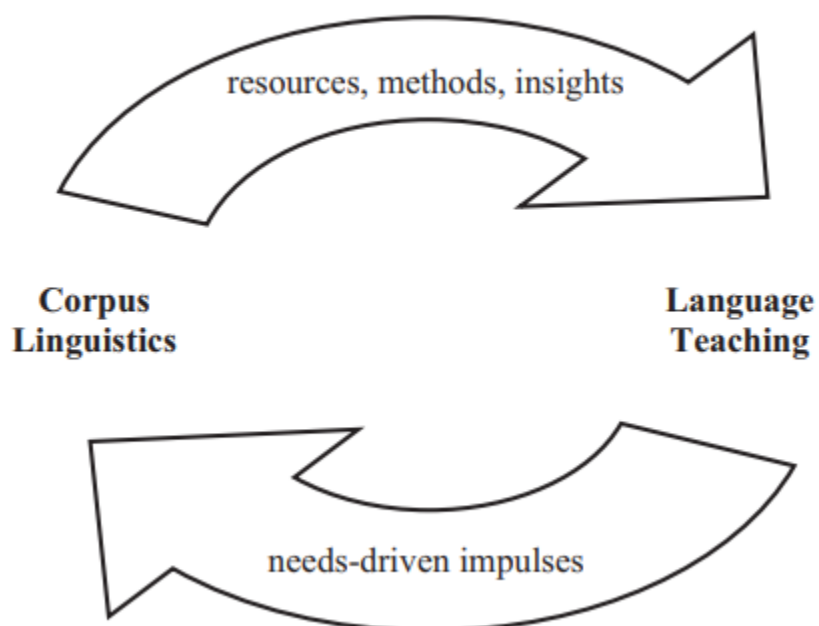
### **2.10 Corpus in Language Teaching and Learning**

The vast field of language teaching (LT) and learning has witnessed quite the change since the 1960s. The emphasis in language teaching has shifted away from the need to teach learners grammatical structures and toward developing learners' communicative competence (Beeching, 2014). This led the emergence of Communicative Language Teaching (CLT) as an approach to language teaching in the 1970s. Consequently, this new approach made language researchers and material designers alike emphasise the importance of authentic materials in the LT situation. The emphasis on authenticity in LT helped with the further exploitation of corpus data and evidence.

Moreover, the role of corpora in language teaching and learning has been rather significant, since corpus analysis provides a better understanding on how genuine language works as well as it provides a closer look on its' intricacies. According to Römer (2008) the relationship between CL and LT is dynamic in nature. Studies in CL provide important insights, methods, and resources that LT benefit from. Simultaneously, the requirements and impulses of LT influence the research projects of corpus linguistics as well as the development of its resources and tools. This concept is illustrated in figure 2.5.

**Figure 2.5**

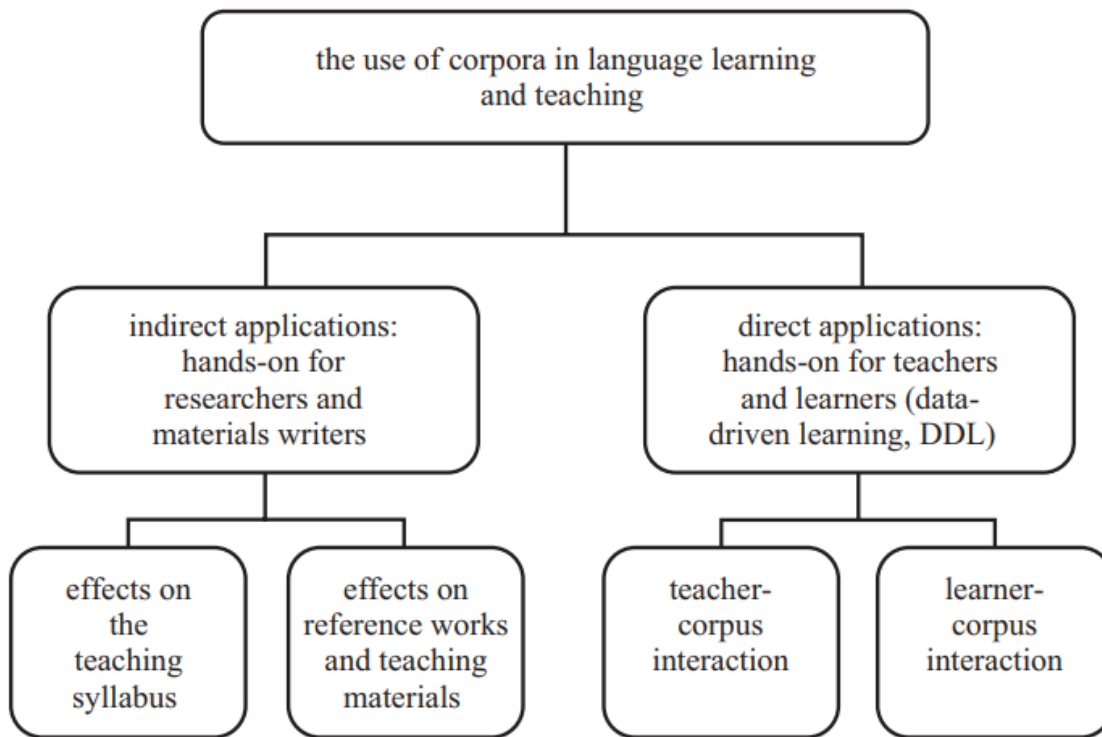
*The relationship between corpus linguistics (CL) and language teaching (LT)*



*Note.* Retrieved from “Corpora and language teaching”, by U. Römer, 2008, p. 113. In A.

Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*. W. De Gruyter.

Although corpora can contribute to language teaching and pedagogy in numerous ways, generally, there are two main methods of applying corpus data into language teaching and learning. On the one hand, there is the indirect use of corpora in LT, which is the use of corpus data for material and syllabus design. On the other hand, there is a direct use of corpora in the learning experience that is also known as Data-Driven Learning (DDL).

**Figure 2.6***Applications of corpora in language teaching*

*Note.* Retrieved from “Corpora and language teaching”, by U. Römer, 2008, p. 113. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*. W. De Gruyter.

### 2.10.1 Syllabus and Material Design

Corpus linguistics contributes to the creation of syllabi and course materials through the data it provides. Beeching (2014) explained: “publishers and researchers can use corpus samples to create language-learning syllabuses and materials.” She continued, “The learners themselves do not have access to the corpus but the corpus informs the way that language is presented to students in learning materials.” (p. 3). In addition, corpora of teaching materials can also be analysed and compared to learners’ L1 corpora in order to evaluate the used materials and enable the selection of more appropriate pedagogical materials (Gabrielatos, 2005).

As it was mentioned earlier in the chapter, the focus on developing the language learners' communicative competence highlighted the need for communicative-focused teaching syllabi. As a result, the use of general large corpora was highly efficient in the process of designing language teaching syllabi which emphasise communicative competence and promote the teaching of real-life communicative events (Römer, 2008).

The language syllabus is usually developed according to what the learners need to know about the target language. A corpus supplies information regarding frequency of occurrence, register, and specific knowledge for specific learners which is helpful in the process of selecting course materials and course planning (Bennett, 2010; Krieger, 2003). Statistical information accessed through corpus can also help teachers in setting priorities for the classroom (Bennett, 2010).

In his book, "Corpus, Concordance, Collocation", John Sinclair (1991) was an advocate for using authentic materials found in a language corpus as means of education. He viewed corpus and concordancers as an essential device for describing modern linguistics as well as for language teaching and learning. Nonetheless, he favoured the use of such materials in syllabi designs and teaching materials rather than first-hand use in the classroom (Wong, 1993).

### **2.10.2 Data-Driven Learning**

Indeed, in its early beginnings linguists, language researchers, and syllabus designers explicitly used corpora and corpus information. Meanwhile, development in Computer Assisted Language Learning (CALL) made it possible for learners to have access to language corpora directly and easily. It allowed language learners to benefit from reliable data extracted from authentic texts and materials without second hand assistance. This notion is known as Data-Driven Learning (DDL). At this point, it is worth noting that DDL can be used directly and

indirectly. The direct use is through guided discovery tasks accomplished with the use of corpus and concordancing software. The indirect use is in form of concordance and concordance line printouts. The use of the term DDL in this paper mainly refers to the direct application of DDL i.e. using a concordance software in the learning context.

Boulton (2017) defined data-driven learning as “the learner’s ability to find answers to their questions by using software to access large collections of authentic texts relevant to their needs, as opposed to asking teachers or consulting ready-made reference materials.” (para. 1). Thus, in DDL Language is viewed as data that can be processed and analysed by the learners while language learners are allowed access to corpus software, acting as researchers themselves. In this sense, corpus is used as a direct language resource or reference (Boulton, 2010). DDL as an approach to LT and learning can be found in numerous works dating back to the 1980s. The most noteworthy of these works is namely the work of Tim Johns. Johns alongside his colleague King first introduced the term data-driven learning in 1991:

[...] an application of computers to language-learning that has come to be known as ‘classroom concordancing’ or ‘data-driven learning’ (DDL) – the use in the classroom of computer-generated concordances to get students to explore the regularities of patterning in the target language, and the development of activities and exercises based on concordance output. (as cited in Boulton, 2010, Introduction, para. 2)

DDL challenges traditional approaches of LT and attempts changing the classroom situation for the better. Hence, the role of the teacher as controller is minimised while his/her role as facilitator is emphasised. Furthermore, the learners are responsible for their own learning and are in fact viewed as an active agent in this approach. Nonetheless, like any teaching or learning approach especially in the context of ELT it was a subject of scrutiny in the field. As there are

multiple advantages to the use of DDL inside the language classroom, there are also some limitations that need to be acknowledged.

### ***2.10.2.1 Advantages of Data-Driven Learning***

One of the most obvious attributes of DDL is that it promotes learner autonomy. Holec (as cited in Smith, 2008) described learner autonomy as, “People's *ability* to take charge of their own learning” (p. 395). Learners in a DDL context benefit from first hand interaction with genuine data. They take control over their own learning process and learn at their own pace. It is true that DDL gives the learners a sense of freedom by reducing the control the teacher has on the learning experience; nevertheless, it does not imply they have complete control. Little (1999) further demonstrated: “freedoms conferred by autonomy are never absolute, always conditional and constrained.” (p.5). That is to say, DDL promotes a learner-centred approach to teaching while preserving the important role of the teacher as a guide in the learning process.

Another advantage of using corpus and concordancers in a DDL approach is the exposure to authentic material. Gavioli and Aston (2001) argued that corpora provide opportunities for learners to witness and use real life discourse themselves. They explained: “the question is not whether corpora represent reality but rather, whether their use can create conditions that will enable learners to engage in real discourse, authenticating it on their terms– and whether this engagement can lead to language learning.” (p. 240).

One major implication that this approach has is that it is inductive in nature. It replaces the traditional teaching paradigm of presentation, practice, production with an identification, classification, generalisation paradigm or some may also suggest illustration, interaction, induction (Boulton, 2010). Following Johns' reasoning, Boulton (2010) also noted that in this approach the language learner is a linguistic researcher, testing and investigating hypotheses, or a



language detective, recognising and interpreting context clues. This initially emphasises critical thinking and inducting skills and helps learners with their development.

The application of corpus in the language classroom was dubbed successful in terms of the acquisition of both explicit and implicit knowledge. Implicit knowledge is the unconscious acquisition of language through “meaning-focused interaction”, while explicit knowledge is in the form of conscious learning through direct instruction (Cheng, 2010). “This explicit instruction can speed up implicit learning processes when supported with language items, patterns, and rules” (Cheng, 2010, p. 320).

Moreover, DDL can increase learners’ motivation by having them participate in a new method of learning that can be interesting and appealing. Learners nowadays tend to use technological tools and applications regardless of the context they are in. they can be more inclined to learn if the teaching environment is on a par with current technological advancements that they are well aware of (internet use, software, applications, etc.). Kettemann (1995, as cited in Römer, 2008) also stressed the exploratory element of DDL and viewed concordancing in the ELT context as “motivating and highly experiential” for language learners (p. 118).

It is therefore noticed that the DDL approach can be highly beneficial especially in the EFL context. Hence, EFL learners can make use of natural occurring language evidence, foster their autonomy, build their inductive reasoning and critical thinking skills, increase their motivation, as well as acquire implicit and explicit knowledge of the English language.

#### ***2.10.2.2 Limitations of data-driven learning***

Like any teaching method, the DDL method has seen its fair shares of setbacks since its emergence. Aside from the obvious issues such as material needs and the appropriate learning

environment, this rather new method still has a long way before it becomes popularised in the EFL teaching context. This is partially due to some requirement issues among other reasons.

The initial setback to the achievement of this approach is ultimately the lack of technological advancement or materials in the LT context. Although, that in itself is dependent on the teaching context (some communities are more advanced than others are). Johns (1988) observed that the achievement of DDL is “dependent not only on the social, cultural and political setting of a particular society at a particular point in time and the development of education within that setting but also on the technology available in the classroom” (p. 13).

Another obstacle that comes to light is the amount of preparation that this approach needs. It is evident that every novice approach needs a significant amount of work and DDL is no exception. It calls for a number of requirements from both teachers and students. According to McCarthy and O’Keeffe (2010, p. 7), “the teacher has to do a lot of preparation work in building up students’ skills of investigation leading to hands-on work with corpora or concordance print outs”. Indeed, teachers must introduce this concept in way that is accessible and clear enough for their students. Not to mention, teachers themselves can be unfamiliar with the use of concordancing software let alone the students. In addition, the selection of the corpus must be done carefully as it must be relevant to the learners’ needs (Krieger, 2003). General large corpora can be used but it can be tricky and the process of learning can be tedious.

Furthermore, the drastic shift of the learning strategy to an inductive one puts some constraints on the part of the learners. Students that were once presented with knowledge are asked to seek knowledge themselves. Additionally, reading a set of concordance lines is essential in the DDL approach and is not some skill that is automatic, “It demands new micro-cognitive

skills whereby the reader moves from phrase pattern to meaning by way of hypothesising and inference.” (McCarthy & O’Keeffe 2010, p. 8).

Krieger (2003) highlighted another student-related issue that has to do with the different learning styles. He stated, “For some students, discovery learning is simply not the optimal approach.” (Problematic Issues Involved, para. 2). Although some students can be highly involved and interested using CALL, some may face difficulties because they rather receive knowledge and not be involved in the learning process.

Even though these limitations can demotivate both teachers and learners, they can be solved with presentence and positive attitudes. Besides, this only stresses the fact that employing a new teaching approach or technology in the classroom needs a considerable amount of planning and open-mindedness.

All the above mentioned obstacles may pose a challenge for the application of DDL; however, it does not negate the impact that it can have. In fact, a multitude of research works proved its positive effect on the four skills of language learning and on learners’ abilities. Especially, its impact on vocabulary knowledge, collocational knowledge, and the writing skill (for example, Rahimi & Momeni, 2012; Yılmaz & Soruç, 2015; Lee et al., 2017).

## **Conclusion**

This chapter attempted to provide an overview on the history of corpus linguistics. It established definitions about corpora and corpora types, as well as about the different approaches involved in corpus studies. It mentioned corpus and the concept of concordances and collocations, and provided a closer look at COCA. The focus in this chapter was on corpus linguistics in relation to language learning and teaching with an emphasis on data-driven learning. The following chapter will discuss the methodological choices that underpin the current

study. In addition, it will display, describe, and categorise the obtained data. Finally, it will summarise and synthesise the findings.

## **Chapter Three: The Research Methodology, Results, and Data Analysis**

### Introduction

#### **3.1 Research Methodology for this Study: Choices and Rationale**

##### **3.1.1 Research Paradigms**

##### **3.1.2 Research Approaches**

##### **3.1.3 Research Design**

##### **3.1.4 Data Collection Methods**

###### **3.1.4.1 The test**

###### **3.1.4.1.1 Structure and aim**

###### **3.1.4.1.2 Piloting and validation**

###### **3.1.4.2 Students' questionnaire**

###### **3.1.4.2.1 Structure and aim**

###### **3.1.4.2.2 Piloting and validation**

###### **3.1.4.3 Teachers' interview**

###### **3.1.4.3.1 Structure and aim**

###### **3.1.4.3.2 Piloting and validation**

##### **3.1.5 Data Collection Procedures**

##### **3.1.6 Data Analysis Procedures**

##### **3.1.7 Population and Sampling**

#### **3.2 Results of the Study**

##### **3.2.1 Results of the Test**

##### **3.2.2 Results of the Students' Questionnaire**

**3.2.3** Results of the Teachers' Interviews

**3.3** Discussion and Synthesis of the Results

Conclusion

## **Introduction**

This chapter represents a rationale for the study at hand and accounts for its methodology namely, the research paradigm, research approach, research design, and the chosen sampling technique. Furthermore, it lists the utilised data collection methods, data collection procedures, as well as data analysis procedures. This chapter also presents the obtained data and its analysis. Besides, it displays and interprets the results of the data gathered from the three data collection tools. Finally, this chapter is concluded by discussing and summarising the major findings of the study.

### **3.1 Research Methodology for this Study: Choices and Rationale**

This section seeks to depict and display the methodological choices of this study and provides a rationale behind each choice. In addition, it attempts to highlight the essential components of research methodology namely, research paradigms, research approaches, research designs, and the population and sampling techniques. Finally, this section accounts for the data collection tools applied in this study, as well as the procedures for the data collection and analysis.

#### **3.1.1 Research Paradigms**

The process of undertaking research encompasses a logical and a systematic set of procedures that need clear justifications and a painstaking framework. In other words, research works need the underpinning of logical processes and philosophical beliefs. Johannesson and Perjons (2014) defined research paradigm as, “A set of commonly held beliefs and assumptions within a research community about ontological, epistemological, and methodological concerns.” (p. 167). It is therefore important to understand what each of these elements denotes in relation to research. Ontology is concerned with the philosophical beliefs about the nature of reality and

social phenomena. Ontological research questions address how things come to exist and the relations that connect them (Johannesson & Perjons, 2014). Epistemology in research manifests in knowledge. It is the foundation of knowledge, its essence, how it can be formed, acquired and how it can be communicated to others (Kivunja & Kuyini, 2017). The methodological aspect is related to questions about how to investigate, collect, and display information in an organised and logical manner.

There are three major paradigms that can be highlighted among others: the positivist paradigm, the constructivist paradigm, and the pragmatic paradigm. The positivist paradigm lies in the belief that social phenomena are ruled by laws that are best presented as facts through applying scientific and objective procedures (Abdul Rehman & Alharthi, 2016). The positivist framework heavily relies on quantitative measures and empirical investigation. Meanwhile, the constructivist paradigm also known as interpretive paradigm emerged as a response to the dominance of the positivist views. Constructivism refutes the idea that reality is binary and verifiable and that it requires the complete removal of oneself and his intuition. Constructivist research is more flexible as it allows for the inclusion of the researcher's perspective and interpretation. According to Grix (2004, as cited in Abdul Rehman & Alharthi, 2016), "Researchers are inextricably part of the social reality being researched, i.e. they are not 'detached' from the subject they are studying" (p. 55).

