A Sequential Explanatory Study of the Learning of Semantically Non-Transparent Words and Collocations

Institute of Education

University of Reading

Sarah Saleh Alamoudi

Thesis submitted for the degree of

Doctor of Philosophy

May 2023

Declaration of Original Authorship

I confirm that this is my own work and the use of all materials from other sources has been properly and fully acknowledged.

Signature:

Print Name: Sarah Saleh Alamoudi

Date: 05/22/2023

Acknowledgements

Firstly, I would like to express my gratitude towards God Almighty for His help and guidance throughout my journey towards completing my PhD. My sincere appreciation then goes to my first supervisor, Suzanne Graham, whose advice and support has been invaluable. She was always approachable and understanding , particularly in the most challenging of times. Furthermore, I would like to also thank my second supervisor, Daisy Powell, who provided me with very helpful feedback, particularly in relation to the statistical portion of my thesis.

My sincere appreciation also goes to the Saudi Ministry of Education and King Abdul Aziz University (KAU), who generously funded this thesis through a scholarship. Furthermore, I would like to thank my colleagues within the English Language Institute (ELI) at KAU for their ongoing support for my project, and also my students who agreed to be participants in this study. Without you this research would not have been possible.

A very special thank you also goes to my family and friends, particularly my late father Saleh Hussien Alamoudi, who passed away in May 2021 while I was undertaking my PhD. He always encouraged me to pursue higher education, was always there to instantly support me, and hoped to see me as a PhD holder. I would not have been where I am today if it wasn't for him and my mother's prayers.

I would also like to thank my four children, Hussain, Jowanah, Ibrahim and Hamza for their patience and understanding during the time I had to spend away from them to undertake this project. Finally, I am eternally grateful to my close friends especially Khadija Nasseef for her emotional support and reassurance.

COVID-19 Impact Statement

I am currently a PhD student in my final year, and I am due to submit my thesis for examination in May 2023. The following statement outlines ways in which my doctoral research was affected by the COVID-19 pandemic.

I initially planned to recruit a larger number of participants for my study. The intended participants were students enrolled in English language courses at a university in Saudi Arabia. My study included multiple time points to carry out the data collection, although due to growing concerns about the pandemic, and fears that the university may close, I was not able to recruit the number of participants I had originally planned to include in my study. When I started my data collection, I was able to recruit a sufficient number of participants from the first English Language course (the first term) to take part in the post-test that followed the intervention. However, not all participants were present for the second post-test and delayed post-test. Recruiting participants from the second English language course (second term) presented even greater challenges, as the university was closed because of a national lockdown.

Abstract

This study explores the complexity behind vocabulary learning in the ESL classroom in Saudi Arabia. It has been suggested that vocabulary learning, particularly when it includes words and collocations which are not semantically transparent, presents a particular challenge for learners. The primary aim of the study was to explore the impact of two instructional methods - contrastive analysis and translation (CAT) and CAT+Corpus - on the short-term recall of semantically non-transparent vocabulary in receptive and productive knowledge. Nineteen vocabulary target items, which included collocations and semantically non-transparent words, were selected for this study.

A total of 56 students agreed to take part. All participants were female first-year university students enrolled in a compulsory ESL course run by the English Language Institute (ELI) at a university in Jeddah, Saudi Arabia. Participants were divided into three different groups: CAT, CAT+Corpus, and control. There were two separate teaching interventions that took place over the course of two weeks. The CAT only group received interventions consisting of reading a passage in English, which included the target items, followed by a reading comprehension task. Afterwards, the teacher explained the similarities and differences between the target items in both the L1 and L2 where the words and lexical system were briefly but explicitly explained. Students also received two worksheets (for the receptive task and the productive task) which required them to match the target items to their equivalent translations. The intervention in the CAT+Corpus

group was similar, however participants also worked with a parallel corpus when completing both receptive and productive tasks. The control group did not receive any treatment.

The study adopted a mixed-methods approach comprising pre- and post-tests in receptive and productive areas, with all three groups. Additionally, 11 students (four from the CAT+Corpus group, three from the CAT group and four from the control group) also took part in think-aloud (TA) interviews while completing the fill-in-the-gaps task. Among these, seven students from the two treatment groups were also asked about their views on the treatment they received.

The findings revealed that the two treatment groups who received CAT and CAT+Corpus instruction made significantly greater vocabulary gains on their post-tests for both receptive and productive short-term recall compared to the control group, which only experienced incidental learning. However, the findings of the immediate post-tests in receptive and productive knowledge suggest that learners who received CAT instruction led by the teacher learned more vocabulary items compared to those who received CAT+Corpus instruction, although the differences were not significant between the two groups. While both groups received explicit CAT for vocabulary instruction, the results suggest that CAT teaching alone could be more effective compared to supplementing it with parallel corpus tasks.

The thesis concludes by discussing these findings in relation to the noticing hypothesis, L1 transfer, and the involvement load hypothesis (ILH), as well as their implications for ESL pedagogy.

Table of Contents

List of Appendices

Appendix A - Tasks, Tests and Intervention Materials

Appendix B - Productive Test

Appendix C - Receptive Test

Appendix D – Ethical Forms Approved in English and Arabic

Acronyms

English as a foreign language (EFL)

English as a second language (ESL)

Second language acquisition (SLA)

First language on native tongue (L1)

Second Language (L2)

Ministry of Education (MoE)

English Language Institution (ELI)

Involvement load hypothesis (ILH)

Working memory (WM)

Contrastive analysis and translation (CAT)

Focus-on-form (FonF)

Focus-on-forms-(FonFs)

Think-aloud (TA)

Stimulated recall (SR)

Arabic English Parallel Corpus (AEPC)

Chapter One

Introduction

1. Introduction and Problem Statement

1.1 English in Saudi Arabia

English is widely used in Saudi Arabia in many places and for many purposes. On the surface, the study of English in Saudi Arabia is often referred to as English as a foreign language (EFL) because it is used in classrooms and in the curricula in schools and universities However, English can also be referred to as English as a second language (ESL) because Saudis have become accustomed to communicating in English with all kinds of people including pilgrims, labourers and tourists who do not speak Arabic. This is in addition to the fact that the English language is in everyday life for Saudis. For example, street names, signs, products and brands usually exist in a combination of both Arabic and English. The media also focus on English and have dedicated television channels for English speakers. Where both ESL and EFL can be used, this study will use the term EFL since the focus will be on classroom instruction and teaching.

Saudi Vision 2030 primarily focuses on developments within the fields of economy, business, and culture (https://www.vision2030.gov.sa/). Nonetheless, there is a strong emphasis on different aims and objectives related to the expansion of all forms of education in the nation. These aims and objectives have been proposed to enhance the training of teachers at all educational levels, and present opportunities to improve the instructional methods and learning outcomes for students. Teaching and learning English was among these aims. In fact, Saudi Arabia's Ministry of Education (MoE) aims to provide students with university-level standards of academic English to prepare them for the needs of an increasingly competitive local and global labour market.

Recently, Saudi Arabia has sought to improve the quality of English instruction in classrooms by using online tools including digital curricula and computers This was implemented through electronic textbooks for the new curriculum, namely the "Madrasati" platform for students, during the academic year 2021–2022. Through the use of electronic textbooks, the MoE seeks to promote and regulate the quality of e-learning, enhance learning motivation, and improve learning outcomes (Ministry of Education, 2021)

1.2 Vocabulary Learning

Where second language acquisition (SLA) is concerned, vocabulary learning has always been of special interest to language teachers who understand the crucial importance of a wide vocabulary for all language skills. In fact, it is important to note that the process of learning new vocabulary is not limited to form and meaning. Since vocabulary learning takes place incrementally and gradually (Schmitt, 2000), Nation (1990) summarises various aspects of learning a new word as:

> a word's spoken form; a word's written form; a word's part of speech, derivative forms, and grammatical patterns; a word's collocations; how frequently a word is used in a language; the stylistic constraints that determine if a word is appropriate in a particular context (register); a word's conceptual meaning(s), and a word's semantic network of associations. (p.31)

However, it has to be mentioned that learners do not become familiar with all of these aspects at once; instead, by having frequent exposure to a lexical item, which is considered key to learning words (Nation, 2001), they may develop an understanding of one or more of them. Moreover, even when having a deep understanding of a lexical item, that understanding may not be accurate when producing collocations. Collocations, in particular, are quite challenging for

non-native speakers to learn and master (Wray et al., 2016). A study by Bahns and Eldaw (1993) involved an investigation into the learning of 15 English collocations in comparison with one-unit words. The subjects were 58 German university students who were advanced English learners. They had to complete two tasks (a German-English translation and a cloze task), in which they were required to produce English collocations to investigate their knowledge of collocations. The findings clearly indicate that EFL learners' knowledge of one-unit words is far broader than that of collocations and that they "are more than twice as likely to select an unacceptable collocate as they are to select an unacceptable general lexical word" (Bahns & Eldaw, 1993, p. 108). Indeed, one reason for the difficulty of learning collocations is the neglect of their instruction in EFL classrooms (Bahns & Eldaw, 1993; Lewis, 1993; Schmitt, 2000). However, it has to be emphasised that some collocations, such as *the upshot and the bottom line,* are considered opaque – that is, semantically non-transparent, as they do not have literal translations in some languages.

In order for language learners to be able to understand and use target collocations, they need to raise their level of awareness to notice and recognize collocations (Lewis, 1993). Therefore, it can appear to be quite challenging for many English learners to recognize these collocations individually without teacher assistance. As a result, utilising these semantically non-transparent words and collocations in oral activities and writing tasks can also be daunting for many language learners, and Saudi second language (L2) students are no exception. In fact, Alqahtani (2015) emphasised the importance of vocabulary knowledge as a whole for students' academic achievements in English courses, and that lacking vocabulary is considered a major problem in Saudi Arabia. He linked this struggle to rote learning in vocabulary without understanding. He also pointed out that Saudi students rely on their instructors' teaching methods,

and that teachers are viewed as “the prominent source of knowledge” (Alqahtani, 2015, p. 23). Similarly, Albousaif (2011) indicated that Saudi students' reliance on their teachers for vocabulary learning is particularly high. This heavy dependence on the teacher for lexical input and a lack of independent learning when it comes to vocabulary acquisition can be problematic. Furthermore, if teachers do not successfully deliver the meaning of an unknown word, the students are less likely to understand it.

Although Alqahtani (2015) addressed vocabulary knowledge in terms of size, there remain first language (L1) transfer issues faced by many L2 learners (Figueredo, 2006), including Arabs learning EFL (Fender, 2003; Saigh & Schmitt, 2012). For instance, when it comes to word recognition, Arabic-speaking EFL learners rely heavily on L1 transfer of phonological processes, which is a problem because of the differences between English and Arabic orthography (Fender, 2003). As Fender (2008) states:

> it is natural that words with more orthographic and spelling pattern complexity will be more difficult for both L1 and EFL to spell. Nonetheless, the problem is especially acute among the Arab EFL participants who seem to struggle with orthographic complexity. (p. 34)

On a similar note, Saigh and Schmitt (2012) found that in relation to reading and writing in English (short vowels), L1 Arabs encounter difficulties in identifying the correct vowels, and this is due to “the orthographic and literacy saliency strategies transferred from their L1” (p. 33). Their study involved 24 native Arabic-speaking participants studying at a British university, whose English proficiency ranged from intermediate to upper-intermediate levels. They completed a test consisting of 80 sentences with correct grammar structure; however, some of these sentences had one spelling error. Their task was to correct the spelling mistakes and write the correct spelling of

the word beside the sentence. Results showed that participants were better able to recognize and learn long vowels in comparison to short vowels.