Furthermore, the main objective of the interpretivist paradigm is not to generalise or provide objective and context free knowledge, it strives to explain social phenomena as interpreted and depicted by individuals (Abdul Rehman & Alharthi, 2016). Thus, the constructivist framework makes use of qualitative data and applies inductive measures as opposed to deductive ones. The third paradigm is known as the pragmatist paradigm. The need



for new approaches that do not eliminate neither logical thinking nor the need for interpretation led to the emergence of the pragmatist philosophy. The pragmatic paradigm viewed the extreme nature of both the positivist and constructivist paradigms as debatable. This particular paradigm supports the practical way of making sense of the world and individuals by applying mixed methods (Kivunja & Kuyini, 2017). The main philosophical grounds of this paradigm are:

- Relationships in research are best determined by what the researcher deems appropriate to that particular study
- There is no single reality and all individuals have their own and unique interpretations of reality
- A combination of quantitative and qualitative research methods (Kivunja & Kuyini, 2017, p. 35).

It is important to know the philosophical beliefs and assumptions that each paradigm advocates in order to accurately choose the appropriate paradigm (Kivunja & Kuyini, 2017). It is essential in locating the stance of ones' research because the research work is bound by those assumptions and beliefs. Consequently, the present work follows the pragmatic paradigm simply because it allows multiple methods to be applied, methods that are most appropriate for the study as viewed by the researcher. The present study makes use of quantitative as well as qualitative measures for a more rounded understanding of the problem at hand. It is guided by logical facts as well as the researcher's own interpretation of the participants' responses.

### **3.1.2 Research Approaches**

Three methods are common among research works. Each research method differs in terms of the data collected and the methods used. Additionally, each paradigm applies the most suitable of these methods. These are quantitative, qualitative, and mixed-methods research

approaches. Generally, quantitative approaches use numerical data and employ experimental instruments, qualitative approaches deal with words and descriptions, and make use of observations and interpretations (Creswell, 2014). A mixed-methods approach as the name suggests combines elements of the two. A mixed-methods research includes gathering quantitative and qualitative data and using different designs that adhere to different philosophical beliefs in order to answer related questions (Creswell, 2014). As a result and due to the nature of the formulated research questions, the most appropriate research approach for this study is the mixed-methods approach. That is due to the need for quantitative as well as qualitative data and not one or the other.

### **3.1.3 Research Design**

The current study follows the case study research design while utilising a mixed-methods approach. It is a non-experimental research design where both qualitative and quantitative research methods are applied without an intervention or treatment of any kind. Hence, the main aim of this research is to gain better understanding of the problems hindering the development of collocational knowledge of students in our EFL context. In addition, it studies the possibilities of integrating DDL and the effect it may have on students' learning generally and their acquisition of collocations specifically.

According to Kumar (2011) a research design is, "A procedural plan that is adopted by the researcher to answer questions validly, objectively, accurately and economically" (p. 96). Thus, a research design is a practical method that allows the researcher to answer his/her research question(s) in an organised manner. Crowe et al. (2011) defined a case study design as, "A research approach that is used to generate an in-depth, multi-faceted understanding of a complex issue in its real-life context." (p. 1). This proves that it can be used in different research

studies for various aims and reasons. This particular research design is used in a variety of disciplines, namely social sciences (Crowe et al., 2011). Although the aim of each research differs generally, case study designs are used for exploratory research. Cook and Kamalodeen (2019) further defined it as, “An in-depth exploration from multiple perspectives of the complexity and uniqueness of a particular project, policy, institution, programme, or system in a real-life context.” (slide, 9). Indeed, case studies are often aimed at tackling new or under-explored phenomena to gain a better and a more inclusive understanding about its nature. That is why it tends to apply a variety of research methods and instruments. Case studies are non-experimental and are categorised as naturalistic i.e. a case study attempts to explore natural events without any form of intervention or manipulation of its elements (Crowe et al., 2011).

A common misconception is that case studies are purely qualitative. In fact, Yin (2003) accounted for the variations of case study research designs. He clarified that the case study design should not be mistaken for “qualitative research” (p. 14). He further explained that case studies can be built upon the combination of both qualitative and quantitative evidence (2003, p.15).

Based on the elements previously mentioned and due to the exploratory nature of the research work, the case study research design is viewed as the most suitable design for the present work.

### **3.1.4 Data Collection Methods**

Data collection methods also known as data collection tools are the instruments employed by the researcher in the course of his/her research journey in order to gather the necessary data that answer the formulated research questions. As previously mentioned, the study follows a mixed-methods approach, that is, it exploits a mixture of data collection tools. As far as this

research is concerned, three data collection methods were adapted. These tools are: a test for students, students' questionnaire, and teachers' interviews.

#### ***3.1.4.1 The Test***

One of the most popular data collection tools are tests. Tests are practical instruments that provide a wealth of numerical data. Cohen et al. (2007) highlighted the importance of tests in education research: "In tests, researchers have at their disposal a powerful method of data collection, an impressive array of tests for gathering data of a numerical rather than verbal kind." (p. 414). Although tests are common tools in experimental design, they can be used in descriptive designs as well. For example, a researcher can aim at assessing the learners' knowledge of an aspect in language for the sole purpose of describing and displaying the results.

Standardised tests are one of the most common types of tests. However, oftentimes standardised tests do not cover some particular aspects of language; they mainly cover language skills (reading, writing, listening, speaking) and they often account for grammar and vocabulary knowledge. Standardised tests are highly recommended to use since they discuss issues of validity and reliability, nevertheless; some aspects in language remain uncovered by this certain type of tests (Johnson & Christensen, 2019). While dealing with collocations, most researchers create their own tests according to their research objectives. Hence, this was the suitable choice for this research project. The test was not fully designed by the researcher nor was it created from scratch. It was adapted from McCarthy and O'dell's 2017 book *English collocations in use for intermediate learners*. The choice was primarily due to the issue of reliability and secondly due to the fact the book provides answer keys, which we followed in the analysis process.

**3.1.4.1.1 Structure and Aim.** The main and perhaps the sole aim of the test was to elicit the level of collocational knowledge of master one (M1) students. The test attempted to answer

research question number one in this study that is “What is the current level of collocational competence of first year master students?” The test was an achievement test that attempted to assess the overall knowledge of collocations that M1 students have. Achievement tests differ from performance tests in that performance tests measure “one's ability to use language for real-world purposes to accomplish real-world linguistic tasks across a wide range of topics and settings.” (*Understanding Language Proficiency*, n.d.). In contrast, achievement tests are typically designed to test an individual or learner’s knowledge of specific information (*Understanding Language Proficiency*, n.d.). In other words, what they know or do not know.

The test consisted of five tasks all adapted from McCarthy and O’dell’s book. Task one dealt with adverb+adjective collocations. They were provided six words in a box and asked to accurately chose which word best replaces the word “very” in the 10 corresponding phrases to form a more appropriate collocation. Task two was connected to task one. It entailed that they fill the blank space in each of the 10 sentences with the most accurate newly formed collocations from task one. Task three consisted of five different sets of collocations. In each set one collocation was incorrect and the participants had to identify and cross out the incorrect one successfully. Task four was in form of a paragraph which contained eight erroneous collocations. The participants had to accurately identify the eight incorrect ones as well as correct them. Task five was the last task and it focused on verb+noun collocations. It consisted of 10 sentences, each sentence had three options of verbs and the participants had to choose the one which best collocates with the noun in context.

**3.1.4.1.2 Piloting and Validation.** The piloting and validation of the research tools are highly important and recommended in undertaking any research. The two processes help in avoiding any major inconsistencies. In addition, they ensure the reliability and validity aspect of

research. After the test was designed, it was sent to three teachers of applied linguistics at Biskra University in order to validate and correct any major mistakes. An opinionnaire section was added for the teachers to answer. The questions were regarding the test's length, its appropriateness in terms of level of difficulty, and the instructions' clarity. All the responses were affirmative and no major remarks were made. However, one suggestion came up from a teacher, which we took into consideration. The suggestion was to add a small introduction that states the objectives of the test. It was in order "to reduce their level of anxiety". The piloting stage followed wherein seven M1 students were asked if they could sit for the test before their session, to which they consented. After they finished the test, they were orally asked if there are any ambiguities or difficult terminology. They all agreed that it was clear enough. Their only complaint was that task one was a bit challenging for them which could not be helped.

#### *3.1.4.2 Students' Questionnaire*

Brown (2001) defined questionnaires as "any written instruments that present respondents with a series of questions or statements to which they are to react either by writing out their answers or selecting from among existing answers." (as cited in Dörnyei & Taguchi, 2010, pp. 3-4). The students' answers and perspective provide integral data for answering the research questions. Therefore, it was important to design and administer a questionnaire for students for a more thorough comprehension of the research problem.

**3.1.4.2.1 Structure and Aim.** The chief aim of the questionnaire was to understand the students' perception of collocations and data-driven learning. More specifically, it strived to uncover whether M1 students were aware of the importance of collocations in language learning as well as if they were open-minded to the integration of DDL in our context. The students'

questionnaire was designed to answer a part of research question number three, which is “How do learners perceive data-driven learning?”

The questionnaire was semi-structured (containing both close-ended and open-ended items). It consisted of three sections that had an overall of 22 questions. The questionnaire also included an introductory section that defined a web corpus and assured the confidentiality of the respondents' information. Section one dealt with personal information about the respondents including their gender, age, and period of studying English as a foreign language. Additionally, section two included questions about collocations and collocational knowledge. Lastly, section three encompassed questions about data-driven learning and corpus linguistics. The questions in section three were asked implicitly i.e. they were questions that related to and revolved around DDL and corpora, this was due to the difficult nature of the topic and to maximise the opportunity of sound answers. For instance, since DDL relies heavily on internet-based activities, the students were asked about integrating internet-based methods inside the classroom and so on.

It is important to state that during the course of data analysis the researcher felt the need to add a question. This question was formulated while dealing with section two i.e. the section about collocations. After asking the respondents about their familiarity with collocations, this question emerged. It reads as follows: “were you familiar with ‘collocations’ before you learned it in the language mastery module or after?” The question was administered through Google forms.

**3.1.4.2.2 Piloting and Validation.** Three teachers of our department including the supervisor of this work validated the questionnaire. They were provided with an opinionnaire including questions about redundancy, ambiguity, the appropriateness of the response categories,

etc. Their responses were mostly positive and had no remarks expect for one teacher who pointed out question 13 in section three. The question used to read, “have you ever used corpus in a language classroom before?” The teacher recommended that we rephrase the question so that it applied more to students as opposed to teachers. We took that into consideration and rephrased the question to “Have you ever learned language explicitly through the use of a corpus in a classroom?”

Meanwhile, the piloting stage took place where seven M1 students were asked to answer the questionnaire, which they accepted. Thanks to the piloting process, we felt the need to provide an instruction section for students, since they appeared to confuse DDL with E-learning that is mainly done at home. In addition, we underlined and put certain words in bold in order for them to comprehend that it has to do with classroom internet-based learning.

#### ***3.1.4.3 Teachers' Interview***

Another common research tool that provides a wealth of qualitative data is the interview. According to Easwaramoorthy et al. (2006), an interview in research is, “A conversation for gathering information. A research interview involves an interviewer, who coordinates the process of the conversation and asks questions, and an interviewee, who responds to those questions.” (p. 1). It was entirely appropriate to conduct an interview with a number of applied linguistics teachers in order to benefit from their teaching experience and their views and beliefs on language teaching in our context.

**3.1.4.3.1 Structure and Aim.** The interview was essential as far as this research is concerned. First, the topic tackles issues related to students' awareness and competence and the teacher is the best observer inside the classroom. Second, due to the novice nature of the subject of DDL in our context, it was extremely preferable to consider the teachers' opinions. The



interview questions attempted to answer research question number two, a part of research question three, and research question four. These questions are respectively: What are the factors impeding students' learning of collocations? How do teachers perceive data-driven learning? To what degree is the application of data-driven Learning and a corpus-based approach achievable?

The interview was semi-structured with open-ended and close-ended questions. Even though it included a set of predetermined question, it was flexible allowing the researcher to probe at times. The interview consisted of 11 fixed questions that were asked to all of the interviewees. It was a face-to-face interview that was recorded after the teachers had consented.

Table 3.1 is displayed for a clear understanding of the aims and structure of the interview.

**Table 3.1**

*The content and aim of teachers' interview*

Section	questions	Content	Aim
Section one	1-2	General information	To gain information about the experience and background of the interviewees
Section two	3-7	Students' collocational knowledge and what may hinder it	To understand the extent to which teachers acknowledge the importance of collocations in students' language learning
Section three	8-11	The corpus and DDL approach in our context	The opinions of language teachers regarding implementing DDL and the factors that may impede its implementation

**3.1.4.3.2 Piloting and Validation.** The interview questions, like the other two data collection tools, were sent to three teachers via email in order to correct any major mistakes or inconsistencies. An opinionnaire was also linked to the interview questions (Appendix H). One teacher assisted us in rewording questions three and four for clearer understanding. Unfortunately, due to time constraints and the teachers' busy schedules it was not possible to pilot the interview it was simply validated and directly conducted with the participants.

### **3.1.5 Data Collection Procedures**

Taking the issues of ethical considerations that any body of research necessitates into account, the participants of this study were informed of what the research entails. Furthermore, they were assured that their data, personal information, and opinions would be treated with the utmost confidentiality and respect. All of the participants in this study had consented to taking part of the process. The teachers signed the informed consent form before the interview took place and was recorded. While students orally agreed to answering the questionnaire and the test.

The process of collecting data was done concurrently. In other words, the data from all three collection methods were gathered at the same time period of one week. Thanks to the help of our supervisor, Master one students of two groups sat for the test and answered the questionnaire in two of their research methodology sessions. Each group had a total of one hour to finish both the test and questionnaire. Meanwhile, four interviews with teachers were arranged according to the availability of teachers and the time they viewed as appropriate. Moreover, all the contacted teachers showed incredible support and interest in partaking in the interview. The four interviews were done in the same week during different days that were set by the teachers. The interviews concluded the data collection procedures for this study.

### **3.1.6 Data Analysis Procedures**

Since the current study follows a mixed-methods research approach, the data gathered by nature were of quantitative and qualitative type. The quantitative part included close-ended items from the questionnaire and the test results. The qualitative part involved open-ended questions mainly from the interviews.

The different kind of data necessitates different analysis procedures to be done. For the most part tests provide quantitative evidence. However, the test, adapted in this study does not provide scores or values but rather frequencies and percentages. Thus, the most appropriate analysis procedure in this case is descriptive statistics. Basic descriptive statistics is also used to analyse and display numerical data collected from the close-ended items in the questionnaire. The data is displayed in form of tables, charts, and graphs for clear understanding.

While dealing with the qualitative part of the data, content analysis was the most suitable option for analysis. Content analysis in its simplest refers to the set of steps a researcher undertakes in order to assign meaning to their written data (Cohen et al., 2007). Content analysis is a mechanical and systematic process that involves the researcher assigning codes in forms of numbers or colours to different segments of the transcripts or writing (Dawson, 2009). This is done in order to identify common themes or trends that may lead to the emergence of various categories as well. Consequentially, we followed the steps of coding and categorising considering that the interview provided rich and varied qualitative data and content.

### **3.1.7 Population and Sampling**

The target population of this research work was the first year master students of applied linguistics at Mohamed Kheider University in Biskra as well as the applied linguistics teachers of the English department.

Convenience sampling is the non-probability, non-random sampling technique where the researcher selects participants that are “readily and easily available” (Taherdoost, 2016, p.22). The main criteria for the selection of participants in this technique are: the availability of participants, accessibility, geographical proximity, and the willingness to participate (Etikan et al., 2016). That is why the non-probability convenience sampling technique was used in the present research. Consequently, 37 master one students comprised the sample in addition to four EFL teachers. The choice came as a result of the current circumstance and because the aim was not to generalise, but rather to gain insights from a sample of participants which were conveniently available. Both the students and the teachers provided valuable input and information to the inquiry, as they were primary references and sources of relevant data.

### **3.2 Results of the Study**

This section aims at analysing and interpreting the collected data from the three data gathering instruments that consisted of a test and a questionnaire designed for students, as well as the teachers’ interviews. As mentioned earlier, the researcher resorted to content analysis and basic descriptive statistics for the analysis and display of the data. The analysis was done manually by the researcher.

### 3.2.1 Results of the Test

#### Task one

**Table 3.2**

*Results of task one*

	Number of answers	percentage
Correct answers	134	45%
Incorrect answers	160	54%
Unanswered	2	1%
Total	296	100%

From the table above it can be concluded that more than half of the answers of the sample at hand (54%) were incorrect collocations, specifically in this case adverb+adjective collocations. However, the analysis process showed that a reasonable number of the participants have good command over collocations (45% accurate answers). The participants particularly displayed their command in successfully identifying “bitterly disappointing”, “highly successful”, and often “ridiculously easy”. It is necessary to mention that during the analysis procedure, an error was detected. Due to a problem in the Word document the word “deeply” did not appear alongside the other suggested words in the box (see Appendix B). Therefore, it was only appropriate to remove the two sentences that involved its use in tasks one and two. In other words, the collocations (task one) and sentences (task two) that included it were dismissed for fair assessment.

**Task two****Table 3.3***Results of task two*

	Number of answers	percentage
Correct answers	95	32%
Incorrect answers	172	58%
Unanswered	29	10%
Total	296	100%

In task two, some students identified only one part of the collocation correctly, mainly the adjective part but not the adverb part, but since collocations is about the frequency of co-occurrence of both of those items it was counted as incorrect. Besides, we followed the answer key provided by the scholars themselves. Furthermore, some students used the incorrect grammatical form (e.g. bitterly disappointed instead of bitterly disappointing); however, in this case, it was counted as correct because we are judging the ability of making word collocations, and not judging the grammatical form. Even though the context is provided, some students missed the entire meaning and selected both of the wrong word combinations.

**Task three****Table 3.4***Results of task three*

	Number of answers	percentage
Correct answers	60	32%
Incorrect answers	122	66%
Unanswered	3	2%
Total	185	100%

In task three, students were given five rows of collocations. Each row contained a set of four collocations with one erroneous collocation, which they were asked to identify. As table 3.4 displays only 32% of their answers were accurate.

**Task four****Table 3.5***Results of task four*

	Number of answers	percentage
Identified + accurately corrected	43	14%
Identified + Inaccurately corrected	35	12%
Identified+ uncorrected	40	14%
Identified (corrected or not)	118	40%
Unidentified	138	47%
Uncorrected (identified and not corrected+ unidentified)	178	60%
Unanswered (ignored completely)	40	13%
Sum of answers	296	100%

Task four was in the form of a small paragraph. The participants were asked to identify the eight incorrect collocations in the written passage by underlining them. They were also asked to correct those inaccurate collocations that they found. The analysis of this task was rather meticulous and quite time-consuming, since the instructions were of two folds. As shown in the table above, participants either: identified the inaccurate collocations successfully and did not correct them at all (14%), identified the inaccurate collocations and corrected them successfully (14%), identified the inaccurate collocations and failed to accurately correct them (12%), neither



identified them nor corrected them (47%). In addition, some participants appeared to be intimidated by the paragraph form and completely ignored it. In order to be more consistent with the other tasks in the test and since a large number of students did not correct the collocations they identified; it was viewed as appropriate to follow the following table:

**Table 3.6**

*Results of task four (simplified)*

	Number of answers	percentage
Identified (corrected or not)	118	40%
Unidentified	138	47%
Unanswered	40	13%
Sum of answers	296	100%

The above table is a simplified version of the table before. The main aim of this task was therefore to test the students' ability to identify incorrect collocations within a written passage. In that case, only 40% of the answers of the 37 participants were accurate identifications of the erroneous collocations in the paragraph while 47% remained unidentified. Meanwhile, 13% were ignored completely.

**Task five****Table 3.7***Results of task five*

	Number of answers	percentage
Correct answers	164	44.32%
Incorrect answers	179	48.37%
Unanswered	27	7.29%
Total	370	100%

Task five concluded the test. The task consisted of 10 sentences and participants were asked to choose one verb from three options to form a sound verb+noun collocation in that context. Here, some students selected more than one answer, in order to be objective it was counted as unanswered. In task five, the students struggled the most with the verb “paid”. They also had issues with “make” and “have”. Overall, the participants did reasonably well in this task especially with collocations such as “pay attention” and “had a dream”.

**The overall test****Table 3.8** *The overall results of the test*

	Number of answers	percentage
Correct/identified	571	40%
Incorrect/unidentified	771	53%
Unanswered	101	7%
Total	1443	100%

The results of the entire test suggest that the level of collocational knowledge of the 37 participants in this study is below average (40%). Moreover, the test was designed for intermediate level students. Thus, this test proved that there is a need for students especially at the master level to learn more about how collocations work in natural language.

**3.2.2 Results of the Students' Questionnaire**

Section one: personal information

## 1. Gender

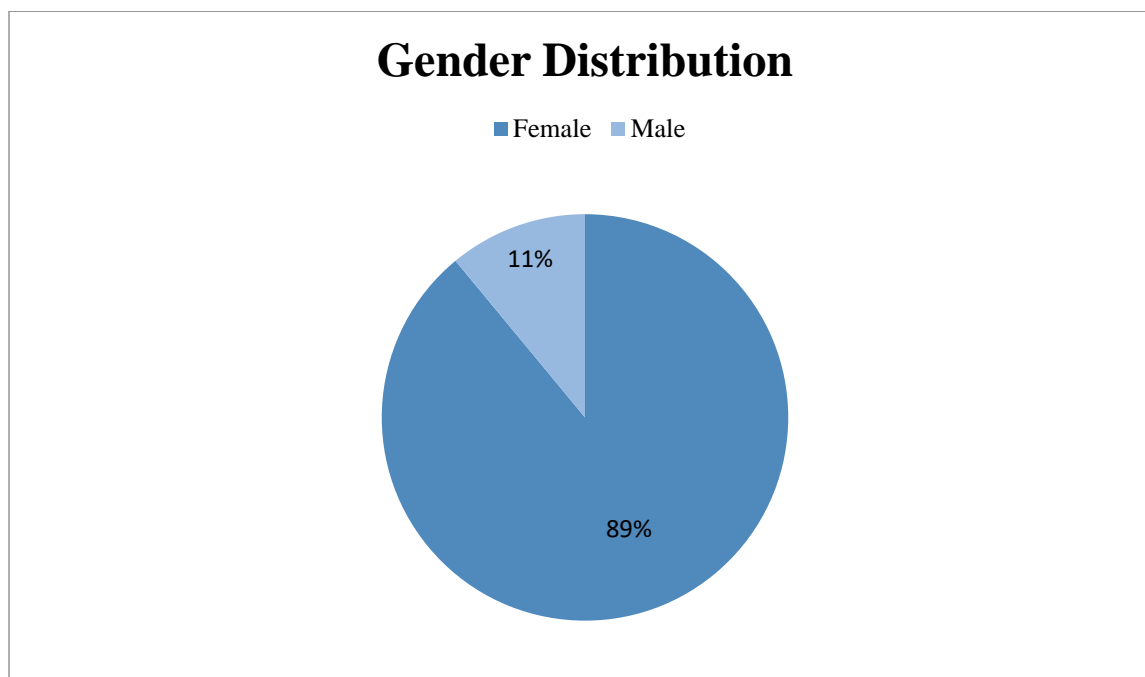
**Table 3.9**

*The respondents' gender*

Option	Number	Percentage
Male	4	11%
Female	33	89%
Total	37	100%

**Figure 3.1**

*The gender distribution of the sample*

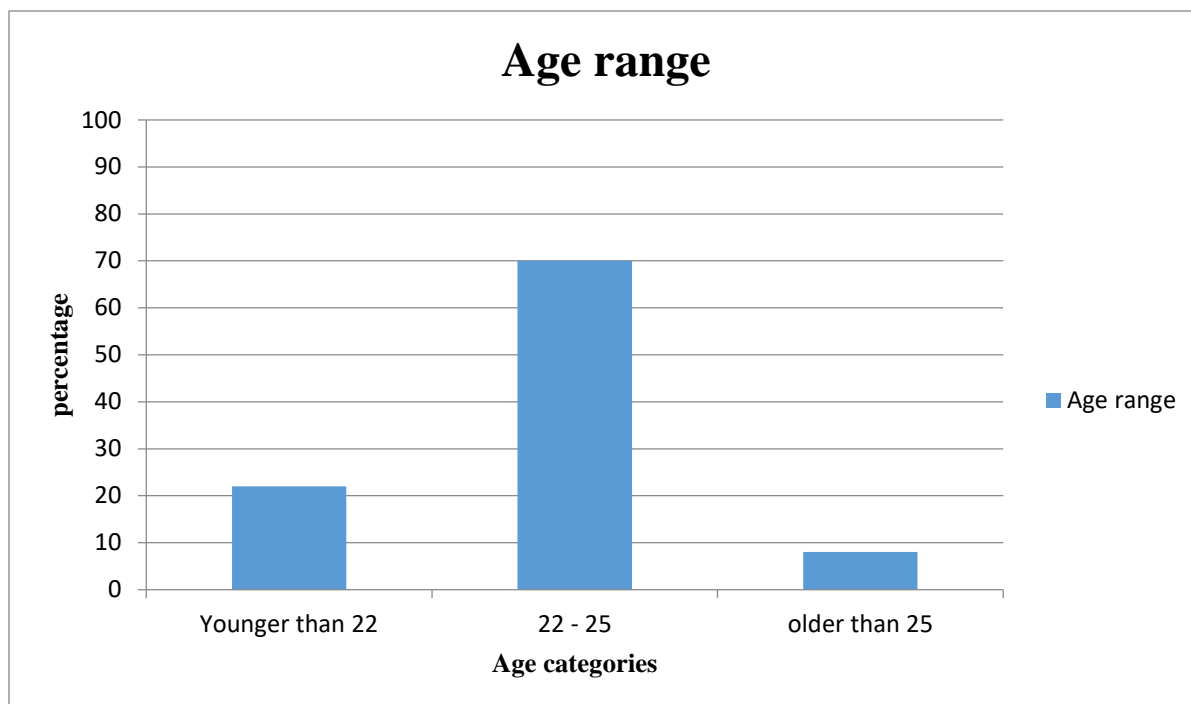


More often than not research surveys and questionnaires will include a general information section. Although questions of gender and age are not integral to the research at hand, they were added in order to better identify and categorise the target sample we are dealing with. As figure 3.1 shows, the percentage of female respondents (89%) dominated that of the male one (11%). This significance may be due to the belief that females are generally more drawn towards language learning and education compared to the other gender.

## 2. Age

**Table 3.10***The respondents' age*

Age	Number	Percentage
younger than 22	8	22%
22 - 25	26	70%
older than 25	3	8%
Total	37	100%

**Figure 3.2***The age range of the sample*

As it was previously mentioned, this question mainly aimed at identifying the age demographic of our sample. The ages of the respondents were between 21 and 37 years old. It shows that master one students are rather diverse in the age category and that higher education is

not only exclusive to younger students. As the table and graph above present, 70% of the respondents were between the ages of 22 and 25.

3. How long have you been studying the English language overall?

**Table 3.11**

*Respondents' years studying English*

Years	Number	Percentage
4 years (university)	12	32%
5 years	3	8%
7 years (since high school)	2	5%
10 to 11 years (since middle school)	18	49%
More than 11 years	1	3%
Left blank	1	3%
Total	37	100%

This question was also open-ended hence, some students only counted their years studying English at the university and not the years before. This question was asked in order to know the extent to which the respondents were exposed to the English language throughout the years in order to link it to their collocational knowledge so far. In a general sense, Algerian EFL learners study the English language since the middle school. It is a part of the educational curriculum for middle school and high school students that enrol for public education. There can be some expectations of course. However, most of the respondents (49%) stated that they were studying English for 10 years or longer. This shows that the respondents have been long exposed to the English language, which can attest to their knowledge of its intricacies.

#### Section Two: collocations

1. Are you familiar with collocations?

**Table 3.12***Respondents' familiarity with collocations*

Option	Number	Percentage
Yes	26	70%
No	11	30%
Total	37	100%

The table demonstrates that 70% of the total number of respondents are/were familiar with collocations. It is a rather significant number indeed, whereas 30% of them also claimed that they were not in fact familiar with the concept of collocations. The respondents were asked to further provide a brief definition in the case they answered with yes. It was in order to assess the level of their knowledge and if they can provide a close or accurate definition of collocations based on prior knowledge. Some of the respondents' definitions were quite accurate regardless of the grammatical errors. For example one definition was: "a pair of words that suit each other, or give better meaning when are together". Another respondent wrote: "the linking between two words to give a specific meaning (two word side by side)". It is safe to say that most of the respondents showed that they know what collocations represent in language. The familiarity of the M1 sample with collocations may be because they learned it in the first semester specifically in the language mastery module.

At this point and precisely, while dealing with this question, the researcher felt the need to clarify whether M1 students knew collocations before they were exposed to it in the language mastery module. Thus, an online questionnaire that consisted only of one question was created in Google forms and distributed via Facebook. Thirty seven M1 students were kind enough to

respond and the form was then closed. The question will be added in this section for a more consistent flow.

2. Were you familiar with collocations before you learned it in the language mastery module or after?

**Table 3.13**

*Respondents' exposure to collocations*

Option	Number	Percentage
Before	19	51.40%
After	18	48.64%
Total	37	100%

The results obtained from this question indicated that a significant number of respondents (49%) did not know about collocations until they were taught in the language mastery course. However, it is also noticed that a slightly larger percentage (51%) did in fact know about collocations before the lessons of language mastery. The results further showed that this question needed to be added to the questionnaire.

3. Do you view learning collocations as important in language learning?

**Table 3.14**

*Respondents' views on collocation learning*

Option	Number	Percentage
Yes	32	86%
No	4	11%
Unanswered	1	3%
Total	37	100%



According to the obtained results, 86% of the sample agreed that collocations are important in language learning while 11% opposed. This further proves that these students are aware of the importance of collocations and the necessity to learn it. In order to further investigate the reasons why they thought it was important (or not), respondents were asked to justify their answers. It can be noted that since the questionnaire was handed to the respondents i.e. that it was a paper and pen questionnaire, a large number of respondents typically ignored the justification questions. Nevertheless, some of the justifications behind why students view collocations as important are worth noting. A respondent wrote, “to improve our writing skill /to sound more like native”. Meanwhile another one justified, “one of the objectives of learning a certain language is to communicate with natives and sound more natural when talking to them. Learning collocations can help achieve this goal.”

4. Have you faced any difficulties while learning collocations in the Language Mastery module in semester one of this academic year?

**Table 3.15**

*Difficulties concerning collocations*

Option	Number	Percentage
Yes	33	89%
No	4	11%
Total	37	100%

The question aimed at discovering the respondents' experience with learning collocations in the language mastery module. This question attempted to understand if the learners in the sample had a hard time learning such a concept or not. As a result, 89% of the sample argued that they did in fact face some difficulties while dealing with collocations, specifically in the

language mastery course. The remaining 11% of the respondents denied facing any difficulties in that vein.

5. How would you rate the difficulty of learning collocations?

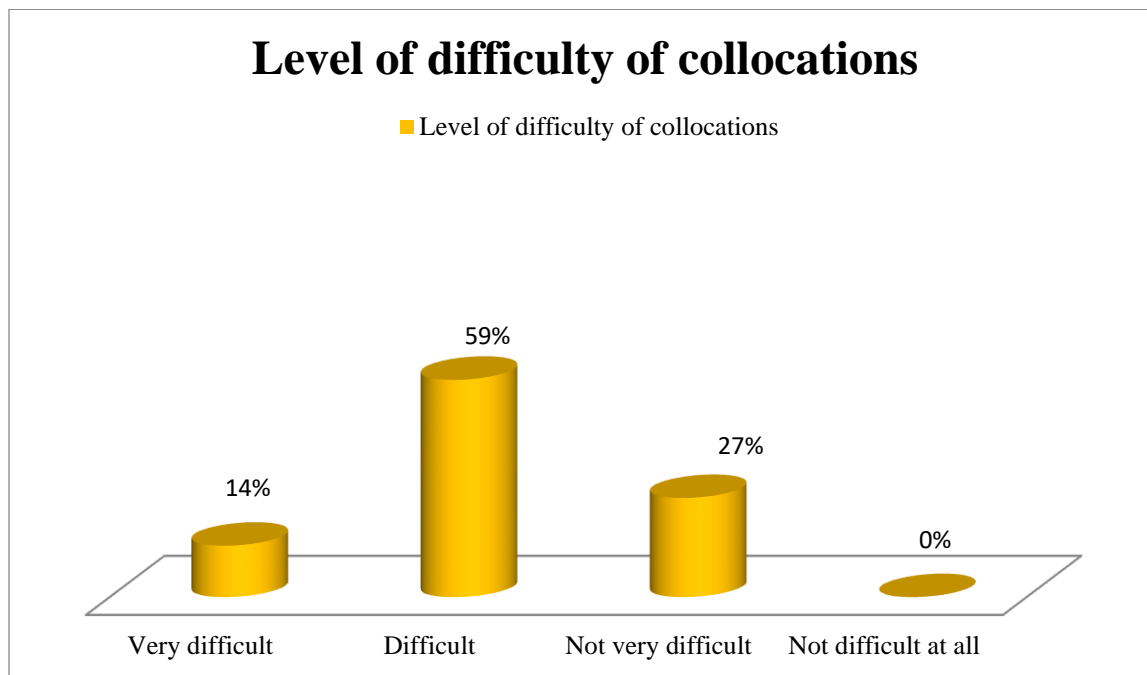
**Table 3.16**

*The level of difficulty of collocations*

Option	Number	Percentage
Very difficult	5	14%
Difficult	22	59%
Not very difficult	10	27%
Not difficult at all	00	00%
Total	37	100%

**Figure 3.3**

*The level of difficulty of collocations*



This question follows through with the preceding question. It aimed at uncovering the level of difficulty of collocations according to the respondents themselves. Notably, 59% of the students in the sample stated that it was “difficult”, followed by a 27% that argued that it is “not very difficult”. Moreover, 14% of the respondents reported that learning collocations was very difficult for them. On the other hand, no one in the sample reported that it was not difficult at all. Overall, learning collocations seemed to pose a challenge for M1 students as far as this sample is concerned.

6. To what degree do you agree with the following statements?

**Table 3.17***The agreement of respondents towards the statements*

Statement	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
a) Learning English collocations will make you sound more fluent and natural while speaking.	57%	35%	5%	3%	0%
b) Learning English collocations will help you express yourself in various ways.	54%	35%	8%	3%	0%
c) Learning collocations can improve your writing skill.	43%	35%	19%	3%	0%
d) Collocations can make you sound more native-like in conversations.	46%	46%	8%	0%	0%

The above table shows the percentage of agreement of the respondents to each statement. The purpose of this question was to elicit the opinions of the sample about each of the statements provided. The statements were also derived from McCarthy and O'dell's book about collocations. They are statements that are commonly associated with collocations (statements

about fluency in speech, accuracy in writing). As it is shown, in statement a 57% strongly agreed that learning collocations will make learners sound more fluent and natural, followed by 35% that agreed to the statement. Overall, 92% agreed to statement a. Meanwhile, 5% remained neutral while 3% disagreed. Regarding statement b, 54% of the sample strongly agreed that collocations help in self-expression in various ways while 35% agreed. Nevertheless, 8% also reacted with “neutral” to the statement, as 3% disagreed.

Statement c entailed that learning collocations can help with developing the writing skill, to which 78% of the respondents agreed (43% strongly agreed and 35% agreed). Nineteen per cent of the respondents remained neutral to this statement and 3% disagreed with it. The last statement, statement d read that “collocations can make you sound more native-like in conversations”. Forty six per cent of the students in the sample strongly agreed to that statement. Equally, 46% also agreed. In this case 8% remained neutral and 0% disagreed. Overall, the results showed that the majority of respondents either strongly agreed or agreed to each statement. As a result, students in the sample seem to realise the benefits learning collocations yields.

### Section three: online corpus and data-driven learning

1. Which of the following materials are used by your language teachers in the classroom?

(Choose an answer for each method)

**Table 3.18***The materials used by teachers*

Materials	Never	Rarely	Sometimes	Often	Always	Unanswered
a) Data shows	57%	32%	8%	3%	00%	00%
b) Computers	59%	22%	13%	3%	3%	00%
c) Language corpus	27.02%	8.10%	27.02%	13.51%	5.40%	18.91%
d) Audio visual aids	46%	19%	24%	8%	3%	00%
e) Hand-outs and worksheets	0%	0%	8.10%	32.43%	59.45%	00%

This question was the first question in section three that deals with questions about corpus methods in teaching and data-driven learning. This question in particular attempted to acknowledge the varied teaching materials or techniques teachers use in the EFL classroom in order to aid students' learning. The option of language corpus was added provided that it was explained in the introduction part of the questionnaire. According to the respondents' answers, the main materials that teachers use are hand-outs and worksheets. A striking majority of English language teachers use hand-outs and paper worksheets either always (59%), often (32%) or

sometimes (8%). In comparison, teachers either never (57%), rarely (32%), or sometimes (8%) utilise data-shows to aid their teaching. Moreover, as far as computers are used inside the classroom 59% of the respondents stated that they are never used, while 22% chose rarely and 13% said sometimes. Forty six per cent of the responses also showed that audio-visual aids are never used inside the classroom. Some respondents stated that they are rarely (19%) or sometimes (24%) used.

From the answers it can be interpreted that the respondents do not fully grasp the concept of learning through corpus which they cannot be blamed for. Hence, there were a percentage of unanswered questions (19%) or the respondents may have randomly ticked an option. This is due to the random choices like ticking always or often while in reality it has not been used at all as far as data-driven learning is concerned. A percentage of the respondents however ticked never (27%) this could be because they fully read the introduction and understood what using an online language corpus entails. To sum up, this question strived to review the diversity of the teaching aids that EFL teachers use. The results from the responses showed that the majority of teachers in our context opt for hand-outs as means of transmitting information to learners. Furthermore, there is a lack in using diverse materials such as ICTs (information and communications technology) inside the language classroom.