These findings help illustrate some of the challenges experienced by Arabic-speaking learners of English. There are different approaches towards understanding the complexity behind acquiring new vocabulary. Vocabulary learning can be explored from a pedagogical perspective (Nation, 2013), a psycholinguistic perspective (Jiang, 2000, 2002, 2004) (see Section 2.3.2), as well as from a cognitive perspective (see Working Memory in section 2.3.3). For example, from a pedagogical perspective, when it comes to different instructional methods, some researchers recommend focusing on the word form and encourage EFL educators of Arabic speakers and EFL curriculum designers for Arabic-speaking students to devote more time to spelling activities. For instance, EFL teachers might consider explaining the negative impact of transferring Arabic strategies to English spelling (Sagail & Schmitt, 2012). Alqahtani (2015) suggests different instructional techniques in teaching vocabulary to EFL students, instead of relying on one technique in vocabulary learning. He refers to several researchers who investigated easier strategies in vocabulary learning. For example, Brewster et al. (1992) used a number of techniques such as objects and visual aids, translation, drawing, illustrations and pictures, contrasting words with opposites, enumeration (lists of items), expressions and gestures, guessing from context, and eliciting. Alqahtani (2015) also refers to strategies used by Schmitt and McCarthy (1997) that include guessing from context, mnemonic techniques, and using vocabulary cards, as well as strategies suggested by Murcia (2001) such as guessing meaning from context, mnemonic devices and keyword techniques, and vocabulary notebooks. These are only a few of the techniques that have been suggested to help students learn vocabulary more effectively. Ultimately, there is a need for further investigation into teaching and learning L2 (second language) vocabulary, especially

given that many language learners find vocabulary learning to be their greatest obstacle (Meara, 1980).

Moreover, when studying vocabulary knowledge, it is crucial to understand that there is a very thin boundary between incidental and intentional learning. Incidental learning refers to the acquisition of words that takes place naturally when interacting with a text, while intentional learning implies acquisition as a result of deliberate effort and study (This is further discussed in section 2.4). Mondria (2003) provides a good example to illustrate the distinction between these two types of learning. In his study, he explored the learning of Dutch students who were taking French lessons. The vocabulary items were selected from their course materials. In the incidental learning experiment, students had to infer the meaning of French words, then check the meaning using a word list. Through translation and memorisation, the intentional learning experiment participants acquired the target items. The findings illustrated that learners who gained vocabulary through incidental learning acquired just as much vocabulary as those who learned it intentionally. This suggests that incidental learning can be just as successful as intentional learning. In relation to the Saudi context, Ahmad's study (2012) examined which vocabulary learning type would be most successful. His findings showed that the incidental group who experienced vocabulary items through contextual clues outperformed the intentional learning group that experienced vocabulary items through word meaning i.e., synonyms. Ahmad's study (2012) contradicts Stæhr's (2008) claim that the most useful vocabulary learning is intentional learning.

'The present study considered pedagogical approaches that included both incidental and intentional learning. For the control group and the two intervention groups, incidental learning was targeted in initial activities that involved reading passages and answering comprehension questions and where the focus was on meaning only. For the two intervention groups in the study only,

follow-up activities were then implemented that intentionally shifted the L2 learners' attention to the target items through a) the teacher's translations and explanations (CAT group) and b) the additional highlighting of target items in a corpus (CAT + Corpus group) (see 1.4).

The learning intervention tasks for the CAT and CAT + Corpus groups therefore drew attention to the target items within a context of incidental learning, since learners also engaged in the reading comprehension exercise, which was meaning driven. All participants were aware that there would be a post-test taking place immediately after the intervention, however only the CAT and CAT+ Corpus groups engaged in activities which intentionally drew attention to the target items after they had completed the reading comprehension exercise. These two groups therefore engaged in incidental learning first and subsequently in intentional learning through CAT and CAT + Corpus activities.

1.3 Problem Statement

1.3.1 Vocabulary Learning Affecting Other Language Skills

A review of the literature shows that many researchers agree on the importance of L2 vocabulary learning (e.g., Alghamdi et al., 2020; Al-Khairy, 2013; Alqahtani, 2015; Laufer & Nation, 1999) for a range of other language skills. At more advanced levels, however, involving language of a more academic kind, L2 learners are likely to encounter vocabulary of a particularly complex or confusing kind, such as semantically non-transparent words and collocations. For L2 learners, learning such vocabulary items that do not have word by word equivalents or translations is more likely to be challenging (Alharbi, 2017) than learning items that have such equivalents. This is because for semantically non-transparent vocabulary learners must know both the

transparent and opaque translations. By doing so, L2 learners can expand and improve their knowledge of vocabulary relevant for academic English study.

Lack of vocabulary knowledge has been shown to affect academic writing for Arabic EFL speakers (Alghamdi et al., 2020; Al-Khairy, 2013). Al-Khairy (2013) explains that Saudi students learning English face several issues in writing academic tasks and that a lack of appropriate use of vocabulary is the first major problem, followed by – in order of difficulty - incorrect grammar, use of irregular verbs, use of prepositions, spelling, use of articles, use of punctuations, and use of question words.

These issues concerning the learning of English have existed for a long time. In fact, earlier studies that looked into the English language proficiency of Saudi students between 1978 and 1980, revealed significant, unexpected results. For example, Al-Guayyed (1997) discussed the total average TOEFL scores of students and pointed out that Saudi students received the fifth worst score out of 474,000 applicants from 143 different nations at that time. All four of the language skills tested were indicative of the weaknesses of Saudi students. Al-Guayyed (1997) partly attributed this to Saudi test-takers' limitation in vocabulary knowledge.

Moreover, these concerns in learning English vocabulary are problematic because they conflict with the aims of Saudi Arabia's Ministry of Education (MoE) which are, according to the MoE website, to acquire the necessary standards of academic English and to connect them with the labour market. These aims are to be achieved through developing the students' English language abilities in speaking, reading, and writing, preparing them for post-secondary study, and ensuring their eligibility for entrance to national and international universities. This is the reason that the Saudi Ministry of Education introduced English from the first grade of primary school

(public and private), starting from 2021-2022 (https://www.moe.gov.sa). Prior to this, English teaching was introduced from the fourth grade at age nine (Alqahtani*,* 2015*).*

The MoE's priority is to equip Saudi learners with the necessary skills to communicate easily with the outside world, especially in the context of globalisation and the rapid lifestyle transformation that has recently taken place in the country. However, there are issues concerning Saudi students in the area of vocabulary learning that have remained unsatisfactory (Alqahtani, 2015; Alshammari, 2020) and the impact has been reflected in speaking and writing skills. As a result, some Saudi students are unable to reach the satisfactory level of EFL that is required internationally and at university-level.

Furthermore, the current researcher has ten years' experience teaching English, and throughout that time, she has seen typical L2 mistakes and vocabulary limitations made by her Saudi students in speaking and writing. Some of these mistakes involve the misuse of articles or incorrect terminology use as a result of L1 interference. Additionally, it was noted that Saudi students frequently rely on a small number of different words, which can be both a restriction and a sign of a limited vocabulary. For instance, students frequently overuse certain adverbs, such as *and, but, however,* and *also.*

This lack of vocabulary knowledge can be a hindrance to speaking and writing. In fact, prior research has shown that Saudi students struggle with English writing, frequently making errors while writing for academic purposes, and typically only engaging in sentence- or, at most, paragraph-level academic writing (Al-Khairy, 2013), which can be partly attributed to lack of vocabulary (Al-Guayyed, 1997). With these well-established language barriers as a backdrop, and in order to help Saudi students improve their L2 vocabulary knowledge, it was decided to investigate their learning and recalling of semantically non-transparent words and collocations.

The purpose for selecting such vocabulary as target items is to explore ways of encouraging learners to develop a better understanding of more advanced words and collocations. One of the primary aims of this study is, therefore, to examine how to better support Saudi students in their L2 learning and vocabulary development, particularly when it comes to semantically non-transparent vocabulary.

1.3.2 Vocabulary Learning with the Aid of Corpora

Vocabulary learning remains of significant interest to schools, educators, and researchers. There are many studies and reviews that have explored the area of vocabulary learning (e.g., Schmitt, 2008; Nation, 1990) and some studies have considered vocabulary learning with the aid of technology (e.g., Abraham, 2008; Nakata, 2011; Plass et al., 1998). A bilingual corpus was also utilised for learning vocabulary such as verb-noun collocations in Reynold's study (2015), and non-congruent collocations in Alharbi's study (2017).

There have, however, been different views on implementing online tools in L2 instructional classrooms (e.g., Abrams, 2003; Yoshii, 2006). Researchers, for example, argue that the question should not be limited to whether technology-based instruction is useful, "but rather under what conditions and for whom" (Chun, 2016, p. 107). This means that studies should consider certain factors when using technology in L2 instruction such as individual differences, materials, learning tasks, and cognitive processes (Chun, 2016).

The idea of using a corpus, in particular, has only been considered in recent years for L2 learners in a Saudi context, and thus studies involving EFL learners' use of corpora in Saudi classrooms is scarce (Pérez-Paredes et al., 2011). In fact, Youssef and Omar (2020) that as of 2016, there was no prominent research on the use of corpora in vocabulary instruction in Saudi Arabia,

but only a few studies that discussed "the idea of forming specialised corpora or checking or using corpora for linguistic analysis" (p. 2). However, soon after, scholars began investigating vocabulary learning employing corpora in Saudi classrooms (e.g., Alharbi, 2017; Alruwaili, 2018; Youssef & Omar, 2020). Alruwaili (2018) discovered that in relation to Saudi, teachers' attitudes - whether they are ignorant of corpora or need persuading that the use of corpora is worth investing their time and effort to comprehend its various applications - can be a major factor in the lack of adopting corpora in L2 classes. Although using a corpus to develop vocabulary knowledge is relatively new in Saudi Arabia there is existing research that supports using corpora in L2 vocabulary learning (Alharbi, 2017; Reynold, 2015) on the grounds that corpora can allow learners to be exposed to authentic data and understand how language is used today and in different contexts. Inspired by these findings, this study used a parallel corpus to investigate its effectiveness for EFL students in Saudi Arabia.

1.4 Contrastive Analysis and Translation (CAT)

Contrastive analysis and translation (CAT) may be defined as "the kind of instruction which leads to learners' understanding of the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system" (Laufer & Girsai, 2008a, p. 696).

Within the English Language Institution (where the present study took place, see section 3.5), translation already exists as a common classroom practice, as I have learnt from my role as an L2 instructor in that context. However translation is used in a quite basic fashion (i.e., literally translating an L2 word to an L1 word). To be precise, if a student would ask for a meaning of an unknown word, the teacher would utter the translation in the L1 and the lesson would proceed without further discussion. In other words, there is no principled method for explaining fully the

vocabulary item as described in the definition of CAT above. That approach seems especially problematic for semantically non-transparent vocabulary items because they are more complex and require more than word-to-word translations. Learners would need an explanation of the form and meaning of these items and a clarification concerning use of the vocabulary items in context. It is also important that learners are not only provided with a translation from L2 to L1 but rather are also offered translations from the L1 to the L2, as this would enhance both their receptive and productive knowledge of the target items (this is further discussed in 2.4.2). Therefore, it was considered logical to adopt CAT teaching based on Laufer and Girsai's (2008a) work in order to extend current practice and investigate its impact on the learning of semantically non-transparent vocabulary items.

Furthermore, Laufer and Girsai's (2008a) work refers to Lexical Focus on Form that is, communicative activities, and is primarily situated within the context of incidental vocabulary learning. However, when non-communicative follow-up activities (i.e., reading comprehension, fill-in-the-gaps, and translation tasks) are employed, these activities help language learners focus on the target vocabulary. As a result, they found that learners who received CAT instruction outperformed the other treatment groups in vocabulary learning. Studies by Zhang (2018) and Alharbi (2017) also showed that L2 learners who received CAT vocabulary explanations outperformed other groups that did not. Using a mixed methods approach, Zhang's (2018) study analysed CAT teaching for vocabulary learning through listening tasks and listening assessments followed by stimulated recall interviews (SRI). He found that the two groups that received CAT approaches combined either with codeswitching or contrastive Focus-on-Form, scored better on the immediate vocabulary tests compared to the third group that received L2 explanations only. Alharbi (2017) conducted a corpus research study involving data driven learning (DDL), CAT

instruction, and non-congruent collocations. Three groups participated and received instruction, namely corpus study with CAT (+DDL +CAT), only CAT (-DDL +CAT), or only corpus (+DDL -CAT). The (+DDL +CAT) group outperformed the other two groups in both the immediate and delayed post-tests for receptive and productive knowledge.