2. What do you normally do when you meet a word that is new to you?

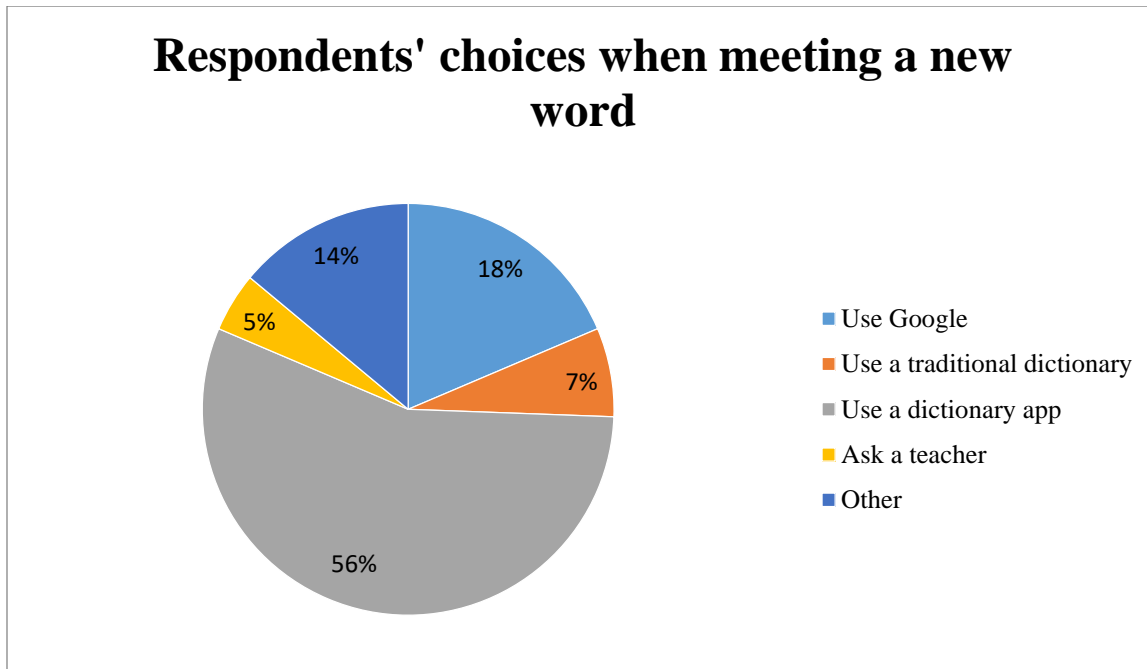
**Table 3.19**

*Respondents' choice when meeting a new word*

Option	Frequency of choice	Percentage
Use Google and search for its meaning	8	18.6%
Use a traditional dictionary	3	6.98%
Use a dictionary application on your phone	24	55.81%
Ask a teacher	2	4.65%
Other	6	13.95%
Total	43	100%

**Figure 3.4**

*Respondents' choices when meeting a new word*





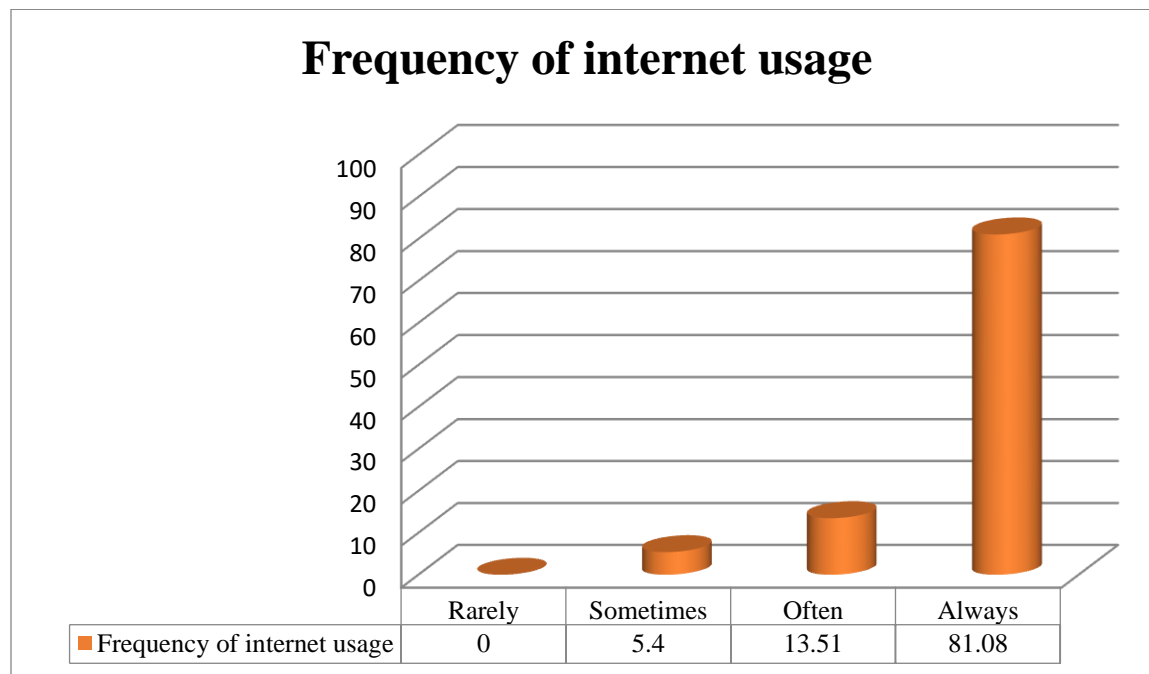
This was asked for the better understanding of students' familiarity with technology-based applications or methods. Since the data-driven approach requires learners to access and possibly learn through software; applications; and the internet, we sought to understand whether the students in the sample already use technological ways of learning. The results obtained showed that 56% of the responses included using a dictionary application as a way of searching new words. This question also required a justification, some of the justification of using a mobile app include, "I think that the dictionary application is more helpful and useful", "I use electronic dictionary is quick and give me synonyms and examples to the word".

Overall, students seemed to appreciate the convenience and content provided by the mobile application compared to a traditional dictionary. Furthermore, 19% of the answers were in favour of "googling" an unknown word to get its meaning. Some of the justifications also involved convenience of time and precision of information. Some respondents argued that they often resort to guessing the meaning from context as a strategy because according to them it also builds vocabulary. Conveniently, corpus software and programmes are a mixture of all those elements stated before. It provides context, synonyms and definition in a clear manner for learners to access.

3. How often do you refer to the internet while doing research or in your studies (at home)?

**Table 3.20***The frequency of consulting the internet*

Option	Number	Percentage
Rarely	0	0%
Sometimes	2	5.40%
Often	5	13.51%
Always	30	81.08%
Total	37	100%

**Figure 3.5***How often the respondents consult the internet*

Again, owing to the fact that DDL highly involves the use of the internet, this question was asked in order to realise the sample's familiarity with internet use. Therefore, it was appropriate to ask the extent to which respondents consulted the internet for learning or research purposes. An overwhelming 81% of the respondents answered that they always consult the

internet, 13% answered with often, while 5% replied with sometimes. Indeed, this generation of language learners or even learners in general references the internet more than any other generation. It further questions the lack of technological materials and equipment in our department and our learning context.

4. Do you prefer:

**Table 3.21**

*Teacher-centred or learner-centred classroom*

Option	Number	Percentage
A teacher-centred classroom	21	56.75%
A student-centred classroom	12	32.43%
Both	3	8.10%
Unanswered	1	2.70%
Total	37	100%

The DDL approach is strongly student-centred. Thus, it was only appropriate to consider this and ask the learners what they preferred. As the results provide, nearly 57% of the respondents in the sample favour a more teacher-centred approach towards teaching. This may be due to issues of anxiety or lack of interest. It could also be related to the belief that teaching needs to be mostly done by the instructor him/herself and that the learner is just a receiver of information. Students were also asked to provide explanations for their choices. A respondent wrote: “because the learners are not required with sufficient information like the teachers”

Overall, from their justifications, it appears that the respondents that favour a teacher-centred classroom believe that the teacher is superior in terms of knowledge and control.

However, 32% of the respondents opted for a student-centred classroom, some of their justifications are as follows: “it gives us opportunities to share our thoughts and ideas”, “I chose the second (student-centred) because I believe that it gives the opportunity to the student to be more active and practice the language more.”. The supporters of the student-centred approach favour discussion and a more independent learning experience. Another category (mainly three respondents) suggested that there should be a mixture of both methods inside the classroom as a collaboration of efforts.

5. Do you prefer learning through:

**Table 3.22**

*Respondents' preference between authentic and adapted materials*

Option	Number	Percentage
Authentic materials	31	84%
Adapted materials	5	13%
Unanswered	1	3%
Total	37	100%

Owing to the fact that DDL supports the use of authentic materials in language learning to gain native-like fluency and accuracy, this question felt relevant to ask. A very significant number of respondents (84%) chose learning through authentic materials over adapted ones (13%). The dilemma of using authentic vs. non-authentic materials has been commonly discussed in the language teaching and learning situation. Although both options contain pros and cons, there is strong evidence that supports the use of authentic materials especially for more profound learning. For instance, Flowerdew and Peacock (as cited in Benavent & Peñamaría, 2011) argued that non-authentic materials fail to account for natural, real-world language use and that “simplified materials often lose some meaning and the real-world situations the learners will

face are best prepared for with authentic texts.” (p. 90). However, it is important to note that the selection of authentic materials must be done carefully according to the needs of the learners.

6. How often do you find yourself losing interest in learning?

**Table 3.23**

*How often do respondents' lose interest?*

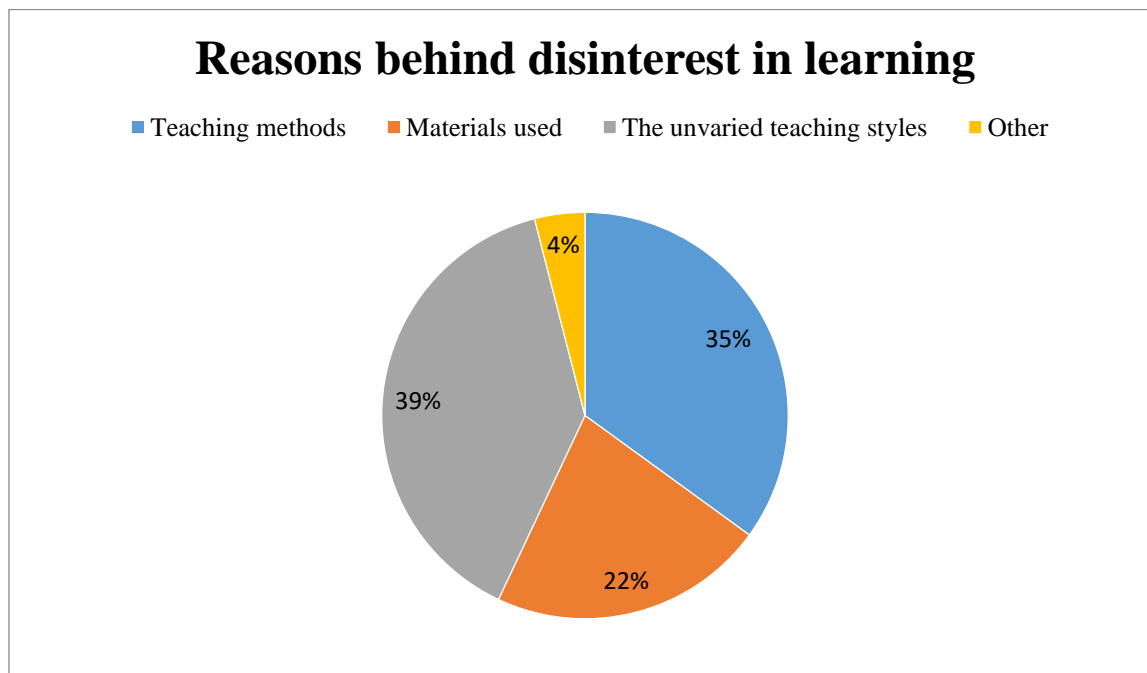
Option	Number	Percentage
Rarely	7	18.91%
Sometimes	22	59.45%
Often	6	16.21%
Always	2	5.40%
Total	37	100%

Moving from experience, the learning process can be rather challenging at times. Therefore, this question aimed at recognising the frequency of disinterest in the language learning process of the respondents. This question is also connected to the coming question relating to the different factors that may lead to learners' disinterest. As the table above displays, 59% of the respondents stated that they sometimes lose interest in learning due to different reasons. In addition, 19% selected rarely as their answer while 16% selected often. Five per cent of the respondents chose always as a response to the question. This indicates that learners sometimes in fact lose interest in learning, and that it is a different experience for every learner. Some students can be highly inclined to learn while others simply are not, depending on a variety of factors.

7. Do you find yourself losing interest in the learning process because of:

**Table 3.24***Reasons behind loss of interest*

Option	frequency	Percentage
Teaching methods	16	35%
Materials used	10	22%
The unvaried teaching styles	18	39%
Other	2	4%
Total	46	100%

**Figure 3.6***Reasons behind lack of interest in learning*

The figure above displays that 35% of the choices of this multiple choice question involve the teaching methods as a factor that leads to disinterest. Another percentage of answers (22%) indicate that the respondents also chose the teaching materials as an impeding factor of learning. Meanwhile, the majority of the answers (39%) accounted for the unvaried teaching

styles that teachers adapt. The teaching styles refer to how teachers decide to present, organize, and monitor the sessions for which they are responsible. Logically, it is impossible to make general assumptions about teaching styles, because the choice of the style depends on various reasons including teacher's personality, the content of the course and even the learners themselves. However, it can be observed that a significant number of instructors specifically in our context favour the directing teaching style. It is the teaching style where teachers promote listening and following directions (Thornton, 2013). Learning is usually in form of oral presentation of lessons from the part of the teacher and taking notes and answering questions from the part of learner.

This question included a justification part as well. One respondent wrote regarding the teaching styles, "the lack of variety in teaching styles makes me bored easily" Another respondent also argued, "The routine, the teacher explain and students take notes"; whereas, another responded regarding the materials used, "Hand-outs are boring" and another agreed stating that, "The used materials are very traditional and no longer effective". Overall, the students in this sample agreed that the traditional teaching methods, materials used, and the unvaried teaching styles play a major role in the aspect of language learning. Some respondents mentioned that the cause of disinterest in learning has to do with the lack of motivation inside the classroom as well.

8. Would you prefer learning through the internet i.e. using internet websites in the classroom?

**Table 3.25***Respondents' views regarding implementing internet-based methods*

Option	Number	Percentage
Yes	26	70
No	10	27
Unanswered	1	3
Total	37	100%

This question is strongly related to the potential of implementing DDL as a teaching and learning method. It aimed at eliciting the respondents' opinions regarding integrating internet-based teaching methodology such as DDL. It can be deduced from the above table that 70% of the sample are open-minded when it comes to implementing such new technology for learning purposes. They were asked to provide a justification to their stance, "strongly helpful and very easy" one responder wrote. Another precise justification was that "the internet is a wide platform and contains various teaching materials" as well as "because sometimes we need to take an idea about something within the context during session". For the most part respondents that agreed to such an approach were appreciative of the amount of information that they could access as well as the ability to access knowledge first hand.

On the other end, respondents who were against this approach (not DDL in particular) constituted 27% of the sample. Their justifications mainly included that, "the session is so limited and the internet may guide to losing attention" and similar answers due to time constraints, and the fact that students may lose focus during sessions because of the internet.

9. Do you think your willingness to learn will improve if teaching (inside the classroom) was through internet access?



**Table 3.26***The respondents' willingness to learn through internet-based methods*

Option	Number	Percentage
Yes	22	59.45%
No	10	27.02%
Unanswered	5	13.51%
Total	37	100%

This was asked to understand whether learners think their willingness to learn will improve if they learned through the internet. It is noticed that 59% of the sample believed that integrating web-based teaching methods such as DDL would improve their inclination to learn the language. Meanwhile, 27% argued against that.

10. Do you believe that there should be teaching methods that allow students to use the internet inside the classrooms?

**Table 3.27***Respondents' beliefs towards learning via internet*

Option	Number	Percentage
Yes	31	84%
No	5	13%
Unanswered	1	3%
Total	37	100%

This served as a follow-up question to the previous one. This is a direct question in which students can state whether they agree with the incorporation of a method such as DDL in their language classroom. The ones that support internet-based learning and in this case DDL (84%) provided some clarifications for their belief. One of the responses was, "The teacher's

information and explanation is not enough, sometimes students do not understand something and they feel ashamed to ask the teacher”. Moreover, one responder wrote, “The use of the internet inside the classroom will make the teaching and the learning processes very easy”. Another response also read, “To improve the way of looking for data through the internet”. Indeed, the language learner at the master stage is also a researcher who needs to acquire research and critical thinking skills. The opposing part of the sample (13%) justified their answers by stating fears regarding implementing web-based approaches. Most of their answers state that the learners will use the internet for personal enjoyment rather than actual learning. However, that can be solved through monitoring and supervision.

11. Are you familiar with corpus linguistics?

**Table 3.28**

*Respondents' familiarity with corpus linguistics*

Option	Number	Percentage
Yes	8	22%
No	27	73%
Unanswered	2	5%
Total	37	100%

This question was asked since DDL is strongly linked to learning through the corpus approach. It was in order to investigate the learners' familiarity with this aspect of linguistics. Corpus linguistics is a lesson in the master one syllabus; however, and as displayed by the table 73% declared their unfamiliarity with the concept. Reasonably, 22% of the sample stated that they were familiar with CL. The difference in percentages proves that there is a need in making learners aware of the importance and benefits of corpus especially in learning. For better clarity of the results, the respondents who chose yes as answer were asked to provide a brief definition

of CL. Only a few answered this question in particular, for example one responder wrote, “A subfield in applied linguistics that studies the language of native speakers”. Another one defined it as “A methodology to obtain and analyse the learning data” the few who answered had quite relevant definitions of corpus linguistics which proves they indeed studied this subject matter and were exposed to it, theoretically that is.

12. Did your teachers ever introduce the concept of language corpora?

**Table 3.29**

*Did teachers introduce the concept of corpora?*

Option	Number	Percentage
Yes	8	22%
No	25	67%
No answer	4	11%
Total	37	100%

This question aimed at explicitly asking the respondents if they necessarily were exposed to corpus content in any shape. As a result, 67% of the sample argued that they were not in fact introduced to CL, whereas 22% stated that they have and 11% chose not to answer. The students that answered with yes were asked to clarify the ways and instances that their teachers introduced corpora. The answers were interesting to say the least. One responder wrote, “I do not really remember” another similarly stated “He did actually but I forgot”. These students among others are proof that the teaching methods need to be more inclusive and interesting in order to help information stick to the minds of learners.