Within these CAT studies, some aspects were not covered. For example, a corpus was used in Alharbi's (2017) study, but there was no qualitative data (e.g., stimulated recall interviews (SRI) or think-aloud (TA) interviews) to understand students' thought processes. Due to a lack of time and space, she did not conduct a follow up interview, though she had implied that a TA interview and understanding learners' perceptions (in qualitative or quantitative form) towards corpora would be useful. Nevertheless, among the research that has investigated English vocabulary in different areas in the Saudi context, her study was the first to focus on CAT instruction.

In order to more adequately comprehend Saudi students' L2 vocabulary learning, this study attempts to build on and expand Alharbi's (2017) research. Although she was the only one to address CAT training within the Saudi context, this study seeks to investigate the efficiency of CAT instruction in all areas that she did not address. As such, this study will utilise a mixed-methods approach and attempt to fill in the gaps, particularly in relation to the qualitative aspect.

Furthermore, this study examines the value of non-communicative activities within CAT instruction for Saudi students in the hope of implementing effective vocabulary learning strategies in Saudi schools and universities that align with the government's goals. To date, only one study, that of Alharbi (2017), has looked at the advantages of these activities with corpora in a Saudi context. In fact, by using corpora, her study draws on the Noticing hypothesis, which suggests that an L2 word's saliency can be improved using CAT instruction, increasing the probability that it will be learnt better. As will be discussed in section 2.6, a key element in vocabulary learning is

'noticing' (Schmidt, 1990), for which learners' attention needs to be drawn towards the item to be learnt, and in different contexts (Schmitt, 2008). Noticing is arguably facilitated through parallel bilingual corpora (Alharbi, 2017), because the parallel translation allows direct L1/L2 comparisons, discussion and highlighting, and hence more saliency, in more than one context. These methods are all adopted in the intervention reported on in this study. Alharbi (2017) stresses the significance and effectiveness of corpora and corpus resources "especially parallel bilingual corpora, for providing the EFL learners with the necessary salience and multiple exposures needed for collocation noticing and intake." (p. 195). This study also explores tasks (with and without a corpus) drawing on the Noticing hypothesis. For instance, when using CAT instruction, I raised awareness of the similarities and differences of both languages. This was done through a brief explanation and highlighting the target items.

Moreover, this study draws on the involvement load hypothesis (ILH) (Laufer & Hulstijn, 2001) as a theoretical framework. The ILH provides an accurate prediction of the relative impacts of tasks on vocabulary learning and can be reported through the calculation of three components; that is, need, search, and evaluation within the involved task (Yanagisawa & Webb, 2021). It has been suggested that the higher the involvement in a task, the better the learning outcomes will be for students (Laufer & Hulstijn, 2001). The three groups in the study had differences in involvement load, with the Control group having the lowest load and the CAT+ Corpus group having the heaviest load. The CAT + Corpus condition was expected to bring better learning compared to the CAT only group, as it involved a heavier 'load', and the CAT group was expected to perform better than the Control group. The relevance of the involvement load hypothesis is further discussed in section 5.2.3 in relation to the findings of the study, which identified that the CAT + Corpus group had the heaviest involvement load in terms of need, search and evaluation,

the CAT group had a moderate involvement load and the Control group had the lowest involvement load (see Section 2.4.1)

1.5 Significance and Outcomes

As an EFL instructor in Saudi Arabia, the current researcher's motivation for exploring the teaching and learning of semantically non-transparent words and collocations is driven by: (1) an inquisitive desire to understand the Saudi EFL context in relation to vocabulary learning; (2) the importance of vocabulary knowledge in learning a language; (3) the challenges of developing EFL students' lexical and collocational knowledge (e.g., L1 transfer).

The current study, therefore, explores CAT instructional methods, particularly in relation to vocabulary learning, as only a limited number of empirical research studies have been conducted in this area, mainly with positive outcomes (Alharbi, 2017; Laufer & Girsai, 2008a; Zhang, 2018). Furthermore, while Laufer and Girsai (2008a) called for the CAT instructional method, which involves interlingual comparison between L1 and L2 and translation to raise learners' awareness in vocabulary learning, using CAT with Computer Assisted Language Learning (CALL) tools, such as corpora, is very limited. Therefore, by using both methods combined, the researcher hopes that the outcomes of this study will also be enlightening and will positively contribute to the vocabulary learning of ESL in general and of Arabic speaking English learners in particular.

The current study adopted two different methodological approaches. The first approach was the gathering of quantitative data through pre- and post-tests. It also included a comparison between CAT instruction and CAT instruction with a bilingual corpus. The second approach involved collecting qualitative data through concurrent reports i.e., TA interviews, in order to understand the learners' thinking process, and what second language learners contemplate while

completing a vocabulary task such as fill-in-the-gaps. This was followed by interview questions to gain a more insightful understanding of the learners' perceptions of the corpus and CAT instruction. The outcomes of both approaches have provided important insights into how L2 learners can maximise the learning of semantically non-transparent words and collocations (form and meaning recall). The implications of CAT instruction and learning L2 vocabulary on Saudi students will also be discussed in detail.

The study makes a significant contribution to the following research fields: CAT instruction, vocabulary learning, EFL in non-English speaking countries and the use of corpora as tools for facilitating vocabulary teaching. This study also specifically researched participants' learning of semantically non-transparent vocabulary, which provides an interesting contribution to the broader area of vocabulary learning. Before undertaking this research, in my own experience as a practitioner, I noticed that even advanced students relied on a limited range of adverbials and collocations. This seemed to suggest that they needed to develop their vocabulary in this particular area. My observation therefore prompted me to explore ways in which the learners' vocabulary could be expanded. In particular, I wanted to focus on semantically non-transparent words and collocations, as these are more challenging to learn. As previously mentioned, research on vocabulary instruction using CAT and corpora suggested that both could be helpful in expanding vocabulary. I therefore decided to explore the effectiveness of both intervention methods in this study. Furthermore, in addition to the quantitative data gathered from the pre- and post-tests used to evaluate the participants' learning of the target items used in the study, the inclusion of the think-allowed interview allowed me to gain better insight into the participants' thought processes while completing the tasks.

1.6. Research Aims and Questions

By examining the impacts of CAT instruction, this research study attempts to address the following:

1.6.1. Research Aims:

From a learning perspective, this study aims to:

1) Investigate the impact of CAT instruction on EFL learners' receptive and productive knowledge of words and collocations (form and meaning), specifically in a Saudi context.
2) Investigate the efficacy of a corpus-assisted CAT approach for semantically non-transparent words and collocations in raising learners' awareness of vocabulary learning, and to compare the results with CAT instruction led by teachers.
3) Critically assess the learning outcomes of L2 students by comparing knowledge of words and collocations at pre-tests and post-tests.
4) Understand the learning process of L2 learners and their learning of semantically non-transparent words and collocations (form and meaning).
5) Conduct think-aloud interviews to investigate the process of learning in light of using CAT, and to better understand CAT learning outcomes and explore learners' strategies and thoughts while completing vocabulary tasks.
6) Explore learners' perceptions of CAT instruction led by a teacher.
7) Explore the perceptions of learners of a corpus-assisted CAT approach.

From a teaching perspective, this study aims to:

8) Statistically understand the impacts of CAT instruction, both with and without using a bilingual corpus, and to determine which of the two approaches is more effective in teaching non-semantical transparent words and collocations.
9) Make recommendations on the teaching of L2 words and collocations of semantically non-transparent words and collocations using CAT instruction.

1.6.2 Research Questions:

RQ1*: What is the impact on short-term learning of semantically non-transparent words and collocations (form and meaning recall) among high-intermediate learners of English of:*

(a) a contrastive analysis and translation (CAT) instruction by the teacher?

(b) a combined approach using both a parallel corpus and direct teacher CAT instruction?

(c) how do the previous two approaches compare with an approach using neither CAT nor a parallel corpus?

RQ2: *What strategies do learners report using to understand semantically non-transparent words and collocations in a think-aloud task?*

RQ3: *What are the views of the learners on the teaching they have received? Do these views illuminate the learners' strategy use and how well they did in the tests*

1.7 Thesis Structure

This Introduction chapter has established the basis of the current study by first providing an overview of vocabulary knowledge in L2 contexts generally, and in a Saudi L2 context, more specifically. By narrowing the scope, the chapter then focused on the problems facing vocabulary

learning, which has a primary objective of enhancing English proficiency for Saudi students. A general overview of relevant theories and empirical studies on vocabulary learning was covered, and the rationale behind this study was explored. Finally, the significance and research outcomes followed by the research questions of the current study were identified.

Following the Introduction, fundamental issues such as what a word is and what it means to know a word - from the perspective of vocabulary learning - have been explored in more depth in chapter two, the Literature Review. This chapter also includes a thorough discussion theories of learning L2 vocabulary. The chapter subsequently discusses vocabulary recall and empirical studies, which led to the identification of gaps in the literature and the formulation of the research questions.

Chapter three, the Methodology chapter, outlines the research design and rationale for adopting a mixed-methods approach. The research instruments and sampling techniques are described. The reliability and validity of the study are also presented here, while details of the pilot phase being provided, followed by a discussion of its implications for the overall study. Finally, procedures for data collection are outlined and ethical concerns are raised.

The findings for each research question are addressed in chapter four, the Findings chapter. The first research question, which focuses on vocabulary recall, will be addressed using quantitative data. The second question will explore the strategies that students use when completing a task, and the third question deals with the learners' perception of the vocabulary instruction provided by the teacher.

The results of each research question will be reviewed in chapter five, the Discussion chapter, with reference to the theories and empirical background covered in the Literature Review. The

analysis in this chapter will examine the significance of the results, and possible explanations will be given when results of the present investigation vary from earlier empirical evidence.

Chapter six, the Conclusion chapter, provides an overview of this study and a summary of the most prominent research findings. The limitations of the study will then be examined, while a discussion of the study's instructional implications will then follow. Finally, last remarks will be made, emphasising once more the importance and contribution of the current study. Relevant ethical forms, materials, and tests will be placed in the Appendix section.

Chapter Two

Literature Review

2.1 Knowing a Word

It has been stated that the average native-speaking, literate adult, has at their disposal language containing "semantic, pragmatic, stylistic, collocational, syntactic, categorical, morphological, phonological, articulatory and orthographic features" (Hulstijn, 2001, p. 259). Grasping the significance of what is typically counted as a word is of paramount importance. In the eyes of Nation (2013), there are two categories in knowing a word, receptive and productive knowledge. When it comes to receptive knowledge, language comprehension is a result of reading and listening activities. Contrary to receptive knowledge, productive knowledge is closely associated with productive skills – that is, writing and speaking. It contributes to the delivery of meaning by producing language in either the written or spoken form. Indeed, a significant number of linguists to whom vocabulary learning is of interest, classify vocabulary knowledge on the basis of productive writing and speaking skills, which are sometimes referred to as active, and receptive reading and listening skills, which are at times referred to as passive (Maskor & Baharudin, 2016; Schmitt, 2010).

Earlier studies have argued that most vocabulary items are learnt receptively via listening and/or reading rather than productively through writing and/or speaking (Jenkins et al., 1984; Nagy, et al., 1987). Moreover, Schmidt (2010) places great emphasis on the lack of a clear relationship between the different aspects of vocabulary knowledge, and elaborates that learners show a higher level of capability to acquire receptive rather than productive knowledge. Having taken the same position, Nation (2013, p. 371) considers receptive tests to be common in learning L2 vocabulary, arguing that "most tests are created especially to measure L2 receptive vocabulary

knowledge", as receptive knowledge tends to grow more quickly than productive knowledge. Thus, Nation (2013) asserts, the volume of receptive vocabulary knowledge will always be greater than that of productive knowledge.

Given that reception precedes production, and receptive vocabulary knowledge usually exceeds productive vocabulary knowledge, Nation (2001) provides four main factors that differentiate between receptive learning and productive learning: the amount of vocabulary knowledge, practice, access, and motivation. Nation (2001) points out that, in general, receptive learning of vocabulary appears to be easier than productive learning. One of the reasons for this is attributed to the fact that for receptive use, learners might only need to know "a few distinctive features of the form of an item: while for productive use the "knowledge of the word form has to be more precise" (Nation, 2001, p. 28). This could be particularly apparent when a learner's L1 and target language are very different from each other (e.g., Arabic and English). Second, classroom practice (e.g., L2 learning tasks) it is receptive rather than productive use that typically receives more practice and training (Melka, 1997; Nation, 2001). Third, there is the access factor, which indicates that while one-on-one meaning translation is only needed for the L2 to L1 receptive use, other parts of knowledge, particularly correct grammatical forms, collocations, and even culture and context, are required for L1 to L2 productive use. The fourth difference is motivation, which some may argue is not directly associated with vocabulary knowledge. Nation (2001), however, emphasises that vocabulary knowledge may fail to be productive if the learner is not motivated enough to use the vocabulary that he/she has already learned.

To further understand the relationship between receptive and productive tasks and how they both contribute to vocabulary learning, Webb (2005) conducted two experiments that yielded interesting findings in relation to productive learning. He compared two experiments that had

similar reading tasks (receptive) and writing tasks (productive). In the first experiment, participants were divided into two treatment groups: a receptive group, in which learners had to understand the meaning of unfamiliar words from a reading passage comprising three glossed sentences, and a productive group, in which learners were instructed to learn the words by writing each word in an English sentence.

It is important to note that the tests in this experiment had a controlled time of 12 minutes, which was based on the duration it took the slowest participant in the pilot study to complete the productive writing activity. Additionally, participants were informed of these follow-up tests. Ten tests were conducted in total, measuring the participants' understanding of orthography, association, syntax, grammatical functions, and meaning and form. The two treatment groups' results revealed an overall significant gain in all areas of vocabulary knowledge, except for grammar functions. Nevertheless, the findings indicate that the learners scored more highly in their reading test compared to the productive test.

The second experiment involved the same procedure, yet with three exceptions. First, there was no time constraint. Second, participants were not informed of the follow-up test but were just instructed to start with the productive task after completing the receptive task. Finally, unlike the first experiment, participants here were not allocated into different groups, and they were required to complete both the receptive and productive tasks. The findings for experiment two revealed that learners showed much greater vocabulary gains from writing compared to reading, and that receptive learning is more effective when learners engage with productive tasks. Furthermore, Webb concluded there was a strong correlation between productive tasks and productive learning, and that this was not the case with receptive tasks and receptive learning. He explains that the productive tasks were more effective on the receptive measures since students were likely to have

spent more time writing rather than on reading activities. Webb also explained the differences in test scores for both experiment one and two, showing that the group in the first experiment had considerably higher scores than the second group. This is because the first experiment just required the group to learn ten target items, while the second group had to learn 20 target items. The demands for the second group were therefore much higher. The results suggested that learning a larger number of words in a short period of time may have negative effects on vocabulary learning.

Uchihara et al. (2022) also looked at the relationship between vocabulary learning in receptive and productive knowledge and particular learning strategies, namely guessing from context and dictionary use. The study found that there was a relationship between consulting dictionaries and improved receptive vocabulary knowledge. The findings suggested that consulting a dictionary could support learning of receptive vocabulary knowledge, as it allows the learner to confirm whether the meanings of unfamiliar words, which can be inferred from context, are correct. This strategy helps the learner to acquire the correct meanings and prevents the retention of misinferred definitions derived from incorrect inferencing. Dictionary consultation can benefit learners more than guessing, as it draws learners' attention to the form of a new word, improves retention, and can help learners infer vocabulary meaning from context by requiring them to focus on the semantic information and the context in which unfamiliar words appear. The study employed a survey-based approach and found that there was no significant relationship between guessing and productive vocabulary knowledge, yet dictionary consultation was associated with improved receptive vocabulary knowledge, which also resulted in indirect productive vocabulary learning.

Knowledge of receptive vocabulary is generally understood as the ability to recognize a word by its form (Laufer et al., 2004); and understand its meaning (Webb, 2008) and is displayed

by being able to translate the word into the L1 or find a synonym (Webb, 2009). Productive vocabulary knowledge, on the other hand, refers to the ability to retrieve the L2 meaning and form of an L1 word (Laufer et al., 2004; Webb, 2008) or to use the appropriate word based on its equivalent meaning in the L1 (Webb, 2009). Productive vocabulary knowledge further comprises recall of the L2 orthographic and phonological forms, which can be tested through a variety of tasks such as sentence completion (Laufer & Nation, 1999), association of words (Meara & Fitzpatrick, 2000), translation from the L1 to the L2 (Webb, 2008), or naming pictures (Uchihara, 2022). Productive vocabulary knowledge can be categorized into controlled and free productive knowledge (Laufer, 1998). While controlled productive knowledge refers to learners' capacity to produce words when prompted, such as in the above-mentioned tasks, free productive knowledge involves the spontaneous use of a word without giving the learner any particular cues, such as in the case of free composition.

The relationship between the development of receptive and productive vocabulary knowledge is complex and there is a debate concerning the importance of recall and use as facets of productive vocabulary tasks. Teichroew (1982) argued that there is inconsistent use of the terms receptive and productive when referring to knowing a word and suggests that it would be more appropriate to regard it as a scale of knowledge. However, the distinction between receptive and productive knowledge does not necessarily need to be understood in this way. For example, Meara (1990) views the distinction between receptive and productive vocabulary as the outcome of a word association, suggesting that productive vocabulary can be activated by other words, while receptive vocabulary can only be triggered by external stimuli. A criticism of this understanding of the divide between receptive and productive knowledge lies in the argument that language use is not only associationally driven, but is rather is driven by meaning (Nation, 2013). Corson (1995)

further suggests that learners may not always make use of the full range of vocabulary knowledge that they possess. He argued that receptive vocabulary includes three kinds of productive vocabulary: "words that are only partly known, low-frequency words not readily available for use, and words that are avoided in productive use" (p. 44-45). These three categories of words would also overlap to a certain extent. This distinction between receptive and productive vocabulary strongly draws on the concept of use as opposed to degrees of knowledge. Learners may know some receptive vocabulary very well but may choose not to use it and hence some receptive vocabulary never becomes productive (Nation, 2013). Furthermore, the development of receptive knowledge is not separate from productive knowledge. As previously mentioned, Webb's (2005) study found that learners showed greater vocabulary gains from engaging in productive writing tasks compared to reading. While the study found a strong correlation between productive tasks and productive learning, this was not the case between receptive tasks and receptive learning, suggesting that, while receptive tasks were helpful, productive tasks were more effective in supporting the development of receptive learning. This does not mean that productive knowledge should be viewed as more important, but rather it exposes the interconnectedness of receptive and productive knowledge and how the two skills can support each other. Rather than viewing receptive and productive knowledge as two separate ways of acquiring vocabulary, it is important to understand how one kind of knowledge complements the other.

Henriksen (1999) draws particular attention to three continua, which he calls three *dimensions of* vocabulary knowledge. The first aspect is *partial*, which includes the level of meaning in a word. The second aspect involves *deeper vocabulary knowledge*, which involves the development from having a basic understanding of the word to having a thorough understanding of the word. The third aspect is *acknowledging receptive and productive continuum.* This is when

a word's representation is either receptive or productive knowledge. To clarify Henriksen's (1999) three dimensions more, it is similar to finding out the exact meaning of words, then developing depth in understanding a word and finally, reaching a level of understanding of these words in receptive and productive knowledge.

2.1.1 Learning Words: Form, Meaning, and Use

Considering both receptive and productive vocabulary knowledge, attention should also be paid to the three different aspects of knowing a word, which are classified as *form, meaning,* and *use*. Nation (2013, p. 49) summarised in detail what is involved in knowing a word by focusing on all three aspects within receptive and productive knowledge, which is presented in Figure 2.1 below:

Figure 2.1 *What is Involved in Knowing a Word? (Nation, 2013, p. 49)*

Form	**spoken**	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	**form and meaning**	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	**grammatical functions**	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use (register, frequency…)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note: R = receptive knowledge, P = productive knowledge.

It should be emphasised that for the purposes of the current thesis, Nation's (2013) model for the concept of knowing a word will form the theoretical underpinning of this study. The focus will be on the three aspects: *form, meaning* and *use*. For example, *form* will focus on what the word looks like and on writing (i.e., spelling) the target item, *meaning* will emphasise understanding the concept and which target item refers to the correct concept, and *use* will be on choosing the correct target item within context.

2.1.2 Learning Collocations

Terms such as collocations, multi-word expressions, formulaic sequences, and lexical chunks are often used interchangeably. Lexical chunks are phrases, and they are made up of conventional grammar and lexis (Becker, 1975). Where language learning is concerned, mastering collocations and using them correctly in context can be quite challenging for students. Also, both listening and reading skills are impacted by the development of collocational learning (Brown, 1974; Hornby, 1974). These challenges can be based on the nature of learners' L1 (e.g., Granger, 1998; Nesselhauf, 2003). Nesselhauf (2003) conducted a research study on the use of verb-noun collocations, such as *take a break*, in participants' free writing. Her findings revealed that advanced German learners of English were found to encounter challenges in acquiring English collocations that were not similar to German collocations. Granger (1998) explained that French learners of English used collocations in amplifiers (such as *completely*, *totally*) far more than English native speakers. She also pointed out that most of the English collocations they used were from direct translation from their native language.

Another challenge is that collocations are either neglected in classroom instruction (Bahns & Eldaw, 1993; Lewis, 1993; Schmitt, 2000) or are taught poorly (Bahns & Eldaw, 1993). Classifying lexical chunks into five groups as: words (e.g. *paper, table*), polywords (e.g. *by the way*, *so long, in spite of*), collocations or word partnerships (e.g. *community service*, *absolutely convinced*), institutionalised utterances (e.g. *nice meeting you*, w*ould you like a cup of coffee?*), and sentence frames and heads (e.g. *the fact/suggestion/problem/danger was* . . .) or even a sixth as text frames (e.g. *on the one hand, on the other hand, not to mention*) (Lewis, 1997a), Lewis (2000) asserts that they should be taught explicitly in classrooms as learners cannot notice them automatically.

The lexical approach by Lewis (1993) gives priority to lexical chunks over individual words. It argues that they are easier to not only recall but also to use. This is an advantage over the attempt to combine parts of words together. Indeed, "instead of words, we consciously try to think of collocations, and to present these in expressions. Rather than trying to break things into ever smaller pieces, there is a conscious effort to see things in larger, more holistic ways" (Lewis, 1997b, p. 204). Taking the same position, Schmitt (2000) points out that collocations are stored in the mind of learners as a whole, so these ready-to-use units can be recalled as single pieces of information. In addition, instead of creating and assimilating new words each time, these formulaic sequences are stored and retained as shortcuts that can be retrieved when needed, as one single unit (Wray & Perkins, 2000). Learners are likely to learn a collocation as a single piece of information rather than many separate words with different linguistic components (i.e., syntax, semantics, morphology, phonology, and pragmatics), which can overload the working memory (WM) according to the cognitive load theory (Cooper, 1998), discussed in section 2.3.3. Therefore, this whole unit of learning becomes less reliant on cognitive and memory capacity.

From an empirical standpoint, Lindstromberg and Boers (2008) explored three different experiments on alliteration in collocation (for example *daydreaming*) in L2 classrooms. The purpose of the study was to understand ways to assist L2 learners in learning collocations. The three experiments were carried out on memorable phrases, noticing lexical chunks, and teachers' brief intervention. The first experiment aimed to explore whether L2 adult language learners were able to remember alliterative or non-alliterative lexical chunks. In both the post and delayed tests, words with phonological repetitiveness (for example *daydream* and *fast food*) were found to be more memorable than phrases with no phonological repetitiveness (for example *phone call* and *fresh air*).

Furthermore, the main aim of the second experiment was to investigate whether 31 language learners in different language majors could spontaneously notice alliteration of lexical chunks (e.g. *green grass and fast food*) and identify other patterns of rhyme (e.g. *name game* and *deep sleep*), of assonance (e.g. *nice try* and *townhouse*), and of no repetition (e.g. *new car* and *brick wall*). Indeed, there was a deliberate reason for choosing the participants; they were fluent in English, according to the *Common European Framework of Reference* (CEFR), and owing to their discipline and major, they were more likely to have "an above-average language awareness and above-average sensitivity to linguistic phenomena in general" (Lindstromberg & Boers, 2008, p. 213). The participants were given five minutes to sort different pattern sets, one of which had a "pleasant sound pattern" (Lindstromberg & Boers, 2008, p. 214), while the other three patterns did not. Most learners were found to fail in sorting the four patterns. Thus, it may be argued that learners need assistance in noticing different lexical chunks and even less noticeable words that share phonological repetitions.