Nevertheless, two more respondents replied to this section of the question. One respondent wrote that s/he was introduced to corpora in form of tweets, “Tweets of politicians (trump)” however, s/he did not specify in what module or provide further detail. Another one

noted that s/he was introduced to CL in the applied linguistics module, “In applied linguistics as a tool of teaching “method” not as a theory”. The results of this question prove that learners need to be provided with better examples and method for learning to be affective.

13. Have you ever learned language explicitly through the use of a corpus in a classroom?

**Table 3.30**

*Instances where respondents learned via corpus*

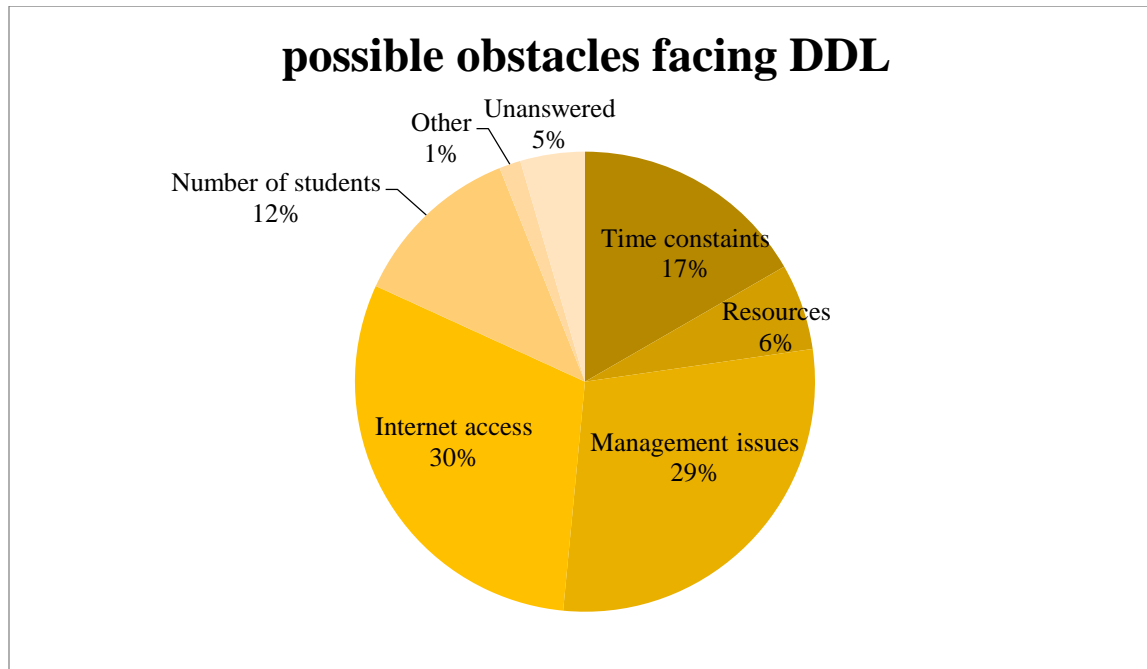
Option	Number	Percentage
Yes	5	13%
No	24	65%
No answer	8	22%
Total	37	100%

Again, moving from the fact that the concept of using an online corpus was explained in the introduction of the questionnaire, this question was asked. Thirteen per cent of the sample answered with affirmation to this question while a more significant percentage of the sample (65%) replied with negation. Indeed, certainly, this approach has not been applied, one indicator may be the unfamiliarity of students with this approach.

14. What do you think are the reasons for not using internet based programmes or corpus in our EFL context?

**Table 3.31***The reasons behind not using corpus programmes*

Option	Frequency of choice	Percentage
Time constraints	11	16.66%
Resources	4	6.06%
Management issues	19	28.78%
Internet access	20	30.30%
Number of students	8	12.12%
Other	1	1.51%
Unanswered	3	4.54%
Total	66	100%

**Figure 3.7***Possible obstacles facing the integration of DDL*

This serves as the last question in the questionnaire. It was asked in order to gain insights from the perspective of respondents on what could be the factors that may prevent the

implementation of new and advanced approaches. According to the answers obtained the major impeding factor is internet access with 30% choice frequency followed by management issues (29%), time constraints (17%), the number of students (12%) and then resources (6%). One other obstacle was stated by one respondent which is the lack of motivation of teachers to adapt new approaches which is quite interesting.

### 3.2.3 Results of the Teachers' Interviews

**Q1.** How long have you been teaching at the university?

**Table 3.32**

*Teachers' period of teaching*

Teachers	Years of Teaching at the University
A	3
B	3
C	13
D	5

The first question in the interview was about the time that the interviewed teachers have been teaching at the university. The question was asked in order to observe how long teachers have been dealing with students and how much teaching experience they have accumulated through the years. From the above answers, it can be assumed that the targeted teachers are semi-experienced and experienced teachers. Additionally, the teachers were asked if they have accumulated quite the teaching experience, their responses were affirmative yet diverse. Teacher A further accounted for his overall teaching experience, he stated, "Well I have been teaching before I mean, well I have been trained in teachers' school for five years and been working at the secondary school for three years and now three years in university. So I am quite knowledgeable

about let's say, teaching English as a foreign language.” Teacher B responded with affirmation and so did teacher C, while teacher D answered with the following, “Not quite, not right there, but I think we can count some experience”

**Q2.** What are the levels and the courses that you are teaching at the moment?

**Table 3.33**

*The courses and levels teachers are in charge of*

Teachers	Response
A	Currently, I teach literature for third year licence and I teach mastery of language as usual for master one.
B	At the moment, we have second year students, the course of reading which is a new course in our programme and master one students, the course of applied linguistics.
C	At the moment, I am in charge of three courses, third year written expression, language mastery for master one and I finished ESP with master two.
D	At the moment, I am teaching written expression and ESP for third year students.

The reasoning behind this question is that only two teachers, A and C, explicitly teach or have taught collocations. Furthermore, unlike other aspects of language such as grammar or writing, collocation does not have its own dedicated modules. Therefore, the teachers were asked to state the current courses, as well as the levels that they are responsible for in order to display the diverse teaching backgrounds of the targeted teachers.

**Q3.** Are you familiar with collocations?

As it was mentioned before, the interviewed teachers were responsible of different courses and not necessarily just the courses that dealt with collocations. At this point, it was essential to elicit how familiar they were with the concept in order to smoothly move to the following related questions. Teacher A responded to the third question by stating that before he had contact with the module of language mastery, he just had a basic idea of collocations until he taught it. Teacher B pointed out that she is familiar with collocations because she especially applied it as a vocabulary building technique for her reading classes. This of course comes in line with various studies that learning formulaic sequences namely collocations help in developing learners' vocabulary repertoire. Teacher C confirmed his familiarity with collocations by stating, "Yeah I am. I am yes. It is a part of the syllabus that I teach for masters". Likewise, teacher D confirmed that he is also familiar with collocations, "Yes, I think to a certain degree"

All of the four teachers showed that they are familiar with the concept of collocations. While teacher B provided a detailed answer, mainly that she used it as a vocabulary learning technique before, the other three teachers were asked to elaborate by providing a brief definition, their summarised definitions can be observed in table 3.34 below.

**Q.** Can you define or explain what collocations are?



**Table 3.34***Teachers' definitions of collocations*

Teachers	Definition of collocations
A	Well, it is a situation in grammar or language where one word follows another but it doesn't happen haphazardly It is A systematic follow up and it has to be that way.
C	It is natural combination of words. The way they occur in a natural way as natives combine them: verbs, noun, adjectives, adverbs.
D	[...] As a definition we can say that it is the co-occurrence of two or more sometimes more lexical items that are frequently used in written or sometimes even spoken language.

**Q4.** While teaching master students, did/do you find their level of collocational knowledge to be sufficient?

One common ground among teachers here was that they cannot generalise the level of collocational knowledge among master students or their students in general. However, the intention behind the question was not to give a general judgment but rather to know whether some or most students face difficulties in terms of collocational use from the perspective of the teachers.

**Table 3.35***Teachers' sample responses toward the learners' collocational level*

Teachers	Sample Response
A	We have catastrophes in terms of academic writing. In speech, I mean fluency is a very big issue. Students you know, tend to keep it simple. They do not use combinations.
B	They use them actually but they do not know that they are called collocations. I mean collocations are frequent stretches of talk so maybe they may use combinations but they are not the right ones they are not the ones which are frequently used among native speakers so they misuse them.
C	You may find some students that are quite familiar with collocations and they use them accurately. Many of them they know that there is something called collocations but in terms of the use they encounter many problems. You can find another category where- another category of students whose knowledge of collocations in other words, it's the first time they get introduced to collocation so we cannot overgeneralise.
D	So I think it depends on the level of the learner and we do not try to generalise, we do not say like high achievers all of them use collocations adequately. It would be very difficult to create such pre-assumptions. However, [...] So yes I think it is used, it is adequate to a certain degree but only and only with high achievers.

The answers provided by the teachers indicated two main things: First that students' collocational competence depends on several elements mainly the learner's overall linguistic competence. Second, some students may use collocations unconsciously i.e. that they are unaware that they are using collocations when and if they do. For example, teacher B and teacher D both agreed that it is a matter of labelling. Nevertheless, it can be concluded that a significant amount of students may not know exactly what collocations are and as a result, they may have deficiencies in collocational use. Most importantly, the teachers highlighted that students do not realise the importance of learning collocations and the benefits it provides. Teacher A stated, "I highly doubt that students like we were I mean really understand the importance of collocations plus having the knowledge of collocations, like putting a word with its appropriate corresponding word. No I do not think they do that"

Similarly, teacher D also responded, "Most importantly I think they do not know the efficiency of collocations in academic writing or sometimes even in writing in general". Teacher B hypothesised that explicit teaching of collocations can yield more affective outcomes: "the problem here, they are not aware and maybe teaching them collocations explicitly is much more affective". In addition, both teacher A and D linked the importance of learning collocations to the development of the writing skill that is strongly argued in the literature related to collocations.

**Q5.** Do you think that students are aware of the concept of collocations? i.e. do they intentionally use collocations while speaking or writing in order to enhance their language production?

Since the language teacher is an assessor and at the same time an observer, this question sought to uncover whether students used collocations consciously as a way to boost their

different language productions. The answers of the teachers displayed interesting revelations.

First teacher A strongly insisted that students must be made aware of collocations through explicit instruction preferably from the first years of their college education. Interestingly, teacher B stated that her second year students are all unaware of collocations or what collocations represent. However, she said that they might use them due to the incidental exposure to native content or input. She further explained that when she assigns them examples and asks them to guess which collocation is more frequent they can sometimes answer correctly:

“sometimes they can identify which one is more frequent and when I ask them why they say: native speakers say this more than this. See. They are aware but again it is a matter of labelling”

Teacher D also happened to point out the idea that incidental exposure to native input can lead to the unconscious use of collocations in students’ language production. He strikingly accounted for the formulaic nature of collocations and that students may learn them as a whole unit without attention to the function of individual words: “those students who are more exposed to native speech or to native talk they usually use them more frequently but they use them as a whole.” He added: “First, they don’t know it is the relationship between two words. They don’t know it is a collocation, and they don’t know even the function of this collocation but interestingly enough they use them correctly”

Teacher C, on the other hand, gave an overall observation about students’ mastery of collocations. Initially, the teacher noted that he could not decide whether they use collocations consciously or unconsciously; yet he noticed that some students have a grasp on the concept. He replied, “In terms of theory, I cannot decide I cannot tell whether they know for example, what is a collocation, what are the types etc., but in terms of practice okay, some students display whether in speech or in their writing let’s say a very satisfactory level of command over

collocations”. However, he quickly went on to say, “but not to a very far extent”. Here he also added another remark, he specified that the students have command over common collocations and they still have some difficulties with more specific or advanced collocations. He argued, “I am talking with common collocations because there still are some ambiguities, some confusions when it comes to the use of intensifying, for example, adverbs with certain verbs like make and do and have.” He concluded his answer by stating that the level of collocational awareness of master student is “okay” to some extent.

**Q6.** Do you think that learning common lexical units and word combinations such as English collocations will help students become more accurate and fluent?

The question was inspired by previous works that linked learning common lexical units found in natural language use to language fluency and proficiency. All of the four teachers seemed to agree with this remark.

**Table 3.36***Interview sample responses about common lexical units*

Teachers	Sample Response
A	Absolutely no doubt! No doubt. The first one accuracy, definitely. I mean academic writing (requires) unity, coherence, cohesion, accuracy. [...] accuracy 100%. Fluency partly. [...] So in fluency yes it would help but it is not the same as accuracy. It is not in the same token as in accuracy because in writing it is obvious.
B	Of course. Sure, for sure, yes. [...] You mentioned here two results: accurate and fluent speakers, of course to reach a native-like fluency and accuracy yes, I think it is highly necessary to teach them such kinds of lexical units.
C	Yeah, no doubt. All what represents the natural English, the spontaneous English, the way natives utter okay. It helps in the development of students' linguistic level.
D	Absolutely yeah, absolutely. This is very interesting, learning common lexical units. [...] I believe that they are very beneficial for learners to use language accurately and to be more fluent and more conversationally competent.

Teacher A emphasised the impact lexical units have on accuracy in writing more than fluency. Here, he also stressed the previous remark that students should necessarily be taught these units before master years. He argued, “I do not like that students get access to this stuff at master.” He also explained that he often tries to make the importance of language mastery clear to his colleagues: “I just try to say that this is not grammar, it is not rules it is not tenses, it is

about language. It is sort of a skill, knowing how to use language, how to manipulate language you see.” As seen in the above table, Teacher B also agreed that teaching lexical units could provide native-like fluency and accuracy. She also noted that teaching students collocations that are more frequent would help with the issue of misusing word combinations, and that it would be “very, very helpful yes and effective among them.” Teacher C also had a similar response. He linked collocations and other word combinations to developing learners’ language fluency and accuracy: “Because the more you use them the better you become fluent and accurate and you sound native-like.” Teacher D had an interesting remark. He stressed that those lexical units must be the native ones in order to build fluency and accuracy because there are non-native common lexical units.

**Q7.** In your opinion what are the factors that may hinder the students’ learning of collocations?

The interviewed teachers answered with important insights regarding the various factors that may impede the learning of collocations. The answered are summarised in the following table:

**Table 3.37***Teachers' responses regarding potentially impeding factors of learning*

Teachers	Potentially impeding factors
A	Students' motivation Failure in teaching collocations
B	Lack of practice Lack of exposure to authentic materials Lack of awareness and motivation of students Failure in teaching students collocations
C	The wide range of collocations Lack of practice Difficulty of some collocations e.g. intensifiers
D	The less frequent collocations are the most difficult to learn Lack of exposure to frequent collocations The substitution of certain lexical items with collocations Where and how to teach collocations

From table 3.37 it can be observed that several answers have some common points.

Therefore, the following categories of factors can be formed:

#### **Lack of exposure to authentic materials**

Lack of exposure of authentic materials or the native speech was one of the common grounds among teachers' responses. This can be due to the lack of reading, listening, or even lack of research from the part of the learners themselves. It can also be due to the lack of teaching and incorporating of authentic input from the part of the teacher.



**Lack of practice**

This aspect has to do more with the learners themselves. As teacher B and C argued, first, students do not practise reading and listening to authentic materials. Second, they do not practise the use of new or innovative language expressions. Teacher C declared, “They always go back to what they know which is their comfort zone. They use this plain English, these common words”. He also reinforced a point, which was previously highlighted in question five. It is that even when students use collocations it could be unintentional and just a matter of a habit built upon incidental exposure. Teacher B also noticed that learners mainly rely on and are satisfied with their basic knowledge of language.

**Lack of motivation and awareness**

Mainly teacher A and B stressed both of these aspects. Teacher A clearly observed that students are often demotivated in the language learning process in general including learning collocations. In order to directly quote him, he said, “Students are not motivated to learn anything let alone learning about collocations.” This is closely related to the issue of awareness that students have. If they are not explicitly made aware of the importance of collocations, they may never see a use in learning such linguistic items. Teacher B established:

[...] There is a purpose for learning anything always. So if we encourage our students to learn collocations precisely collocations to achieve something in their let's say learning experience, journey, they will absolutely I mean be more interested in learning collocations. But since they are not aware of the importance, the effectiveness of learning collocations, to using them appropriately to achieve fluency, accuracy, more proficiency in their language they will never be motivated to learn collocations. So I think that it is our job, it is our duty to introduce them to this concept and to encourage them.

Indeed, perhaps if teachers tried to raise students' awareness of the benefits that collocations have on their overall linguistic and communicative competence, they would be more inclined to learn about collocations.

### **The nature of collocations**

Interestingly enough, both teachers C and D highlighted the challenging nature of collocations. Teacher C explained that the wide range of collocations might be an impeding factor in itself. He replied, "Collocations are sometimes, I think they are unlimited. There are many, many collocations you cannot learn them at once". He also specified that the intensifying adjectives are the most problematic type of collocations for the learners. Teacher D argued that the less occurring collocations pose the most trouble for students. He also linked the lack of exposure to such items to their difficulty. He replied, "The first characteristic of a formulaic expression or a collocation is the occurrence of this item." He continued, "So if they do not see it, if they do not encounter it, if they are not exposed to this collocation more often I think, usually it would be very difficult for them."

### **The inadequate teaching of collocations**

Collocation is often referred to as the neglected phenomena in language teaching. Indeed, teaching such an important lexical unit as a lesson in a unit at the master level can be viewed as rather problematic. Teacher A again encouraged the teaching of collocations right from the beginning of the academic journey: "Again, I would have loved if students were taught from day one, because they are still there as *tabula rasa*. I mean they are fresh you can mold them as you want." Moreover, according to teacher D, teachers are open to incorporating this item in the curriculum, but rather do not know where and how to teach it: "I think it is very difficult because as teachers you do not know where to put this item, or how to teach this item. Because usually,

we do not have a session called collocations” The researcher at this point asked teacher D if he thinks there should be more importance given to teaching collocations to which he responded:

Absolutely, absolutely because when we talk about language structure, usually we have grammar and we have written expression [...]. So I do not think we have a course in which we focus more- I think maybe language mastery but it is too late I mean language mastery M1, I think it is too late now. [...] For the second year I think we should include even a subdivision of a course like grammar and vocabulary not only grammar, so vocabulary, because vocabulary is very important.

Notably, teacher A earlier in the interview mentioned that he would appreciate if it was taught earlier in the module of written expression as an example. Furthermore, teacher B as mentioned before stressed that teachers should introduce such a concept to learners in order to raise their awareness of its importance.

### **Students' linguistic bias**

Teacher C mentioned that students like to use the words which are familiar with and feel comfortable using, which leads to plain and unnatural sounding language. He further explained that even if they are taught collocations they often simply decide not to use it. He explained, “Though they are introduced to such a linguistic item like collocations but they do not often practise it and they always go back to what they know which is their comfort zone.” Teacher D also mentioned that substituting those common words with more frequently used collocations may be difficult for the learners. For example, they stick to the use of the common adverb “very” instead of other more appropriate intensifying adverb and adjective collocations.

**Q8.** Do you think that the teaching materials and methods may have an effect on the quality of students' language learning?

This question intended to uncover the teachers' opinions about the effect that teaching methods and materials may have on learning. It was not specified whether the effect is positive or negative it was left for the teachers to express what they believed.