Last but not least, the third experiment aimed to investigate whether it is useful for classroom instruction to explicitly draw learners' attention briefly (for seconds) to salient patterns of lexical chunks. This was a quasi-experimental study with an experimental group and a control group of college students who had taken the same English courses for two years. The participants were exposed to a three-month period of listening to authentic materials, with the experimental group experiencing brief interventions by the teacher to draw their attention to phonemic repetition in lexical chunks. Both of the groups were given post-tests without previous notice. The tests focused on seven alliterative lexical chunks (*play a part, put to the test, clear-cut, from pillar to post, pink and perky*, *mind over matter*, and *short shrift)*; one rhyming phrase (*wear and tear*); and 12 phrases involving non-salient phonological repetitive items - that is, phrases made up of words

with different phonological sounds initially in each word (*state of the art, over the counter, that's the long and short of it, wreak havoc, clean/empty slate, ethnic cleansing, turn a blind eye, toss/flip of a coin, in the wake/ aftermath of, cutting-edge, the thin end of a wedge, burned at the stake*). The experimental group was found to outperform the control group in recalling the salient lexical chunks directed by the teacher.

However, the difference between the two groups in relation to the salient lexical chunks was not statistically significant. In fact, the control group performed slightly better than the experimental group on non-salient phrases. This could be because the control group might have been familiar with or had background knowledge of the target items, as revealed in their EFL exam scores. Accordingly, the control group had an initial advantage over the experimental group. If indeed, the experimental group students were weaker than the control group, then, it is possible that "the brief interventions by the teacher were worth the minimal effort" (Lindstromberg & Boers, 2008, p. 218).

2.1.3 The Role of Learning Strategies in Vocabulary Learning

Learning strategies play an important role in vocabulary learning. Acquiring vocabulary items in a new language can present quite a challenge and therefore it is important that learners adopt effective learning strategies. Language learning strategies (LLS) are understood as a process which is "consciously selected by learners and which may result in action taken to enhance the learning or use of a second or foreign language, through the storage, retention, recall, and application of information about that language" (Cohen, 1998, p. 4). LLS have drawn interest from researchers since the 1970s, for example by trying to identify effective learning strategies adopted by successful learners to enhance their second language (L2) (see Rubin, 1975). Since then, a number of studies have explored various issues in the context of vocabulary

learning strategies (see Fan, 2003; Gu & Johnson, 1996; Schmitt, 1997, 2000). Gu and Johnson (1996) identified two types of strategies, which they classified as metacognitive and cognitive and each type comprises several related strategies. Within the metacognitive category, he identified the strategies such as self-initiation and selective attention and cognitive strategies included, for example, contextual guessing and use of dictionaries for learning new words. Schmitt (1997) further identifies two categories, described as: a) discovery strategies, which refer to ways to learn new words, and b) consolidation strategies, which include strategies adopted to consolidate the learning of words already acquired. *Discovery strategies* are subdivided into two classifications: determination and social strategies, while *consolidation strategies* have four classifications: memory, cognitive, metacognitive, and social strategies. According to Schmitt (1997), determination strategies refer to when learners independently discover the meaning of new words, for example by drawing on the context, structural knowledge or by using reference material. Social strategies, on the other hand, require collaboration with others. Learners may ask others the meaning of unknown words. Learners may use social strategies to discover the meaning of unknown words and also to practise the meaning of newly acquired vocabulary.

Fan (2003) further proposed an additional categorization, which includes nine categories of strategies: management, sources, guessing, dictionary, repetition, association, grouping, analysis, and known words. More recently, Nation (2013) proposed a taxonomy of Vocabulary Learning Strategies (VLS) consisting of four categories of learning: *planning, source strategies, processing strategies* and *skills in use*. *Planning* refers to the learner's ability to determine what to learn and when (e.g. selecting words and appropriate strategies). *Source strategies* refers to identifying information about the words, such as guessing the meaning of a word from its context. *Processing strategies* includes mastering the vocabulary knowledge acquired, for example through

detection and retrieval and *skills in use* refers to the strategies adopted through input (reading and listening) and output (speaking and writing) and developing these four language skills.

Table 2.1 A taxonomy of vocabulary learning strategies (Nation, 2013, p. 328)

General class of strategies	Types of strategies
Planning: choosing what to focus on and when to focus on it	Choosing words Choosing the aspects of word knowledge Choosing strategies
Sources: finding information about words	Analysing words Using context Consulting a reference source in L1 or L2
Processes: establishing knowledge	Noticing Retrieving Generating (creative use)
Skill in use: enriching knowledge	Gaining in coping with input through listening and speaking Gaining in coping with output through reading and writing Developing fluency across the four skills

Nation's (2013) VLS is further discussed in chapter 5, as it informs the interpretation of the study's findings. The following section discusses the literature on vocabulary testing.

2.1.4 Vocabulary Testing

In relation to knowing words and collocations, it is vital to point out that there are different ways of measuring vocabulary knowledge through tests. Webb (2008) referred to two instruments for vocabulary testing, namely the Vocabulary Levels Test (Nation, 1990) for receptive knowledge, and the Productive Vocabulary Levels Tests (Laufer & Nation, 1999) used for productive

knowledge. He described the former test as a matching test where one of six answers should be selected to provide the correct meaning for the three given definitions. The latter test involved cued recall testing that includes the first letter of the target item in a sentence and participants are required to complete these target items. He found these testing formats (i.e, Vocabulary Levels Test knowledge and the Productive Vocabulary Levels Tests) to support receptive knowledge more than productive knowledge, suggesting some flaws within the tests. For one, in the Vocabulary Levels Test, guessing correctly without knowing the target items is possible, whereas there is no chance of guessing in the Productive Vocabulary Levels Test. Second, the Vocabulary Levels Test assesses knowledge of form and meaning, while the Productive Vocabulary Levels Test includes all three aspects of knowing a word: form, meaning, and use. The latter test requires participants to write the correct spelling and grammatical function of each target item. Third, the Vocabulary Levels Test involves recognition of the target item whereas the Productive Vocabulary Levels Test uses recall. Recognition is easier and allows guessing, whereas recall is more difficult, and thus there is no comparison between them. Furthermore, Webb (2008) makes reference to previous research (e.g. Morton, 1979) which suggests that tests that provide the first letters of the target items are, in fact, testing receptive knowledge, indicating that such tests are more appropriate for testing receptive knowledge, rather than productive knowledge.

For the current study, learning took place through receptive tasks and productive tasks during the intervention, while both receptive and productive knowledge in recall was tested and scored. Learners were required to complete tasks for the think-aloud interviews, which assessed recognition, but were not scored. The following section will cover the issue of semantic and non-semantic transparency which will help in understanding the types of target items selected in the study.

2.2 Understanding Semantic Transparency in One-Word Form and Collocations

In order to provide a greater understanding of the target items used in the current study, this section will review the background on semantic transparency within one-word forms, the constituents of which are either adverbs, such as *thereby* or *henceforth,* or nouns like *upshot* or *outcome*. Furthermore, it focuses on semantic transparency within multi-lexical vocabulary items, such as *on the other hand*. It is important to note that words with multi-morphemic forms (hereinafter referred to as compound words) are normally defined as transparent "when the meaning of the compound word is consistent with the meanings of the constituents (e.g., *car wash*). In contrast, a compound word is defined as semantically opaque when its meaning cannot be constructed by directly combining the meanings of the individual constituents (e.g., *pineapple*)" (Pollatsek & Hyönä, 2005, p. 262). Furthermore, semantically opaque meaning is not only predominantly influenced by language development over time, but also by moving away from the meaning of the constituents (Aornoff, 1976).

In today's English, a very common compound word is *therefore*. There are, however, less commonly used compound words associated with *there*, for example, *thereupon, thereafter, therein,* and *thereby*. In the eyes of Österman (1997), who studied the development of the *there* compounds, early compound adverbs had gone through different semantic developments and changes up until the advent of Modern English; so much so that they became more and more abstract and grammaticalized. In fact, the local meaning - that is, the original meaning of *there* compounds - has changed. For instance, the meaning of the word *thereabout*(s) has changed "from a local about or near that place to stating approximation (or so)" (Österman, 1997, p. 268).

From a psycholinguistic viewpoint, a distinction needs to be drawn between semantically transparent and opaque (non-transparent) compound words, as compound words are characterised

by an inconsistency in morphological representations and processing (Libben, 1998). For example, the word *blackboard* is "routinely activated during word recognition" because its constituents are semantically transparent, whereas the processing of the word *department* would not contain "such constituent activation" and thus, is semantically opaque (p. 31). Libben (1998) finds compounding to often be a productive, useful, and straightforward morphological process, and therefore the chances of encountering new compound words, for example *slushfoam*, are high. He states that they can be "ideal candidates for routinized morphological decomposition" (p. 33). This is largely because they can be easily recognized and understood, as their comprehension is basically achieved through the meaning of their constituents.

However, there are some occasions when *semantic drifts* occur, which is when there is change of meaning over time (Beinborn & Choenni, 2020). Libbon (1998) suggests that compound words are more prone to semantic drift and therefore often show "high degrees of semantic opacity" within these words (Libben, 1998, p. 34). Thus, according to their fundamental degrees of compound, Libben (1998) and Mattiello and Dressler (2018) classify words with compounds into four categories: transparent – transparent (TT), as in *doorbell*; opaque – transparent (OT), as in *strawberry*; transparent – opaque (TO), as in *jailbird*; and fully opaque (OO), as in *humbug*.

Where collocations are concerned, some researchers attach significance to semantic transparency, so much so that it has been considered to be either the only factor or at least one of the most important factors in understanding and differentiating some expressions as idioms and collocations (Cruse, 1986). Kurosaki (2012), for example, investigated the recognition and production of semantically transparent and semantically non-transparent collocations. Two different tasks were implemented respectively: multiple choice questions to recognize

collocations, and translation from L1 to L2. Following these tasks, Kurosaki (2012) investigated the use of English collocations within free writing in language corpora.

The study examined French and Japanese participants learning four lexical categories of collocation namely, [verb + noun], [delexicalised verb + noun], [adverb + adjective], and [adverb + adjective]. The participants received teaching in 71 target collocations that consisted of 1) semantically transparent collocation (literal meaning), example *tell the truth, keep records*, 2) semantically non-transparent collocation (figurative meaning) such as, a*nswer/pick up the phone, bad/ill/poor health,* 3) restricted combination [-ResComb] in a collocation, which, according to Kurosaki's (2012) study, are words like *make/take a decision*, where both *make* and *take* can be combined with the phrase *a decision*. He explains that having more than one option for a verb with a noun means this collocation is not restricted, and 4) restricted combination [+ResComb] in a collocation, which means there is one possible word to combine with other words.

The first task involved a comparison between the restricted combination and semantically transparent collocations, and that of the non-restricted combinations and semantically non-transparent collocations for each of the four lexical categories of collocations. The results showed no distinct difference in the recognition accuracy of both the restricted combination and semantically transparent collocations, and that of the non-restricted combinations and semantically non-transparent collocations for the French and Japanese learners. However, within each lexical category of collocations, i.e. [verb + noun], [delexicalised noun + noun], [adjective + noun] and [adverb + noun], there were significant differences in the restricted combination and semantically transparent collocations, as well as the non-restricted combinations and semantically non-transparent collocations between the responses of the French learners and the Japanese learners.

Furthermore, among the four lexical categories of collocation, the lowest accuracy was found in the [adverb + adjective] category in the interaction between the two language learners.

These differences between the groups suggest that the L1 influence might have taken place among learners. There is also the assumption that French learners' L1 influence might be greater because of the similarities between English and French, although this was only the case in some categories as Japanese learners revealed greater L1 influence in the [adjective + noun] category. For this task particularly, it could be said that L1 influence played a role to some degree, but other factors such as developmental errors, which were unique to each learning group, also had an effect on the learners.

When it comes to Kurosaki's (2012) second task involving translation of L1 to L2, the [verb + noun] and [delexicalised verb + noun] category collocation results show that learners performed better and showed higher accuracy production in their answers in the restricted combination and semantically transparent collocations in comparison to the non-restricted combinations and semantically non-transparent collocations. Interestingly, in the [adjective + noun] category of collocations, the Japanese learners had a higher accuracy (66.5%) in producing the restricted combination and semantically transparent collocations than in the non-restricted combinations and semantically non-transparent collocations (34.6%), whereas the French learners showed a marginally lower accuracy (47.8%) in the restricted combination and semantically transparent collocations than in the non-restricted combinations and semantically non-transparent collocations (49.8%).

In the last lexical category of collocations [adverb + adjective], the results contrast with the previous categories in this task as both learner groups showed accuracy in the responses of the non-restricted combinations and semantically non-transparent collocations in comparison to their

responses in the restricted combination and semantically transparent collocations. In fact, similar to the first task, the overall accuracy of [adverb + adjective] collocation for both learner groups were lower than the previous lexical categories. This could imply that [adverb + adjective] collocation is more complex for language learners.

The second task clearly revealed L1 influence for both French and Japanese learners, although the extent of L1 influence depended on L1 background. For example, the L1 influence was greatly found in the use of prepositions for the French learners, whereas L1 influenced the use of nouns for the Japanese learners. The results of the second task suggest that depending on the language background, L1 can have different influences on L2 semantic outcomes.

Kurosaki's (2012) final finding involved learner corpora of both French and Japanese. The aim was to understand the tendency of collocations written in English essays by French and Japanese learners. Though the target items found in the corpora were small in number, the results showed that both groups of language learners were able to frequently produce collocations which have the corresponding meaning and similar syntax of word order to their L1s. Where the influence of the combinability and transparency of collocations is concerned, results showed that both groups were able to produce a larger number of collocations which were non-restrictive and non-transparent [-ResComb, -Transp] compared to collocations which were restrictive and transparent [+ResComb, +Transp]. This contradicts the results of the second task in the area of combinability and transparency of collocations. Overall, the above outcomes suggest that Kurosaki's (2012) findings varied between tasks, language learners, and lexical categories, which should all be taken into consideration when teaching collocations to L2 learners.

From a different perspective, Ramadhan's (2017) longitudinal study discussed semantically transparent collocations along with other criteria. He examined the learning outcomes of 252 Kurdish EFL learners and employed two tests to measure receptive and productive knowledge and development. Moreover, four criteria were assessed in relation to teaching collocations: semantic transparency, frequency of occurrences, congruency with L1, and syntax. Frequency was found to have the main effect on learners' collocational knowledge and development, followed by syntactic structure, then congruency with L1, while the semantic feature was deemed to have the least impact of all four criteria. Although frequency of occurrences influenced receptive and productive knowledge, participants' productive knowledge increased more than receptive knowledge. The most interesting observation, however, was that the receptive knowledge of the lower frequency collocations appeared to have significantly advanced over the school year compared to that of the higher frequency collocations. This could be explained with reference to the students' emphasis on learning the less frequent collocations—that is, there appeared to be a shift among the participants towards learning more challenging collocations rather than the more frequent collocations.

Another interesting point is that test results of semantically transparent collocations showed that participants' knowledge of semantically non-transparent collocations was significantly better than their knowledge of semantically transparent ones. This could be explained with reference to the fact that idioms and collocations can have figurative meanings and have "varying degrees of semantic transparency," and that "one might assume more transparent meanings are learned before less transparent meanings" (Macis & Schmitt, 2017, p. 324). In fact, such collocations (i.e. semantically non-transparent and figurative meanings) can often be learned before semantic transparent collocations (Kurosaki, 2012; Ramadhan, 2017), while a meaning of less or semi-

transparent collocation can be less easily noticed compared to a non-transparent figurative collocation (Macis & Schmitt, 2017), which can cause problems for L2 learners (Nesselhauf, 2005).

In summary, studying semantically transparent or non-transparent words and collocations has its challenges; that is to say, what appears easy and clear may be complicated for L2 learners. Different aspects need to be considered such as frequency of occurrences, type of learning tool used (e.g. corpora), complexity of lexical categories, and language background (including age and proficiency level) of L2 learners. The aim of this study is to further our understanding of semantically non-transparent collocations in respect to words and collocations.

The following section will discuss issues concerning L2 learning with regards to L1 influence, the cognitive stages and processes that occur within L2 learning, working memory capacity, and semantic fossilisation.

2.3 Issues in L2 Vocabulary Learning

2.3.1 L1 Influence

Using L1 in language classrooms has always been the subject of heated debate among language authorities. In fact, in earlier studies, the use of L1 in second language learning was assumed to have a negative effect; as a result, the terminology L1 *interference* was introduced by Weinreich (1953). In more recent studies (e.g., Laufer & Shamuel, 1997) language researchers have reported that the L1 exerts significant influence, and claimed that the ubiquitous nature of L1

influence prepares the ground for the argument that it is both reasonable and practical to exploit it when it is in our advantage to do so (Schmitt, 2008).

Furthermore, Swan (1997) argued that L1 has a major role in L2 vocabulary learning, and thus language learners should be aware of the fact that although exact and equal translations can exist between L1 and L2, this is not always the case when it comes to "more marginal or metaphorical ways" (p. 158). Indeed, "the mother tongue can support, fail to support or actively hinder someone who is learning or using the vocabulary of a second language" (Swan, 1997, p. 156). Singleton (1999, pp. 189-190) pointed out that both L1 and L2 words have connections either "directly, between individual L1 and L2 lexical nodes, or via a common conceptual store (or both)". In addition, Sunderman and Kroll (2006) asserted that in relation to words that have similar forms in L1 and L2, both languages have some influence, even though only one language is being performed. In fact, L1 is found to be active not only for beginners, but also for more advanced learners (Hall, 2002; Sunderman & Kroll, 2006).

When it comes to learner output, Hemchua and Schmitt's (2006) study revealed that Thai students writing 20 compositions made errors, with approximately one quarter of these mistakes being lexical. Their findings were based on asking students to write an argumentative composition of 300-350 words without the support of dictionaries. The most common errors were words seen in suffix types (e.g., *polluted* instead of pollution), near synonyms pairs (e.g., using *penitent/contrite* criminal instead of the word regretful), and preposition partners (e.g., some channels *in* television instead of using the preposition on). These errors can be attributed to the influence of L1. However, they are also the result of the unique properties of the English language that often lead to these learners' confusion. Examples include the abstractness of words, and some English words being excessively long. of (Schmitt, 1997).

With regards to using translation in learning English as an L2, Hemchua and Schmitt (2006) argue that translating paves the way for the acquisition of both English language skills and systems, particularly words, idioms, expressions, and phrases. Moreover, it is possible to learn more new vocabulary items (form-meaning) with the aid of L1 translations compared to L2-based definitions (Laufer & Girsai, 2008a; Laufer & Shmueli, 1997). Ramachandran and Rahim (2004) argued that by providing easy access to meaning, the use of L1 can facilitate learning the relationship between form and meaning. As a result, cognitive resources can focus more on learning the form.

2.3.2 Stages and Processes in L2 Vocabulary Acquisition

Jiang (2000) proposed a psycholinguistic model of L2 vocabulary acquisition, which draws on the learning conditions unique to such learners. L2 learners are not exposed to as much contextualised input which complicates the integration of lexical meanings and their extraction. Furthermore, adults or adolescents are different in comparison to children in learning language. This is because L2 adults and adolescents already have a well-developed lexical system, for which there is often a corresponding concept or translation in the learners' L1. Therefore, adults do not necessarily need to learn new ideas and meanings, especially in the initial stages of L2 acquisition (Jiang, 2000). According to Jiang (2000), when children acquire new vocabulary in the L1, they are also learning about new concepts, which results in form and meaning often becoming inseparable. Adults or adolescents, however, do not associate L2 vocabulary with conceptual or semantic development. When adults or adolescents acquire new terms in the L2, their learning process actively involves the language and conceptual systems of their L1. Based on this, Jiang (2000) proposed a three-stage model of adults' L2 vocabulary acquisition. The first stage requires learners to identify a word in its phonological or orthographic form. Learners understand its

meaning by linking the word to an existing semantic structure in their L1. Learners associate the word with its translation in the L1 to help them remember the meaning. This process allows the new word to enter the learner's mental lexicon.

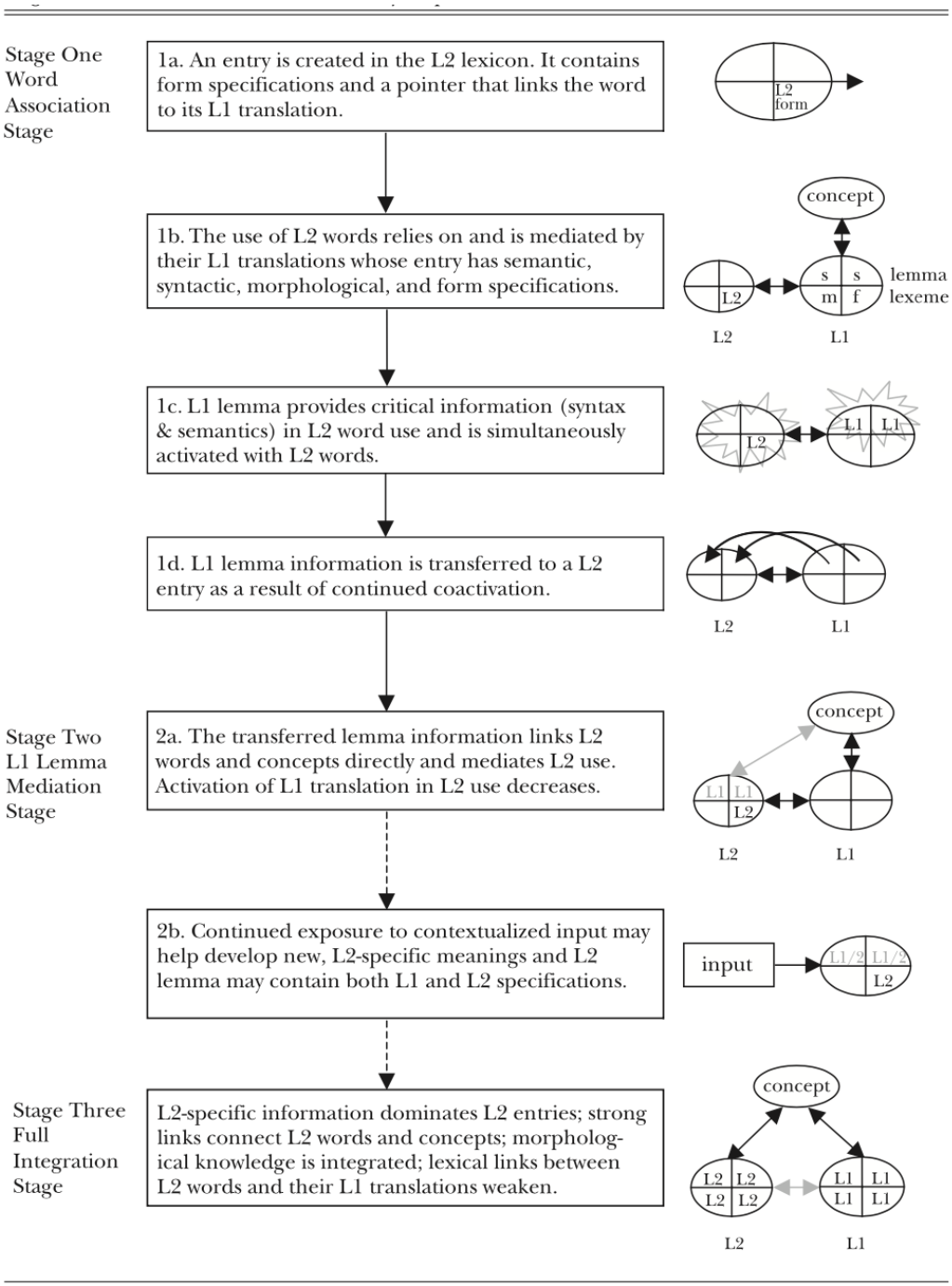
However, unlike words in the L1, whereby the learner associates words with the lemma structure which combines meaning and syntax, and morphology and phonology, and orthography in the lexeme (Levelt, 1989), the new term in the L2 is only understood in terms of *form knowledge* (i.e. phonology and orthography). The L2 entry is also directly associated with its L1 translation. L2 production at this stage relies on recalling the meaning of words in the L1 as no direct link is established between L2 concepts and words, or these links are not strong.