While interviewing teacher A, he had mentioned students' motivation as an answer to question seven, the researcher then linked that point to the current question: "so you mentioned motivation, do you think that the teaching methods or materials play a role in that?" To which he responded, "Like material selection you mean? Which may probably motivate students to learn collocations? Absolutely, yes. I mean this is the role of the teacher and the role of the designers". He then revealed that neither teachers are motivated nor students are motivated. He proceeded to link this condition of demotivation to the teaching and learning environment and conditions: "Let's not forget that these people are just human beings and you can see the conditions they are working in. You teach English and you do not even have the labs, the essential materials to teach the language."

Teacher B agreed that teaching methods and approaches can impact the process of language learning. She confirmed, "For sure yes. Yes I do believe that the teaching materials and methods have a strong effect on the quality of language learning. Of course yes." She was asked if it could be positive or negative to which she replied, "Now negatively I do not know it depends, because again the teaching process is not just about the teaching methods." She added that, "Many, many different factors interact together to form the teaching process, to create, to shape, the teaching process. So relying on, judging only the teaching method as the only indicator of success or failure in teaching is not fair." However, she also stated that "Opting for more effective teaching methods will absolutely create let's say a progress in the teaching learning progress. Yes."

Teacher C asked the researcher to be more precise. We elaborated: “yes, this is a general question that later will be specified. But I mean in terms of for example using traditional mediums of teaching just like hand-outs or printed materials”. The teacher then focused on the use of technology and teaching aids as teaching methods. According to him the best result comes from the combination of visual and auditory methods. He emphasised that a mixture or a variety of teaching methods and styles is the best kind of teaching because learning styles also vary. He explained, “Personally, I prefer this kind of combination. It is a mixture sometimes this, sometimes that. Students need to be exposed to a wide variety of teaching methods so they can locate themselves as learners.”

Teacher D also agreed that the choice of teaching methodology and material selection both have a role to play in the learning experience. He specifically focused on teachers’ attitudes towards integrating new and innovative methods of teaching. He expressed that one of the chief issues that they face is that, “Teachers do not like to collaborate. Usually they prefer to work individually, etc. and some teachers are somehow you can say less active, it means they do not prefer to bring something new.” He continued, “They like just to keep repeating the same traditional methods and as far as the materials are concerned I have a textbook or I have a set of examples in a textbook I will keep using it for the upcoming five years.” He concluded by stating that an update is more than necessary. He advised that teachers need to re-evaluate and to assess the teaching materials and methods every educational period for instance.

**Q9.** Are you familiar with the corpus and data-driven learning approach?

From the answers of this particular question, it can be deduced that all of the teachers are familiar with the corpus approach, mainly corpus analysis and corpus-based studies. However, they were not that familiar the data-driven approach or DDL.

Teacher A explicitly stated that he was not informed of this approach until he had contact with this research work. Teacher B expressed her personal experience with corpus analysis and corpus-based studies in relation to sociolinguistics. She was asked if she ever used an online corpus before to which she replied by reference to her previous study of Algerian and French code switching found in television shows that were collected from YouTube videos. She then was asked if she used any English online corpus (data-driven) she then stated that it was a future project for her and that she has not yet used an English corpus. Teacher C deliberately stated that he is only familiar with the corpus approach but not DDL.

Teacher D on the other hand was familiar with the corpus approach and somehow with DDL. He decided to compare it with the competency-based approach (CBA). He explained, “It is like when we talk about CBA or competency-based approach, usually when we ask teachers do you know it? They say no, but when we try to elaborate or explain the process they say yeah yeah I remember I am using this even.” He then was directly asked if he is familiar with DDL, he then responded, “The concept is clear but the application of the concept I think not like CBA. CBA we use it, most of us actually use it without even knowing what is CBA.” He then proceeded, “But as far as data-driven in terms of research, language teaching, specialty teaching, etc. I think we are not really using it, but I think it is efficient. I think it is efficient.”

At this point, the researcher explained the concept of DDL, particularly in relation to foreign language teaching. It was in order to ensure that the teachers understood what it entails for better chances of answering the following questions. They were also asked what they thought of such an approach to language teaching; all of their responses were positive.

**Q10.** Do you believe that such an approach to foreign language teaching is achievable in our context?

The teachers' answers to this question were quite diverse and interesting. One teacher expressed his frustrations with the current circumstances; some teachers were more optimistic and hopeful of integrating such new approaches. Another teacher was more detail-oriented and provided criteria for the application of such an approach in our context.

Teacher A expressed his frustration regarding the teaching conditions and lack of administrative support. He specified, "I am always in for innovation, any new stuff. Sometimes it's a bit frustrating. I mean look at the classes, look at the conditions. No data shows, no air conditioning, no electricity, no nothing." He was very interested in the DDL approach and recommended that we should make a suggestion of this as a teaching pedagogy. He also displayed some signs of discouragement himself in another response he said, "I mean although I am all in for technology, then I say no one cares just do the job as it is in the conditions which are there. Do your duty." He also pointed out that even existing and rather important classes tend to have a lack in material support: "you know in many oral classes they do not have labs so that is a big problem there."

Teacher B was rather optimistic. At first she assumed that DDL can be applied at home; however, it was explained that ideally this approach would be applied inside the language classroom as a teaching approach. She then replied, "Yes we hope so; we hope so yes. But whether it is really achievable or not in our context I have no answer. Why? Because I am not a decision maker anyway." We then made it clear that the opinions of teachers should be accounted for because they are a valuable agent in the teaching and learning experience. Indeed, the opinions and insights of the language teacher are highly important especially in this case. She then expressed, "My opinion is that I hope so, I wish. Yes."

Teacher C provided a wealth of knowledge based on his teaching experience. To summarise his point, he insisted on setting clear objectives right from the beginning. He listed a number of requirements that include: provision of materials, of a clear syllabus, of administrative commitment, instructors' commitment and students' commitment. He stressed that the availability of resources alone is not an indicator of success and that the human factor also needs to be addressed. To answer the question of whether or not this approach can be achieved in our context he asserted that if there are good intentions to boost the level of education and if there is a collaboration of the before mentioned elements that it would yield positive outcomes. He persisted, "For sure if there is this collaboration of all those agents that I have mentioned earlier. For sure the percentage of success will start from a very insignificant percentage and then starts to increase till all objectives will be achieved". He then mentioned that the starts are always difficult with so many pitfalls and setbacks but with adjustments and revisions it will lead to better results.

When DDL was first explained to teacher D he called it demanding yet doable. When asked if it is achievable in our context he replied, "I think yes, I think it is achievable to a certain degree yeah." He then proceeded to account for the mentality that our community has towards applying new technology: "Nowadays you see when we are faced with a problem like this global pandemic, suddenly everyone knows how to use a computer". He explained, "when we are faced with a problem it means we do not have any options suddenly we know how to do everything. So I think it is the same thing here." Teacher D realistically observed that, in our context, a problem needs to exist in order for change to happen. He concluded by arguing that we need to be more decisive and that if we want to try a new method, we should take the initiative and face the



challenges that may arise. He then mentioned that this approach can be achievable: “I think it is doable, it is very achievable.”

**Q11.** In your opinion, what are the obstacles that may prevent the application of such an approach in our EFL context?

The teachers provided important insights and highlighted different issues that may hinder the application of DDL in our current context. Their responses are summarised in table 3.38 below.

**Table 3.38***Potential obstacles of integrating DDL*

Teachers	Potential obstacles
A	Teachers' and students' attitudes Infrastructure Management issues Lack of trained teachers and syllabus designers Extrinsic and intrinsic motivation
B	Lack of awareness Psychological reasons (unwillingness) Lack of the necessary materials
C	The administrative obstacle Various requirements Inconsistent participants and commitment levels Time management issues and constraints Students' attitudes and motivation The provision of materials, syllabus selection and design
D	The physical materials Teachers' and students' attitudes

Based on the table, it can be observed that the teachers provided various and important answers based on their perspective and experience. Some common themes emerged yet also some essential details that need to be acknowledged. Nevertheless, these commonalities can be grouped into a number of categories with various factors.

**Teachers' and learners' attitudes**

This particular obstacle seems to be an essential one. Teacher A pointed out that both the teacher and student are involved in the process of including new approaches yet he emphasised the role of the teacher. He stated that the student could be shaped to accept certain decisions and outcomes because the teacher has the upper hand: "I would say it is the responsibility of teachers to get this into the classroom." Teacher D mentioned that both the teachers and learners have a say in adapting new approaches. He stated, "Do actually these teachers and students, will accept, I mean their attitudes towards this method would it be like positive or negative." Indeed the teachers and students' willingness to change and to adapt to new approaches is highly important and need to be taking into consideration.

**Administrative issues**

This important point was strongly emphasised in the answer of teacher C. He mentioned that if the administration is not willing to provide the needed financial, moral, and even human support that the operation would be quickly dismissed. He spoke, "So if the administration does not provide this it is the first pitfall; it is the early end of the operation." He also revealed that the administration is not always open to change: "The administration sometimes does not welcome these initiatives and they see it as an obstacle in itself." Indeed, having new and demanding teaching approaches integrated may be viewed as "disruptive", especially if the administrative staff has a set of methods that they refuse to change.

**Management issues**

Teacher A listed management issues as one of the obstacles that can be encountered while applying a new approach. Similarly, teacher C stated that the participants must be consistent in terms of members and commitment level. He also mentioned an important factor

that is the time factor. So, the time dedicated to studying this application needs to be well managed. According to him these kinds of operations need extended periods not just a period of a semester, because it cannot be an indicator of success or failure. Teacher C also accounted for other logistical issues such as the planning of the syllabus material itself and the ways this approach can be applied.

### **Lack of resources**

The infrastructure, lack of computer-assisted labs, lack of the appropriate teaching space and environment all play a significant role in the success of this approach. Teacher B mentioned that she will encourage her students to learn via DDL but at their homes. When she was asked why not here at their university, she replied, “Because we do not have the materials. We do not have the necessary materials.” Teacher A and D also mentioned the current conditions and physical environment as an impeding element.

### **Lack of awareness and motivation**

Teacher B mainly mentioned the lack of awareness that teachers have in terms of knowing the importance of such new approaches. On the contrary, when teacher B was asked if he thought it was a problem of awareness he denied and argued that it is a matter of motivation and ethics. Teacher B mentioned that both extrinsic and intrinsic motivation play a crucial role in adapting such methods. He demonstrated that extrinsic motivation is the idea of support. It is the support that teachers and students need mainly from the administration, whether moral or physical (rewards, training programs, field trips). Intrinsic motivation according to him is the psychological conditions of the participants, whether they are internally motivated or not. Teacher C mentioned that the students’ motivation also needs to be accounted for in this case. He explained, “The third one [obstacle] is related to students themselves. Again, are they motivated

to experience something new or not, or they feel quite comfortable with the way they currently learn.”

### **Lack of trained professionals**

Teacher A and C highlighted this point. Teacher A described the lack of training in didactics and pedagogy as a “handicap” to progress. He mentioned that he would appreciate the inclusion of trained professionals in syllabus and material design to be in charge of method and material selection. Meanwhile, teacher C listed a team of trained professionals as a requirement for the sound application of any new approach. He mentioned that teachers could not take on such a daunting task without professional help and assistance. He specified, “I would love to have for example, a teacher expert in the idea of corpus-based learning so that I can know all what are the pros and the cons of the idea.” He also added that having experts in those kinds of specialties might even effect the teachers’ motivation to collaborate and as a result incorporate the intended approach.

Since the interview was a semi-structured interview i.e. that the questions where more flexible, some additional questions became known and as a result some additional response categories have been added.

Since the idea of using authentic materials and texts is central to the DDL approach, teachers were asked whether they were in favour of using these kinds of materials in teaching. These are their responses:

**Table 3.39***Teachers' responses to using authentic materials*

Teachers	Response
A	We have to use authentic materials because the learners have to have access to natural language. [...] we should keep our hands off to the maximum from interfering into the materials unless it's like extremely, extremely necessary.
B	It is better. For me, yes. Especially in teaching collocations.
D	for authentic materials yes I think authentic materials provide much, much more than non-authentic materials for learners, but they require much more effort from the teacher because authentic material needs to be modified

Teacher C mentioned that DDL is a combination of both merits developing autonomous learners as well as linguistically competent learners, thanks to the use of authentic materials. He observed:

This is a mixture of two merits where students get first of all autonomous, more than that they will acquire authentic language with varied contexts this is what I mean. They are going to develop their linguistic knowledge okay to reach wider understanding of language in its nuances like collocations.

Furthermore, the researcher asked the teachers about the reasons behind sticking to the use of traditional methods of teaching. They responded with the following:

**Table 3.40***Reasons behind using traditional materials*

Teachers	Reasons
A	Yes, I mean the teacher in the classroom yes he should let the flow go, assign students tasks and stuff but he should exert efforts outside the classroom in preparation, in material selection, in designing syllabus tasks, researching. That is his job because at university you are not solely a teacher, you are a researcher.
B	They want to stay in their comfort zone. They do not want to change they do not want to go beyond what they are already familiar with.
C	We have a fear, the main I mean, chief obstacles in revolution, in making a revolution of the system, we have two things: the obstacle of the fear, and the resistance of change. We are afraid to experience something, because we are afraid of the consequences. We do not know where to head.
D	I think it is just because it is easier that is all. Because to apply something new, it requires much more effort and energy etc., resources, and I think to a certain degree most of us are very lazy to create to initiate you know new methods or approaches. [...] So, I think for teachers, it depends on teachers' personality, perspective, etc. so we cannot generalise, I think it is impossible to generalise.

To account for teachers' perceptions of the DDL and the corpus approach, the researcher asked each interviewee about their thoughts regarding this approach. The following table represents a sample of their answers.

**Q.** do you find this approach interesting/beneficial?

**Table 3.41**

*Teachers' opinions about DDL*

Teachers	Thoughts about DDL
A	Yeah yeah very much I mean, I am in for any pedagogy that focuses on the student. That gives more space, more room in class for the student and minimizes to the maximum the role of the teacher.
B	of course of course. Yes very very beneficial and helpful by experience as a learner not a teacher as a learner. By experience, this approach is great of course. Because again we are dealing here with authentic language and the target language is English so authentic language in English that would be very powerful. Yes.
C	The way you explained it, it's amazing that students experience something out of the box something not traditional, something as I said related to the technology-based learning plus corpus-based learning.
D	I think its demanding but it's beneficial [...] to a certain degree I think it's very beneficial and if only applied correctly.



It can be deduced that all the teachers perceive the integration of the DDL approach in a positive light. Furthermore, they see it as a beneficial and interesting method especially in teaching a foreign language. However, as teacher D stated, it must be applied correctly and in a logistical way in order for it to reach its full effect.

### **3.3 Discussion and Synthesis of the Results**

In brief, the study at hand had a set of objectives to reach including assessing the current collocational competence levels of master one students, as well as understanding the hurdles that prevent the application of modern approaches such as data-driven learning in our EFL classrooms. In addition, following the case study design, this work attempted to recognise the learners and the teachers' perceptions towards implementing DDL. This exploratory study benefited from the pragmatist views thus, applying a mixed-methods approach that yielded rich qualitative and quantitative data obtained with the help of the participants in this study. The study exploited three different data collections tools that were a test about collocational knowledge for students, students' questionnaire, and teachers' interviews. The data obtained were used to address and possibly answer the four central research questions. The answers of the four questions, as well as the corresponding hypotheses will be discussed in accordance to the findings obtained from the data analysis.

Research question one sought to uncover the current level of collocational knowledge of master one students. The test was the most suitable tool to answer this question thus; a test was specifically designed for this study. It was achievement test adapted from McCarthy and O'dell's book English collocations in use the intermediate version. It consisted of five tasks of varied content. The findings of the test revealed that more than half (53%) of the answers of the sample (37 participants) were incorrect which proves their level of collocational knowledge to be

insufficient and further calls for the need of teaching collocations in language classrooms.

Hence, that answers our first research question, and subsequently confirms research hypothesis number one that the participants will have deficiencies in collocational knowledge.

It was essential to consult four experienced and semi-experienced teachers in the course of conducting our research because the teacher is an integral part of the language classroom, and may be aware of most its problems. Consequently, four face-to-face interviews were arranged with teachers of applied linguistics at the section of English. The main objectives were to know the factors that may hinder the development of collocational knowledge of their students, to understand the teachers' perceptions of data-driven learning approach, and to elicit their opinions regarding the potential application of this approach in our EFL context.

The data obtained from the interviews were rich and diverse leading to interesting revelations and insights. First, factors that may hinder learners' acquisition of collocations include: the complicated nature of collocations itself, the lack of practice and exposure to authentic input, the lack of awareness of the importance of collocations from the part of the learners, and the inadequate teaching of collocations.

In addition, the teachers agreed that the teaching methods and materials play a significant role in the quality of language learning, in general, and collocation learning specifically. This involved the answer to research question number two and led to the confirmation of the second research hypothesis. Furthermore, the teachers were well aware of the benefits collocations have in language learning; they were also supportive of teaching common lexical units such as collocations to language learners in order to boost their fluency and accuracy. They even recommended that students must be taught collocations from the early stages of their academic careers and not until master level.

Another question that the interview tackled was how teachers perceive the idea of integrating DDL. According to the teachers' answers and reactions, it can be deduced that the four teachers had positive reactions towards the idea and were remarkably interested and open-minded towards it. Moreover, interview questions aimed at listing the chief obstacles that could prevent the integration of this approach in our context. A common observation that could be made is that the major obstacle in the application of any new method or technique is the human obstacle. In other words, according to the four teachers, the human factor plays a major role in the smooth implementation of new technology and teaching methods. It has to do with motivation, awareness, willingness to change and adapt, in addition to the attitudes and acceptance of administrative staff and even learners themselves.

Needless to say that the physical obstacles remain an issue as well, issues such as monetary support, resources unavailability, and infrastructure. All this presents the answer to question number three, and consequently confirms research hypothesis number three that the teachers will have positive perceptions regarding DDL.

Meanwhile, the last question that the interview attempted to uncover was to what extent is the application of the DDL approach possible in our context. Regardless of the impeding factors, the answers of the interviews displayed that it is indeed possible to a certain extent. This also comes in line with what we previously anticipated in research hypothesis four. Nevertheless, as one of the interviewees insisted, it has to be in a well-organised and well thought out manner. Additionally, the effectiveness and success of this operation depends on a set criteria and a combination of elements that need to be present. The elements include all of the before mentioned aspects.

The third data collection tool was the students' questionnaire. The questionnaire aimed at answering the remaining questions that were left unanswered in this research. Mainly, it sought to answer the second part of research question three on how do learners perceive data-driven learning. From the answers of learners in the sample, it can be concluded that not only are students aware of technological innovations and already use them, they are also open-minded towards the implementation of internet-based approaches such as the DDL and corpus approach in the classroom. This also confirms another part of research hypothesis three that learners indeed perceive DDL positively.

Moreover, the students in the sample agreed that the out-dated teaching methods, unvarying teaching styles, and the lack of ICT aids largely and negatively affect the learning experience. Other factors that impeded the students' learning of language and specifically collocations may include lack of motivation and interest in mastering the language.