Repeated exposure to L2 word use, however, results in a "continued coactivation" of a word in the L2 and its corresponding L1 translation (Jiang, 2004, p. 417). The transfer of semantic and syntactic information in the L1 translation causes processing and lexical representation of an L2 word to shift. The second stage in lexical development now integrates both L2 form specifications as well as L1 transferred semantic and syntactic specifications, which occurred from L1 translations. The fact that semantic content is now present in the entry means that the L2 word is now directly connected with conceptual representations. One might expect learners at this stage to use L2 words more fluently, because the direct connection to conceptual representations means that activation of L1 translation is not required anymore. However, just as in the case of L2 word acquisition in the first stage, there is still considerable influence from the L1 in L2 word use; so much so that the lemma information in its L1 translation, which is now a component of the L2 entry, can still mediate lexical processing and production. From a processing standpoint, this level is therefore referred to as the L1 lemma mediation stage, but from a representational standpoint, it may be considered as a "hybrid entry-stage" (Jiang, 2004, p. 417). This is because an L2 entry at

this point includes both L1 meaning and syntactic information as well as L2 form. In the third stage of lexical development, the information from the L1 is removed and the lexical knowledge that is specific to the L2 word is included into its entry. The L2 words can therefore be used more fluently, with more automaticity, but also in idiomatic ways with little influence from translation from the L1. Yet, as suggested by Jiang's (2000) model (see Figure 2.2), many words may not reach the third stage and L1 lemma mediation could become a normal state in advanced L2 learners' lexical processing.

Figure 2.2

Stages and Processes of Adult L2 Vocabulary Acquisition (Jiang, 2000 in Jiang, 2004, p. 418)



Jiang (2002) examined the proposition that lexical items in the L2 are often mapped on to their existing semantic context in the learners' L1, rather than onto a new semantic specification of their own. According to the hypothesis for semantic transfer, "content residing in L2 words is transferred" from learners' L1 (Jiang, 2004, p. 420). Therefore, Jiang (2002) explained that when learners are presented with two English words, which share the same translation in Chinese, the semantic representation in the minds of Chinese EFL learners should be identical or very similar. However, given the variance in the Chinese translations, translation pairings should appear less similar. The hypothesis therefore predicts that Chinese EFL learners would provide more accuracy for the same-translation pairs compared to pairs with different translations. If the hypothesis is accurate, Chinese EFL students should also answer more quickly to same-translation pairs because shared semantic elements are easier to identify.

Jiangs' study (2002) required both native and non-native English speakers to complete two semantic judgement experiments, for which they had to establish the degree of relatedness of English word-pairs (experiment 1) or determine whether two English words had related meanings (experiment 2). Non-native speakers were found to respond faster or score higher on L2 word-pairs which shared the same L1 translations, compared to those which did not. These findings thus suggest that L2 lexical entries include L1 semantic information, which plays a significant role in L2 vocabulary acquisition.

Advocates of the L1 lemma mediation model (Jiang, 2000, 2002, 2004) argued that advanced language learners map L2 words onto their first language concepts, while advocates of the revised hierarchical model (RHM) (Kroll & de Groot, 1997) suggested that the more experience in the L2, the more learners "remap the L2 words to their L2 meanings" (p. 482). To further explain, the RHM argues that learning a second language's vocabulary is a developmental

process in which L2 beginners map L2 words with L1 words that exist in L1 concepts using L1 translation. As the learner progresses in L2 and the connection is revisited between an L2 lexical form and its L1 concept through L1 translation, the learner will eventually deactivate the L1 translation and use the L2 word just like natives by linking the L2 word and its L2 concept directly. This is when the L2 semantic concept has finally been incorporated.

These two models have inspired Alshehri's (2021) research. In her study, 26 Arabic L2 speakers and 26 native speakers of English were given 76 semantically related word-pairs and were requested to immediately assess their semantic similarity on a 5-point Likert scale. The findings indicated that word-pairs with similar Arabic translations received ratings that were noticeably higher than word-pairs without similar Arabic translations. The study concluded that the results supported the L1 lemma model because even advanced L2 learners still relied on the L1 in order to make sense of the meaning of the L2 words.

The study identified differences between same-translation pairs and pairs with different translations. Learners rated same-translation pairs higher because of the greater semantic overlap between the words. For instance, there is no difference in meaning between the words *present* and *gift*, which share only one translation in Arabic. Pairs, such as *present/gift,* were rated high on semantic relatedness by non-native speakers, thereby confirming the prediction made in Jiang's lemma model, which suggests that even advanced L2 learners will associate L2 words with their L1 translation. This is because L1 lemma information is transferred and copied into the lexical entry of the L2 word. The different-translation pairs, however, received lower scores because each entry contained different L1 information, which made the semantic relationship for non-native speakers weaker. Alshehri's (2021) study provides evidence in support of Jiang's lemma model (2000, 2002, 2004), although conclusions are not generalisable as they are based on a rather small

study at one university. Nonetheless, the findings help illustrate the relevance of Jiang's lemma model to understanding L2 vocabulary acquisition, and emphasise the importance of the L1 in the learning process.

Furthermore, Alkhudiry (2018) investigated the relationship between L2 English learners' processing of new words in a task and their acquisition and recall of the meaning of those words; and if participants were still mapping Arabic lemma content onto the English L2 form or whether they had begun to construct lemmas with L2 word meanings. The participants in Alkhudiry's (2018) study were English learners from Arab countries, mostly from Saudi Arabia. They were enrolled in undergraduate courses at a UK university. Based on Jiang's (2000) model, Alkhudiry found that the L2 learners could link the L2 meaning to the L2 form in the immediate post-tests. However, these results disappeared one week later when conducting the delayed post-test. This is because L2 students relied more on the Arabic semantic representation of the target words' meaning, which may support the notion that L2 students continue to translate the L1 lemma information onto the L2 form a week later.

When it comes to the processing of collocational knowledge, research shows that congruent collocations are easier to process than the non- congruent collocations (e.g., Wolter and Gyllstad, 2011, 2013). Yamashita and Jiang (2010) followed Jiang's framework applying a recognition and whole-collocation acceptability-judgment test to evaluate the process of congruent and non-congruent English collocations for two groups; the first were 28 advanced Japanese ESL speakers, and the second were 20 native English speakers. With regards to the response times or error rates, native speakers did not reveal any significant difference between the congruent and the non-congruent collocations. However the advanced Japanese ESL speakers showed the same response time for both groups of collocations but had more non-congruent

collocation errors than congruent collocations. The results also suggested that the process of acquiring L2 collocations is influenced by both the LI congruency and L2 exposure. They also found that congruent collocations are quicker to be regarded as legitimate "in the L2 mental lexicon than incongruent collocations" and that when the latter are stored in the memory, " they may be processed as wholes without going through word by- word L1 mediation" (Yamashita & Jiang, 2010 p. 130). The study found that EFL learners made a greater number of errors when dealing with incongruent collocations compared to congruent ones, suggesting that the former were more challenging. These findings have implications for the understanding of how L2 vocabulary is acquired and help illustrate the challenges that learners may experience especially when learning incongruent collocations. This links back to Jiang's L1 lemma mediation model (2000), which focuses on the importance of L1 in the stages of L2 learning.

2.3.3 Working Memory

Working memory (WM) is defined as a unit of components of the mind, which contain a "limited amount of information" that is "temporarily in a heightened state of availability for use in ongoing information processing" (Cowan, 2017). According to Cooper (1998), the capacity of WM can be quite restricted, and there can be a cognitive overload if verbal WM is required to process many elements. He explained his definition of cognitive load as the "total amount of mental energy imposed on WM consumed by an instance in time" (Cooper, 1998, p. 10).

Accordingly, WM can have an effect on vocabulary learning (e.g., Yang et al., 2017). This was observed in Yang et al.'s study (2017) where WM significantly impacted two of their treatment groups; the comprehension only and the gap-fill groups in an immediate post-test. The participants in these two groups learned new words by understanding their definitions and the linguistic contexts in which they were used. This was not the case, however, with the sentence-writing group

where participants were not only required to understand the new words and the linguistic contexts, but also to produce these words in new contexts. It was speculated that the WM of the last group might have been overridden due to the fact that they had to produce new contexts for the use of the newly learned vocabulary.

Ruiz et al. (2021), drawing on Baddeley and Hitch's model (1974), also explored the effects of WM on instructed second language vocabulary learning. WM is understood here as a system that indicates storage subsystems, which hold and process both visual-spatial and verbal information for a short term. If weight is placed on assessing the storage capacity (particularly the phonological loop), then this is often referred to as "phonological short-term memory" (Ruiz et al., 2021, p. 414). However, where emphasis is placed on both elements of storage and processing of WM (the phonological loop plus the central executive), it is referred to as "complex working memory" (Li, 2017 in Ruiz et al., 2021, p. 512). Ruiz et al. (2021) focused on complex (executive) working memory and found that there was a relationship between WM and lexical learning of targeted phrasal verbs, in that WM predicts learning outcomes. The participants in the study read the form-focused content and completed multiple-choice gap-fill exercises that contained the phrasal verbs. According to the study's findings, variables like instructional contexts can influence how WM affects L2 learning. The study also found that WM interacts with instructional treatments. Additionally, the results by Ruiz et al. (2021) are also aligned with the theoretical hypothesis that WM is more important when instruction is more cognitively challenging (e.g., Linck et al., 2014).

When it comes to learning foreign language vocabulary, Sweller (2017) argued that teaching should take into account ways to reduce WM load and that it is very important to make sure that split-attention is avoided. This normally occurs when there is a need for multiple sources

of information to be processed simultaneously. For example, in tasks where L2 learners are provided with vocabulary translations, instruction should remain as close as possible to the task without overloading the learner's WM. If a task is designed in an electronic format, being able to click on a word to view its translation would be helpful. In a paper-based task, the corresponding translation of the words could be indicated with arrows to facilitate the search. Having to access external sources, such as a separate dictionary, may also increase learners' cognitive load. Thus, value should be placed on avoiding redundancy as unnecessary information brings about additional cognitive loads for no good reason. Therefore, it can be argued that "the redundancy effect may lead to the expertise-reversal effect" (Sweller, 2017, p. 9). In his paper, Sweller (2017) explained that *the expertise-reversal effect* is when teachers provide certain tasks for L2 beginner learners such as translation, but these tasks are no longer needed when proficiency increases, as these tasks might become useless and redundant for advanced learners.

According to the above, therefore, the greater the WM consumed by a task, such as a challenging L2 writing assignment, the less mental capacity is left for alternative cognitive tasks. Nawal (2018) explored the implications of cognitive load theory on L2 academic writing in female learners studying for a degree programme in Saudi Arabia. The study found that avoiding the need to access resources in the L1 during L2 writing tasks may result in improved performance. This is because eliminating the requirement to access external sources avoids splitting learners' attention, which would increase the load imposed on the WM (Cooper, 1998).

Another important item to pay attention to is that instructional manipulations may have the potential to pave the way for desirable difficulties (DD) (Bjork, 1994). Simply stated, certain kinds of difficulties can be created to not only improve learning, but also to slow learning loss (Schneider

et al., 2002). One of the challenges that L2 teachers face is that approaches towards instruction, which result in a rapid improvement in performance, often fail to support long-term retention and transfer, while approaches towards instruction, that appear to pose difficulties for the learner by reducing the apparent rate of learning, often lead to improved long-term retention and transfer (Bjork & Kroll, 2015).