An interesting revelation occurred when comparing the test results with the answers of the questionnaire. It appears as if learners as far as this sample is considered, have a strong theoretical understanding of collocations but significantly inadequate knowledge when it comes to practice or application. This can be due to the great emphasis on memorisation as a way of learning and testing in our context.

In conclusion, the findings of the current study showed that both participants i.e. teachers and students are aware of the importance of collocations in language learning. Thus, it can be stated that there should be more emphasis in teaching lexical units, in general, and collocations specifically in order for the learners to build vocabulary, develop fluency, and obtain accuracy. Furthermore, teachers and students in this sample perceived the integration of DDL positively and have hopes for its implementation in the future. Overall, there need to be more diversity in

the teaching methodology opted for in the EFL classroom to prevent possible disinterest of learning and loss of motivation. Finally, it is possible to apply DDL as it is with other internet-assisted methods as long as careful and organised planning is involved in the process and the human factor is willing to make the efforts involved.

### **Conclusion**

This chapter provided an overview and rationale of the methodological choices of this study. It highlighted research paradigms, research approaches, and research designs. Moreover, the current chapter discussed the data collection tools, their structure and aim, the data collection procedures, as well as the data analysis procedures. It also discussed the sampling technique used in this study. Additionally, results of each data collection tool were displayed, analysed, and interpreted. Finally, the results were summarised and discussed in the final section of the chapter.

### **General Conclusion**

Language learning largely involves accomplishing language proficiency and mastery. The wealth of vocabulary that a language learner possesses not only proves useful in receptive language skills but also productive ones. It is evident that vocabulary knowledge is not limited to knowing long list of words and their meanings, but that it also includes a multitude of aspects such as the collocational aspect of those words.

Collocational competence entails that the language learner understands how certain words are combined and used together in a native-like manner. The development of such a skill can be challenging hence the explicit instruction of collocational patters remain highly recommended. Furthermore, it was noticed that a significant percentage of master students often faced issues while attempting to produce sound collocational sequences, as a result they often resorted to the use of basic word combinations and common vocabulary items. It was hypothesised that this impediment can be a consequence of various factors that may include the lack of emphasis on teaching collocations and the deficiency in opting for new and interesting teaching methods and materials. Thus, the use of an online corpus programme was suggested as a possible solution and teaching approach.

In order to fulfil the practical part of this work, a sample of 37 master one students of applied linguistics were chosen through the convenience sampling technique. In addition, four teachers of the same branch also contributed to the study. The philosophical framework followed in this study was the pragmatist one. Hence, the mixed-methods research approach was utilised. As for the research design, the case study research design was the most suitable for this work due to the exploratory nature and objectives of the latter.

At an attempt to reach the research objectives and in order to answer the research questions, three data collections tools were used. These include: collocations test for students, a questionnaire for students, and an interview with teachers. The three instruments were thoughtfully designed and carefully validated to ensure the reliability of the data. For the sake of making sense of the rich gathered data. Both content analysis and descriptive statistics were employed to analyse, interpret, categorise, and display the results.

The results obtained from the test further confirmed that the level of collocational competence of the students in the sample was insufficient which in turn further emphasises the need for the explicit instruction of collocations. Moreover, the results of the questionnaire proved that the students are indeed familiar of collocations theoretically, and that they are aware of its importance in language learning. They also seemed to face difficulties while learning collocational patterns; however, they were open to learning such patterns in order to improve their accuracy and fluency. Additionally, the answers from the questionnaire attributed the deficiency of collocational knowledge to different reasons including the lack of motivation, the lack of interesting and updated teaching methods and materials, and the unvaried teaching styles.

The questionnaire also revealed that the students of the sample are equipped with technological competence as they are familiarised with internet learning and the use of ICTs. The respondents also appeared to be supportive of the integration of internet-based teaching methods such as the DDL approach.

The teachers' interview also disclosed interesting findings. The interview results aligned with what the questionnaire discussed mainly that the learners can be aware of collocations in theory but struggle in terms of practice. In addition, the potential impeding factors of learning collocations that were discussed include: the difficult nature of collocations, the lack of exposure

to authentic materials, and the lack of practice. Like the case of students, the interviewed teachers also agreed that implementing DDL may yield positive learning outcomes and that it is recommended overall. The teachers were very interested in this approach as it can bring new technology and variety into our English teaching situation, despite the possible obstacles that may prevent its application. The teachers concluded that it is possible to integrate as long as some conditions are applied such as administrative support and the acceptance of the learners and teachers alike.

### **Implications and Recommendations**

Following the results obtained by the study, a number of recommendations can be drawn for further considerations especially in the EFL context. This list of recommendations chiefly provides suggestions for the implementation of collocations in the teaching curriculum specifically through using the DDL approach due to its positive impact as proven by the literature.

- Collocations should be given more importance in the EFL teaching scene due to its apparent benefits such as vocabulary acquisition, speech fluency, accuracy in writing, and overall language proficiency.
- Students' should be exposed to and made aware of collocations right from the beginning years of learning English as a foreign language and not until the master level.
- Vocabulary should be taught as an individual subject or module, where students actively acquire vocabulary items through various strategies such as learning collocations.
- Language learners need to be exposed to a wealth of authentic input and materials for a more genuine acquisition of the English language.



- Educators should opt for diverse teaching styles and updated teaching materials for the sake of raising learners' involvement and motivation.
- There should be more room for learners to partake in the language learning experience in order to foster their autonomy and critical thinking skills as well as to practice the language and experience it first-hand.
- The EFL classroom should be less teacher-centred and the teacher should be more of a guide and less of a controller.
- Technological and modern teaching methods should be integrated inside the language classrooms which meet the modern day learner's expectations and standards.
- Internet-based approaches generally and the data-driven approach specifically should be integrated in the EFL classroom because of the positive effects it provides.
- The students should be made aware of the existence of such techniques in learning in order to freely explore their intricacies at their own time.
- The English language teachers should exploit corpus content more, giving that it is easily accessible and available for teaching purposes.
- Corpus software such as COCA should be introduced to the language learners in the future as a way to access and learn authentic language.
- The administrative staff should encourage and sponsor the integration of innovative and modern teaching methods by providing the needed physical and human resources.
- There should be trained professionals in the aspect of syllabus and material designs that provide the essential outline and suggestions for integrating new techniques.
- Thoughtful and logistical planning is required by the educational staff if or when implementing new teaching approaches.

### **Limitations and Suggestions for Future Research**

The possibility of facing limitations and constraints is always present in the course of conducting any research work. These may include issues of availability of participants, time constraints, and even lack of resources. Even though the study's overall objectives were met, its limitations must still be addressed. Furthermore, based on what was observed and realised in this work, suggestions for future works can be made.

The chief issue that faced this work was the limitation of time. The work was done in a period of five months which is a rather limited time frame for research. Moreover, another obstacle was related to methodology, this study was initially intended to be a quasi-experimental study, yet due to the limited time and resources such as lack of computer labs and internet access in the university, the study was approached from a different perspective. Thus, it is recommended for future research dealing with collocations and online corpus programmes to opt for a quasi-experimental research design, especially since the variables are highly cause and effect oriented.

Furthermore, the researcher could not locate any standardised tests for collocations hence; we resorted to designing one for the sake of this study. It is therefore suggested to use a standardised collocational test if an appropriate one is found.

Another and possibly, final suggestion for future research would be to include a larger number of participants in order to gain a more thorough understanding of the research problem under discussion and to potentially generalise the findings to a broader context.

### References

- a glossary of corpus types*. (2017, March 15). In Other Words. <https://wmtang.org/corpus-linguistics/a-glossary-of-corpus-types/>
- Abdul Rehman, A., & Alharthi, K. (2016). An Introduction to Research Paradigms. *International Journal of Educational Investigations*, 3(8), 51–59.
- Anderson, R. C., & Freebody, P. (1981). Vocabulary Knowledge. In Guthrie, J. T. (Ed.), *Comprehension and Teaching: Research Reviews*. (pp. 77-117). International Reading Assosiation.
- Attar, E. M., & Allami, H. (2013). The Effects of Teaching Lexical Collocations on Speaking Ability of Iranian EFL Learners. *Theory and Practice in Language Studies*, 3(6).  
doi:10.4304/tpls.3.6.1070-1079
- Beare, K. (2018, January 21). *Learn about collocation in English*. ThoughtCo.  
<https://www.thoughtco.com/what-is-collocation-1211244>
- Beeching, K. (2014). Corpora in language teaching and learning. *Recherches En Didactique Des Langues et Des Cultures*, 11(1). <https://doi.org/10.4000/rdlc.1672>
- Benavent, G. T., & Peñamaría, S. S. R. (2011). Use of authentic materials in the ESP classroom. *Encuentro*, 89–94. <https://doi.org/1989-0796>
- Bennett, G. R. (2010). *Using corpora in the language learning classroom : corpus linguistics for teachers*. University Of Michigan Press, Cop.
- Benson, M., Benson, E., & Ilson, R. F. (2010). *The BBI combinatory dictionary of English : your guide to collocations and grammar* (3rd ed.). J. Benjamins Pub. Co., Cop.

Biber, D., Conrad, S., & Reppen, R. (1996). Corpus-Based Investigations of Language Use. *Annual Review of Applied Linguistics*, 16, 115–136.

<https://doi.org/10.1017/s0267190500001471>

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python*. O'reilly.

Bonelli, E. T. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics*. Routledge.

Boulton, A. (2010). Data-Driven Learning: The Perpetual Enigma. In S. Roszkowski & B. Lewandowska-Tomaszczyk (Eds.), *Explorations across Languages and Corpora*. (pp. 1-18). Frankfurt: Peter Lang.

Boulton, A. (2017). Data-Driven Learning and Language Pedagogy. In *Language, Education and Technology: Encyclopedia of Language and Education* (Vol. 3, pp. 1–12).

[https://doi.org/10.1007/978-3-319-02328-1\\_15-1](https://doi.org/10.1007/978-3-319-02328-1_15-1)

Cheng, W. (2010). What can a corpus tell us about language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics*. Routledge.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). New York: Routledge.

*Communicative competence*. (n.d.). TeachingEnglish | British Council | BBC.

<https://www.teachingenglish.org.uk/article/communicative-competence#:~:text=Communicative%20competence%20refers%20to%20a>

Cook, L. D., & Kamalodeen, V. J. (2019). *Mixed Methods Case Study Research* [Slides]. Mixed Methods International Research Association. <https://www.ualberta.ca/international-institute-for-qualitative-methodology/media-library/international-institute-of-qualitative-methods/webinars/mixed-methods/2019/cook-kamalodeen-webinar-final.pdf>

- Corpus types*. (2016, August 19). Sketch Engine. <https://www.sketchengine.eu/corpora-and-languages/corpus-types/>
- Creswell, J. W. (2014). *Research design : qualitative, quantitative & mixed methods approaches* (4th ed.). Sage.
- Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. *BMC Medical Research Methodology*, 11(1).  
<https://doi.org/10.1186/1471-2288-11-100>
- Dawson, C. (2009). *Introduction to research methods : a practical guide for anyone undertaking a research project* (4th ed.). Robinson.
- Demiray Akbulut, F. (2018). Effects of Teaching Collocations on Academic Writing. *Turkish Online Journal of English Language Teaching (TOJELT)*. 3(2), 64-82.
- Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research : construction, administration, and processing* (2nd ed.). Routledge.
- Easwaramoorthy, M., Zarinpoush, F., & Imagine Canada. (2006). *INTERVIEWING FOR RESEARCH* (pp. 1–2). Imagine Canada.  
[http://sectorsource.ca/sites/default/files/resources/files/tipsheet6\\_interviewing\\_for\\_research\\_en\\_0.pdf](http://sectorsource.ca/sites/default/files/resources/files/tipsheet6_interviewing_for_research_en_0.pdf)
- El-Dakhs, D. A. S. (2017). Collocational Competence in English Language Teaching: An Overview. *SSRN Electronic Journal*, March 2015. <https://doi.org/10.2139/ssrn.2834432>
- English Corpora: most widely used online corpora. Billions of words of data: free online access.* (n.d.). [Www.English-Corpora.Org](http://www.English-Corpora.Org). [https://www.english-corpora.org/coca/help/coca2020\\_overview.pdf](https://www.english-corpora.org/coca/help/coca2020_overview.pdf)

- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4.
- Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*, 67(June), 155–179. <https://doi.org/10.1111/lang.12225>
- Gabrielatos, C. (2005). Corpora and Language Teaching: Just a Fling or Wedding Bells?. *Tesol-Ej*, 8(4), 1–37.
- Gavioli, L., & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT Journal*, 55(3), 238–246. <https://doi.org/10.1093/elt/55.3.238>
- Gee, R. W. (2015, March). USING THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH TO LEARN ABOUT NEAR SYNONYMS: COLLOCATIONS. *The Newsletter of the Computer Assisted Language Learning the Internet Section*. Retrieved from <http://newsmanager.commpartners.com/tesolcallis/issues/2015-03-03/7.html>
- Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. *Phraseology: An Interdisciplinary Perspective*, September 2014, 27–49. <https://doi.org/10.1075/z.139.07gra>
- Guz, E. (2014). Formulaic sequences as fluency devices in the oral production of native speakers of Polish. *Research in Language*, 12(2), 113–129. <https://doi.org/10.2478/rela-2014-0004>
- Haji, S. (2015). *Collocation: Theoretical Considerations, Methods and Techniques for Teaching It*.

- Hamed Mahvelati, E., & Mukundan, J. (2012). The Role of Cognitive Style in the Collocational Knowledge Development of Iranian EFL Learners through Input Flood Treatment. *English Language Teaching*, 5(10). <https://doi.org/10.5539/elt.v5n10p105>
- Hasan, K., & Shabdin, A. A. (2016). Conceptualization of depth of vocabulary knowledge with academic reading comprehension. *Pasaa*, 51(August), 2016.
- Haywood, S. (n.d.). *Using concordance lines*. [www.nottingham.ac.uk](http://www.nottingham.ac.uk).  
<https://www.nottingham.ac.uk/alzsh3/acvocab/concordances.htm#:~:text=A%20concordance%20line%20is%20a>
- Hodne, S. (2009). *Investigating word combinations in two English textbooks for Norwegian upper secondary school students*.
- Johannesson, P., & Perjons, E. (2014). *An introduction to design science*. Springer.
- Johns, T. (1988). Whence and whither classroom concordancing? In T. Bongaerts, P. de Haan, S. Lobbe, & H. Wekker (Eds.), *Computer applications in language learning* (pp. 9–27). Dordrecht: Foris.
- Johnson, R. B., & Christensen, L. (2019). *Educational Research: quantitative, qualitative, and mixed approaches*. Sage Publications.
- Kivunja, C., & Kuyini, A. B. (2017). Understanding and Applying Research Paradigms in Educational Contexts. *International Journal of Higher Education*, 6(5), 26.  
<https://doi.org/>
- Koester, A. (2010). Building small specialised corpora. In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics*. Routledge.
- Krieger, D. (2003). Corpus Linguistics: What It Is and How It Can Be Applied to Teaching. *The Internet TESL Journal*, Vol. IX(3). <http://iteslj.org/Articles/Krieger-Corpus.html>

- Kumar, R. (2011). *Research methodology: A step-by-step guide for beginners* (3rd ed.). London, MA: SAGE.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399–436. <https://doi.org/10.1111/j.0023-8333.2004.00260.x>
- Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning and Technology*, 21(2), 32–51.
- Leech, G. (2005). Adding Linguistic Annotation. In M. Wynne (Ed.), *Developing Linguistic Corpora: a Guide to Good Practice*. AHDS Literature, Languages, and Linguistics. [http://icar.cnrs.fr/ecole\\_thematique/contacti/documents/Baude/wynne.pdf](http://icar.cnrs.fr/ecole_thematique/contacti/documents/Baude/wynne.pdf)
- Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory into Practice* (1st ed.). Heinle ELT.
- Lexical approach*. (2019). TeachingEnglish | British Council | BBC. <https://www.teachingenglish.org.uk/article/lexical-approach>
- Little, D. (1999). *Definitions, issues and problems*. Authentik Language Learning Resources.
- Manca, E. (2012). *Context and Language*. Università del Salento: Coordinamento SIBA.
- Martyńska, M. (2004). Do English language learners know collocations? *Investigationes Linguisticae*, 11, 1. <https://doi.org/10.14746/il.2004.11.4>
- McCarthy, M. (2004). *Touchstone: From Corpus to Course Book*. Cambridge: Cambridge University Press.
- McCarthy, M., & O'dell, F. (2017). *English collocations in use : how words work together for fluent and natural English: self-study and classroom use. Intermediate*. Cambridge University Press.



- McCarthy, M., & O’Keeffe, A. (2010). Historical perspective: What are corpora and how have they evolved? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics*. Routledge.
- McEney, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- McEney, T., & Xiao, R. (2011). What Corpora Can Offer in Language Teaching and Learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning*. Routledge.
- McGee, I. D. (2006). *Lexical intuitions and collocation patterns in corpora*. June, 382.  
<https://search.proquest.com/docview/1370376569?accountid=11005%0Ahttp://pqdd.sinica.edu.tw/twdaoapp/servlet/advanced?query=U584816>
- Meyer, C. F. (2004). *English Corpus Linguistics : An Introduction*. Cambridge University Press.
- Miščin, E., & Pavičić Takač, V. (2013). Exploring the collocational competence of non-native users of medical English. *JAHN-European Journal of Bioethics*, 4(7), 235–256.
- Moghadam, S. H., Zainal, Z., & Ghaderpour, M. (2012). A Review on the Important Role of Vocabulary Knowledge in Reading Comprehension Performance. *Procedia - Social and Behavioral Sciences*, 66, 555–563. <https://doi.org/10.1016/j.sbspro.2012.11.300>
- Mudraia, O. (2001). Lexical approach to second language teaching. *Eric Clearinghouse on Languages and Linguistics*, 1–8.  
<http://www.cal.org/content/download/1546/16373/file/LexicalApproachtoSecondLanguageTeaching.pdf>
- Nation, I. S. P. (2000). *Learning Vocabulary in Another Language*. Cambridge University Press.

- Nemoianu, A. M. (n.d.). *The Lexical Approach*. [Http://Www.Tesl-Ej.Org/](http://Www.Tesl-Ej.Org/). <http://www.tesl-ej.org/wordpress/issues/volume1/ej02/ej02r3/>
- Ngula, R. (2018). CORPUS METHODS IN LANGUAGE STUDIES. *Perspective on Conducting and Reporting Research in the Humanities*, 205–223.  
[https://www.researchgate.net/publication/327234896\\_CORPUS\\_METHODS\\_IN\\_LANGUAGE\\_STUDIES](https://www.researchgate.net/publication/327234896_CORPUS_METHODS_IN_LANGUAGE_STUDIES)
- Nordquist, R. (2019, March 2). *What Is the Lexical Approach?* ThoughtCo.  
<https://www.thoughtco.com/what-is-a-lexical-approach-1691113>
- O’keeffe, A., Mccarthy, M., & Carter, R. (2007). *From corpus to classroom : language use and language teaching*. Cambridge University Press.
- Rahimi, M., & Momeni, G. (2012). The effect of teaching collocations on English language proficiency. *Procedia - Social and Behavioral Sciences*, 31, 37–42.  
<https://doi.org/10.1016/j.sbspro.2011.12.013>
- Rao, V. C. (2018). The Importance of Collocations in Teaching of Vocabulary. *Journal for Research Scholars and Professionals of English Language Teaching*, 2(7), 1689–1699.
- Römer, U. (2008). Corpora and language teaching. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: An International Handbook*. W. De Gruyter.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford Univ. Press.
- Sirkel, K. (2017). the Importance of the Acquisition of Formulaic Language for the Development of Communicative Competence and Learner Autonomy. *Estonian Journal of Military Studies*, 4, 35–50. [www.ksk.edu.ee/publikatsioonid](http://www.ksk.edu.ee/publikatsioonid)
- Smith, R. (2008). Learner autonomy. *ELT Journal*, 62(4), 395–397.  
<https://doi.org/10.1093/elt/ccn038>

- Taherdoost, H. (2016). Sampling Methods in Research Methodology; How to Choose a Sampling Technique for Research. *SSRN Electronic Journal*, 5(2).  
<https://doi.org/10.2139/ssrn.3205035>
- Thornton, P. B. (2013, September 30). *Three Teaching Styles*. Faculty Focus | Higher Ed Teaching & Learning. <https://www.facultyfocus.com/articles/philosophy-of-teaching/three-teaching-styles/>
- Understanding Language Proficiency*. (n.d.). [www.languageproficiency.com](http://www.languageproficiency.com); Language Testing International. Retrieved June 6, 2021, from <https://www.languageproficiency.com/lti-information/understanding-proficiency>
- Varlamova, E. V., Naciscione, A., & Tulusina, E. A. (2016). A study on the phenomenon of collocations: Methodology of teaching English and German collocations to Russian students. *International Journal of Environmental and Science Education*, 11(6), 1275–1284. <https://doi.org/10.12973/ijese.2016.398a>
- Webb, S., & Kagimoto, E. (2011). Learning Collocations: Do the Number of Collocates, Position of the Node Word, and Synonymy Affect Learning? *Applied Linguistics*, 32(3), 259–276. <https://doi.org/10.1093/applin/amq051>
- Williams, G., & Millon, C. (2012). Palmer, Firth and Internet: Drawing together collocational threads. *Proceedings of Corpus Linguistics 2011*, 1–34.
- Wong, I. F. (1993). Book Reviews: Corpus Concordance Collocation. John Sinclair. Oxford: Oxford University Press. 1991. xviii, 170pp. *RELC Journal*, 24(2), 100-104.  
[doi:10.1177/003368829302400207](https://doi.org/10.1177/003368829302400207)

- Wood, D. (2006). Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency. *Canadian Modern Language Review*, 63(1), 13–33. <https://doi.org/10.3138/cmlr.63.1.13>
- Wood, D. (2019). Classifying and identifying formulaic language. *The Routledge Handbook of Vocabulary Studies*, 30–45. <https://doi.org/10.4324/9780429291586-3>
- Wynne, M., & Berglund Prytz, Y. (2012). Types of corpora and some famous (English) examples. In *Corpus Linguistics course*. OUCS Hilary. <http://tinyurl.com/669o4zt>
- Yan, H. (2010). Study on the Causes and Countermeasures of the Lexical Collocation Mistakes in College English. *English Language Teaching*, 3(1), 162–165. <https://doi.org/10.5539/elt.v3n1p162>
- Yılmaz, E., & Soruç, A. (2015). The use of Concordance for Teaching Vocabulary: A Data-driven Learning Approach. *Procedia - Social and Behavioral Sciences*, 191, 2626-2630. doi:10.1016/j.sbspro.2015.04.400
- Yin, R. K. (2003). *Case study research : design and methods* (3rd ed.). Sage Publications.
- Zhao, Y. (2009). *An empirical study of L2 learners ' use of lexical chunks and language production*.

## Appendices

### Appendix A: Consent Form for Teachers

#### Interview Consent Form

Dear Teacher,

I am conducting a research about exploring the use of a web corpus for learning collocations. You are kindly asked for an interview, which serves as a data collection tool for this study. Your answers will help answer the formulated research questions regarding the teachers' attitudes towards collocational competence and the use of a data-driven approach. Additionally, this research study seeks to obtain deeper insights on the difficulties that face master students in acquiring collocations and further suggests the use of an online corpus as a solution. The responses and the data you provide will be anonymous and will be used for research purposes only. If you consent to partake in a recorded interview, please sign this consent form. Your help and collaboration are highly valued.

Yours sincerely,

Hamidane Zineb

#### **Contact details:**

**Email:** zainabhamidane@gmail.com

Mohamed Kheider University of Biskra, Faculty of Letters and Foreign Languages

Section of English

I have read and clearly understood the researcher's request. I consent to the participation in a recorded interview for the research work undertaken by Hamidane Zineb.

**Name:**.....

**Date:**.....

**Teacher's Signature:**

**Appendix B: The Test**  
**Test on Collocational Knowledge**

Dear Master students, Thank you for agreeing to sit for this test, your time and efforts are highly appreciated. This test is just for research purposes you are not being evaluated. Your answers will serve as data for answering one of the formulated research questions in my Master dissertation. Rest assured that all your information, answers, and data are well protected and will remain anonymous. Thank you once again for your cooperation and understanding.

**Task One:** Choose an adverb from the box to replace **very** in each of these expressions.

Example: very cold : bitterly cold

utterly, strongly, bitterly, ridiculously, highly

- |                       |                       |
|-----------------------|-----------------------|
| 1) very ashamed       | 6) very disappointing |
| 2) very cheap         | 7) very opposed       |
| 3) very controversial | 8) very ridiculous    |
| 4) very stupid        | 9) very easy          |
| 5) very successful    | 10) very concerned    |

**Task Two:** Use a collocation from the newly formed collocations (your answers) to complete each of these sentences.

1) The flight from London to Rome was .....  
 ..... . It only cost 20 euros.

2) Some people love her new book, others are very angry about it. It is  
 .....

3) His father was a peace campaigner all his life and was .....  
 ..... to war.

4) The exam results were ..... for the whole class. We had all expected to do much better.

5) When I realised how much my selfish behaviour had upset everyone, I was .....

6) In the 1990s she ran a(n) ..... company which made outstanding profits.

7) Everyone got more than 95% correct in the test; it was .....

8) You must apologise immediately. It was a(n) ..... remark to make.

9) She has always been ..... about the environment and would like to work for a conservation agency.

10) That you would even think that I would steal money from you is ..... ! You must be crazy!

**Task Three:** In each of these sets of phrases, one is not a correct collocation. Cross out the **incorrect** one.

1) strongly recommend  strongly influence  strongly love  strongly dislike

2) highly educated  highly profitable  highly unusual  highly exhausted

3) bitterly regard  bitterly regret  bitterly resent  bitterly criticise

4) absolutely convinced  absolutely tired  absolutely devastated  absolutely absurd

5) deeply unhappy  deeply religious  deeply successful  deeply committed

**Task Four:** Underline and correct the **eight** collocation errors in this text.

In the morning I made some work in the garden, then I spent a rest for about an hour before going out to have some shopping in town. It was my sister's birthday and I wanted to do a



special effort to cook a nice meal for her. I gave a look at a new Thai cookery book in the bookshop and decided to buy it. It has some totally easy recipes and I managed to do a good impression with my very first Thai meal. I think my sister utterly enjoyed her birthday.

**Task Five:** Choose the correct collocation.

- 1) She had / took / paid attention to what I told her and started working harder.
- 2) I had / made / took over a hundred photographs on my trip to Antarctica.
- 3) She made / paid / brought me a nice compliment yesterday.
- 4) I got / made / had a bad dream last night and woke up sweating.
- 5) The President made / gave / paid tribute to all the people who had supported him.
- 6) I got / took / had a liking to my new doctor the moment I met her.
- 7) I gave / made / had a feeling I had met Richard before, but I couldn't remember where.
- 8) I went to Douglas Farnham's funeral to give / take / pay my last respects to a fine man.
- 9) Shall we make / get / have a party for Ruby? She's moving to Manchester next week.
- 10) I had / got / took a feeling that he was trying to hide something from me.

**Thank you very much for your time and cooperation!**

Reference:

McCarthy, M., & O'dell, F. (2017). *English collocations in use : how words work together for fluent and natural English: self-study and classroom use. Intermediate*. Cambridge University Press.

**Appendix C: The Opinionnaire for the Test**  
**The Opinionnaire**

Dear teacher,

The previous test aims at eliciting the level of collocational competence of first year master students. It was adapted from McCarthy and O'dell's English collocations in use. After reading the test, you are kindly asked to answer the following questions as they serve to clarify and correct any mistakes. Your willingness to help and your time are greatly appreciated!

1. Is the test of reasonable length?

Yes  No

If no, should it be:

Shorter  longer

2. The test is for intermediate level students, is the level appropriate for first year master students?

Yes  No

If no, should it be:

More difficult  less difficult

3. Are the instructions clear enough or is there any ambiguity?

Yes  No

If yes, which task questions need reformulation? Please state them below.

.....  
 .....

If you have any additional comment or suggestion, please feel free to state it.

.....  
.....  
.....

**Thank very much for your time and cooperation.**

## Appendix D: The Students' Questionnaire

### Questionnaire

Dear Master students, you are kindly asked to answer the following questionnaire. It is about learning collocations through web-based corpus. A web corpus is a free online corpus of the English language that allows students to access data collected from authentic language use in order to learn language aspects such as collocations. I personally assure you that all your collected data will be confidential and protected. There are no right or wrong answer, feel free to answer however you like. Thank you for your time and cooperation as your answers will serve as data for answering the related research questions.

**Instructions before Answering:** Dear students, before answering please read the questions carefully. Some of the questions asked regarding internet use are about internet use in the university context. In other words, the use of internet inside the classroom for learning purposes and not at home. Thank you for your understanding.

#### Section One: Personal Information

1. Gender:

Male                   Female

2. Age: .....

3. How long have you been studying the English language overall?

.....

#### Section Two: Collocations

1. Are you familiar with collocations?

Yes  No

If yes, can you provide a brief definition?

.....  
.....

2. Do you view learning collocations as important in language learning?

Yes  No

If yes, please explain why

.....  
.....

3. Have you faced any difficulties while learning collocations in the Language Mastery module in semester one of this academic year?

Yes  No

4. How would you rate the difficulty of learning collocations?

Very difficult

Difficult

Not very difficult

Not difficult at all

5. To what degree do you agree with the following statements?

a) Learning English collocations will make you sound more fluent and natural while speaking.

Strongly agree  Agree  Neutral  Disagree  Strongly disagree

b) Learning English collocations will help you express yourself in various ways.

Strongly agree  Agree  Neutral  Disagree  Strongly disagree

c) Learning collocations can improve your writing skill.

Strongly agree  Agree  Neutral  Disagree  Strongly disagree

d) Collocations can make you sound more native-like in conversations.

Strongly agree  Agree  Neutral  Disagree  Strongly disagree

**Section Three: Online Corpus and Data-Driven Learning**

1. Which of the following materials are used by your language teachers in the classroom?  
(Choose an answer for each method)

	Never	Rarely	Sometimes	Often	Always
Data shows					
Computers					
Language corpus					
Audio visual aids					
Hand-outs and worksheets					

2. What do you normally do when you meet a word that is new to you?

- Use Google and search for its meaning
- Use a traditional dictionary
- Use a dictionary application on your phone

Ask a teacher

Other: .....

Please justify your choice:

.....

3. How often do you refer to the internet while doing research or in your studies (at home)?

Rarely

Sometimes

Often

Always

4. Do you prefer:

A teacher-centered classroom (the teacher is in control of the learning process)

A student-centered classroom (the students are responsible for most of their learning)

Justify:

.....

.....

5. Do you prefer learning through:

Authentic materials (materials based on the native use of the language)

Adapted materials (modified or artificial language made for learning purposes)

6. How often do you find yourself losing interest in learning?

Rarely

Sometimes

Often

Always

7. Do you find yourself losing interest in the learning process because of:

- Teaching methods
- Materials used
- The unvaried teaching styles

Other: .....

Justify your answer:

.....  
.....

8. Would you prefer learning through the internet i.e. using internet websites in the classroom?

Yes  No

Justify:.....  
.....

9. Do you think your willingness to learn will improve if teaching (inside the classroom) was through internet access?

Yes  No

10. Do you believe that there should be teaching methods that allow students to use the internet inside the classrooms?

Yes  No

Justify:.....  
.....

11. Are you familiar with corpus linguistics?

Yes  No

If yes, can you please provide a brief definition?

.....  
.....

12. Did your teachers ever introduce the concept of language corpora?



Yes       No

If yes, in what way or in what instances?

.....  
.....

13. Have you ever learned language explicitly through the use of a corpus in a classroom?

Yes       No

14. What do you think are the reasons for not using internet based programmes or corpus in our EFL context?

- Time constraints
- Resources
- Management issues
- Internet access
- Number of students

Other: .....

**Thank you for your time and efforts. Best of luck!**

**Appendix E: The Online Questionnaire**  
**Students' Questionnaire**

Dear master one students, you are kindly asked to answer this simple questionnaire that consists of one question. It will serve as a source of data in accomplishing my master dissertation about collocations and data-driven learning. Rest assured that all your data and information will be kept confidential and safe. Thank you so much for your time and cooperation.

**Collocational Knowledge:** Collocations are common word combinations that frequently occur together in speech or writing. For example: deeply unhappy, strongly believe, commit a crime, etc.

1. Were you familiar with "collocations" before you learned it in the language mastery module or after?

- I knew collocations before language mastery
- I did not know collocations until I learned it in language mastery

**Appendix F: The Opinionnaire for the Questionnaire**  
**The Opinionnaire**

1. Are there any repetitive questions?

Yes  No

- If yes, please specify them.

.....  
.....  
.....

2. Did you find any spelling or grammar mistakes in the questions?

Yes  No

- If yes, please specify them below.

.....  
.....  
.....

3. Are there any irrelevant questions that need to be removed?

Yes  No

- If yes, please provide the number of the question(s) below.

.....  
.....  
.....

4. Is the questionnaire of reasonable length?

Yes  No

- If no, does it need to be:

Shorter  Longer

5. Are there any ambiguous questions that need to be reformulated and / or clarified?

Yes  No

- If yes, please indicate the questions that require reformulating.

.....  
.....  
.....

6. Are the response categories appropriate?

Yes  No

7. Is the order of the questions appropriate

Yes  No

- If No, which questions need reordering?

.....  
.....

8. If there are any questions that you believe are of close relevance to the purpose of the questionnaire but were not included, please write them below.

.....  
.....  
.....

9. If there are any additional comments or suggestions please notify them.

.....  
.....  
.....

**Thank you very much for your time and collaboration.**

**Appendix G: Teachers' Interview Questions**

Q1: How long have you been teaching at the university?

Q2: What are the levels and the courses that you are teaching at the moment?

Q3: Are you familiar with collocations?

Q4: While teaching master students, did/do you find their level of collocational knowledge to be sufficient?

Q5: Do you think that students are aware of the concept of collocations? i.e. do they intentionally use collocations while speaking or writing in order to enhance their language production?

Q6: Do you think that learning common lexical units and word combinations such as English collocations will help students become more accurate and fluent?

Q7: In your opinion what are the factors that may hinder the students' learning of collocations?

Q8: Do you think that the teaching materials and methods may have an effect on the quality of students' language learning?

Q9: Are you familiar with the corpus and data-driven learning approach? (if yes please explain)

Q10: Do you believe that such an approach to foreign language teaching is achievable in our context?

Q11: In your opinion, what are the obstacles that may prevent the application of such an approach in our EFL context?

**Appendix H: The Opinionnaire for the Interview Questions****The Opinionnaire**

Dear teacher,

The current study is entitled “exploring the use of a web corpus for learning collocations”. It seeks to answer research questions such as: What are the factors impeding students’ learning of collocations? How do teachers and learners perceive data-driven Learning? To what degree is the application of data-driven Learning and a corpus-based approach achievable?. By answering this opinionnaire, you will be helping in the process of refining and validating one of the data collection tools which is the teachers’ interview. Thank you very much for your time and considerations.

1. Is the number of questions acceptable?

Yes  No

If no, do they need to be:

Less  More

2. Are the questions asked comprehensive?

Yes  No

3. Are the questions clear and direct?

Yes  No

4. Are the response categories appropriate?

Yes  No

5. Are there any redundant or unnecessary questions that need to be removed?

Yes  No

If yes please state them below.

.....  
.....

6. Are there any questions that need reformulation?

Yes                       No

If yes, please state them below.

.....  
.....

7. If you have any additional remarks and or suggestions, please feel free to note them.

.....  
.....  
.....

**Your time and collaboration are highly valued. Thank you.**

## ملخص الدراسة

معرفة المفردات تعتبر ركيزة في تعلم اللغة. تشمل هذه المعرفة مفاهيم أكثر عمقاً، وبالتالي، فإن أحد العناصر الأساسية لمعرفة المفردات هو معرفة المتلازمات اللفظية أو كيف تظهر تركيبات كلمات معينة بشكل طبيعي في اللغة. ومع ذلك، لوحظ أن غالبية طلاب الماستر يعانون من حيث المفردات المتنوعة و التجميع الدقيق للكلمات. كان من المتوقع أيضاً أن الأساليب والمواد التقليدية المستخدمة قد تساهم في التعلم الغير كاف للمتلازمات اللفظية. انطلاقاً من هذه النقطة، حاولت هذه الدراسة تقييم مستوى الكفاءة اللفظية لـ 37 طالباً من طلاب الماستر في جامعة بسكرة الذين تم اختيارهم كعينة للدراسة. علاوة على ذلك ، هدفت الدراسة إلى استكشاف إمكانية دمج التعلم القائم على البيانات (DDL) كنهج للتعليم واكتشاف العقبات المحتملة التي قد تمنع تطبيقه في سياقنا. من أجل الوصول إلى هذه الأهداف، تم تطبيق نهج مختلط يتبع تصميم دراسة الحالة. وبناءاً عليه، تم تصميم اختبار واستبيان للطلاب وأجريت مقابلات مع أربعة من معلمي القسم. أظهرت النتائج أن المهارة اللفظية لدى غالبية الطلاب في العينة أقل من المتوسط. علاوة على ذلك، قد تكون عدة عوامل قد أدت إلى مثل هذه النتائج بما في ذلك التدريس غير الكافي للمتلازمات اللفظية ونقص التنوع في طرق التدريس. كشفت النتائج أيضاً أن المشاركين في المشروع كانوا متحمسين لإدماج نهج DDL. أخيراً، قد تتنوع المعوقات التي قد تواجه تطبيق هذا النهج وتشمل مسائل تنظيمية وإدارية. ومع ذلك ، فإن تطبيق DDL في سياقنا ممكن وموصى به في ظل التخطيط الجيد والتنظيم المدروس.