While Cooper (1998) argues that a cognitive overload can lead to the capacity of WM to become fairly restricted, Schneider et al. (2002) suggests that, although conditions of instruction that create difficulties for learners may apparently slow down the learning process, they often result in retention and transfer being optimised. Having taken the same position as Schneider et al. (2002), Bjork (2018) places great emphasis on experiencing optimal levels of difficulty during practice and introduces the term *desirable difficulty.* What he means by *difficulty* is including various unpredictable practices e.g., using tests for learning purposes instead of presentations to pose challenges for learners as it helps to maximise their post-practice retention and transfer in the long-term. These challenges are *desirable difficulty* (DD), as through these demanding practices, effective and long-term learning takes place. Therefore, it might be argued that although creating desirable difficulties for learners has the potential to considerably improve long-term retention and transfer, at times it may affect initial learning, such as slowing down learning or even reducing accuracy rates.

From the literature above, we can understand that when presenting tasks to learners, both WM and desirable difficulty aspects should be considered. This is why the current study focuses on CAT instruction with an objective of balancing between WM and desirable difficulty perspectives. CAT instruction in the present study involved learners in active participation within learning interventions, which required them to understand the meaning of non-semantically

transparent target items. Learners had to remember both the meaning of the L2 target items in the L1 as well as being able to use the L2 target items correctly in tasks. Being able to recall the meaning of the target items and use them correctly in production constituted a desirable difficulty. In order to maximise learning and reduce WM overload, the list of target items was kept to a total of nine or 10 words. Additionally, the intervention for the CAT group provided vocabulary scaffolding through brief explanations of the meaning and lexical system of the target items both in English and in Arabic. Learners were also provided with examples of how the target items were used in context. For the CAT+Corpus group the DD was higher because learners had to work with several examples within a large corpus. Learners were facilitated in the process of selecting the correct meaning and usage of the target items in the corpus during the teaching intervention. Despite this attempt to simplify the task, it was still considerably challenging and more demanding compared to the one given to the CAT only group, which might have led to WM overload.

2.3.4 Semantic Fossilisation

Fossilization is understood as the persistence of errors made by learners, which are difficult to eradicate despite sustained efforts on behalf of the instructors (Richards, 2008). Jiang (2004) argued that the semantic system of the L1 is the main source for L2 lexical development hindrance, which may result in fossilisation of most L2 words. Jiang (2004) also argued that semantic restructuring can help overcome semantic fossilisation. Semantic restructuring is initiated when the learner starts to notice semantic differences between a word in the L2 and its L1 translation. Alternatively, this could take place between two L2 words that are translated the same way in the L1. Semantic differences between an L2 word and its L1 translation or between two L2 words are usually minor; hence, it is possible that natural contexts are not always sufficient to pick up on these subtle changes. Therefore, significant semantic restructuring and development may not result

from simply exposing the learner to the target language in a natural context. Jiang (2002) argued that intentional instructional intervention may be required and useful to help learners recognise such subtle variations in meaning. This is so that L2 learners can gain significantly from a variety of vocabulary methods that help them focus on the differences in semantics between L2 words and its L1 translation as well as differences between L2 words with similar meaning. Jiang (2004) suggested that providing learners with explicit teaching may aid in their understanding of specific terms as well as their ability to notice that translation equivalents do not always convey the same meaning.

Having identified explanations for L1 issues within L2 vocabulary learning in terms of L1 influence, the processing of two languages, the WM capacity of the learner, and the semantic fossilisation of repeated errors made by the learner, it is important to now focus on L2 vocabulary learning from a pedagogical point of view. The following sections cover Incidental Learning, Intentional Learning, Noticing, and the use of corpora.

2.4 Incidental Vocabulary Learning: An Overview

Incidental vocabulary learning is understood as a "by-product" of learning practices that do not specifically target lexical acquisition (Laufer, 2003; Richards & Schmidt, 2002). As such, learning words takes place incidentally when learners have not been told that they would be tested on their learning (Uchihara et al., 2019). It has been suggested that if learners are aware that they will be tested on the vocabulary, they will pay more attention to it and, therefore, take part in intentional rather than incidental learning (Hulstijn, 2003). A further way to define incidental vocabulary learning is to view it as an outcome of meaning-focused activities (Hulstijn, 2003).

One aspect of acquiring incidental vocabulary has looked at the impact of repetition and word frequency on learning; the more a word is used in context, the more probable it is that the learner will understand its form and meaning (Horst et al., 1998; Webb, 2007). This positive outcome is caused by the fact that frequent exposure to the target words makes learning new words easier by introducing them gradually (Schmitt, 2008). Researchers agree that repetition plays a significant role in contributing to incidental vocabulary learning, however studies show considerable differences in the number of times a word is encountered and what effects emerge on learning (Uchihara et al., 2019). Therefore, it is argued that research into incidental vocabulary learning should focus more on understanding the complexity of the relationship between word frequency and incidental vocabulary learning rather than trying to identify a threshold number of encounters for acquisition to take place (Uchihara et al., 2019). For example, this could include the role of L1 and semantic transparency in vocabulary learning.

2.4.1 Incidental Vocabulary Learning and Involvement Load Hypothesis (ILH)

The involvement load hypothesis (ILH) by Laufer and Hulstijin (2001) was proposed to predict the usefulness of instructional tasks for incidental L2 vocabulary learning (Yanagisawa & Webb, 2021). The ILH argues that involvement or engagement with language has motivational and cognitive components, which include *need, search, and evaluation,* as these three factors are believed to influence involvement and therefore vocabulary learning (Laufer & Hulstijn, 2001). Whether or not the ILH provides an accurate prediction of the relative impact of tasks on vocabulary learning has long been the subject of heated debate. This includes, for example, the deliberation about the impact of these tasks on working memory (see section 2.3.3). Indeed, a significant number of studies have been conducted to give an insight into that. Although there are

some researchers who claim that their findings match the predictions of the ILH quite accurately (Eckerth & Tavakoli, 2012; Hulstijn & Laufer, 2001; Kim, 2008), there is an ongoing argument that these predictions are not always consistently accurate (Keating, 2008; Yanagisawa & Webb, 2021). It is important to note that this inconsistency has been mainly attributed to the time allocated to tasks, the frequency of using a target word, and the weightings of the three components – that is, *need*, *search*, and *evaluation* (Keating, 2008; Kim, 2008; Laufer & Hulstijn, 2001).

According to Laufer and Hulstijn (2001), relative task efficacy in vocabulary learning can be reported through the calculation of the involvement load (IL) for each task. In doing so, the points for the three components above need to be added together. First of all, there is the motivational factor which is the *need* component. Given that this component refers to the necessity of understanding unknown words for task completion, three different point levels are considered for need. Zero (0) points are given on occasions when need is *absent*. Simply stated, an unknown word is not needed for the task to be completed. One (1) point, however, is given in situations where it is considered *moderate*. This might be, for example, when learners are asked to either understand or use the word given by an external agent such as a teacher. For instance, the teacher asks his/her learner to use an unknown word in a sentence. Two (2) points are given where need is rated *strong*. This applies to a situation that imposes unknown or new words. For example, a situation in which a learner wants to use an unknown word in his/her speech or written work and looks it up in a bilingual dictionary in order to be able to do so.

Furthermore, the second component *search* is always conceived of as a cognitive factor. This is largely because it refers to the attempt that has to be made to look for either the L2 form or meaning of a vocabulary item. Unlike *need*, only two different point levels are taken into consideration for search: *presence* or *absence*. The first applies to situations in which learners have

to use external resources such as teachers or dictionaries to find an L2 form or meaning. In such cases, (1 point) is typically given. The latter refers to situations in which both the L2 form and meaning are furnished in a task and therefore search is not present; instead, it is absent.

Finally, there is the third component *evaluation*. This occurs when it comes to deciding on the most appropriate option for a context through a comparison of the L2 form or meaning of an unknown word, with other possible words or meanings. Interestingly, this component resembles the first component where giving points is concerned. Like the need component, three different point levels are applied for evaluation. *Zero* (0) points are given in situations where evaluation is *absent*. That is to say, there is no need for a decision to be made on which word or its sense is to be used. *Moderate*, however, calls for (1 point). This is the case when a context is provided. An example can be when learners need to respond to a gap-fill activity by selecting the most appropriate answers from a box which provides them with several options to choose from. Finally, there is the third situation when there is a need for an authentic context where learners spend time on a task and connect to the new vocabulary (Huang et. al., 2012). For example, learners are asked to write the new target items in a composition where they take time to process and connect to their understanding of the lexical system (see Webb, 2005). In such a situation, evaluation is considered *strong* and (2 points) are given.

In Laufer and Girsai's (2008a) CAT study, they argued that translation tasks had higher involvement than the other tasks i.e., the need component was present, the search component was moderate, and the evaluation was moderate in the receptive translation task (from L2 to L1) and strong in the production translation (from L1 to L2). For the current study, these components were included with and without corpus. However, due to the nature of the corpus facilitating a contrastive analysis and translation approach, the corpus-tasks presented in the study were

expected to result in higher involvement load and cognitive processing compared to tasks which did not make use of a corpus. The corpus-tasks required a higher involvement load from learners because they had to understand not only the form and meaning of the target items, but also had to engage with an extensive authentic text, which included many examples of the target items used in context both in the L1 and L2.

It is also important to mention, not all findings are in favour of the ILH. There is an argument that greater vocabulary learning gains are not necessarily achieved through tasks or activities with higher ILs than those with lower ILs (Yang et al., 2017). Yang et al. (2017) for example, explored the effects of task-induced involvement and WM in vocabulary acquisition in post-reading, word-focused activities such as fill-in-the-gaps, sentence-writing, and comprehension. Their participants were advanced L2 university learners from China, completing their first-year major in English. These participants were assigned to three different groups: the fill-in-the-gaps and sentence-writing groups, the comprehension-only group, and the control group. Each group received different tasks after reading a comprehension task. The first group completed word-focused exercises, the second completed one essay question excluding any type of form-focused instruction, and the third group did not complete any of the two learning tasks, and instead, only completed the post-tests and delayed post-tests. By using the Vocabulary Knowledge Scale developed by Paribakht and Wesche (1997) and a Reading Span Test developed by Daneman and Carpenter (1980), the researchers tested two areas—that is, the treatment effects between groups, and the learners' WM capacities. Their findings supported the involvement load hypothesis, but only partially, as the results indicate that in both the immediate post-tests and the delayed post-tests, the sentence writing group equally outperformed the other two groups. In their

analysis, they found that WM played a major role in predicting the gain scores of the comprehension-only and the fill-in-the-gaps groups on the immediate post-test.

Some even went further by pointing out that at times, some tasks with even lower ILs are found to lead to similar outcomes as those with higher ILs (Bao, 2015). For example, in Bao's study (2015), translation tasks were moderate and writing tasks were strong for the evaluation component, however both tasks revealed similar results especially in receptive vocabulary knowledge. Moreover, with reference to the ILH, Uchihara et al. (2022) found that learners who consulted a dictionary to look up the meaning of unknown words had strong intrinsic motivation (*need = two*). This study was among the few which identified *need* as strong. This was linked to learners having to exclusively rely on consulting the dictionary to find the meaning of unknown words. This contrasts with a typical L2 classroom, where learners might rely more on the teacher when searching for unknown words.

Furthermore, there have been some studies that compared the ILH with other frameworks such as technique feature analysis (TFA) (e.g., Hu & Nassaji, 2016). In their study, Hu and Nassaji, (2016) explain that TFA was developed by Nation and Webb (2011) to overcome the involvement load hypothesis' shortcomings by incorporating additional measuring criteria for depth of processing beyond those found in the ILH. Essentially, TFA was modified from an earlier framework for learning vocabulary, which claimed that acquiring vocabulary consists of three components, which include *noticing, retrieval*, and *generating* (Nation, 2001). The revised TFA framework added the components *motivation* and *retention* with the purpose of expanding on the "number of the elaboration parameters" and offering criteria "to assess each component" i.e., the five components (Hu & Nassaji, 2016, p. 30).

Additionally, Hu and Nassaji (2016) compared ILH with the revised TFA. They recruited 96 EFL participants in their study and grouped them into four different groups. Each group was assigned a different vocabulary task using the ILH and TFA frameworks. Results showed TFA had stronger effects in predicting vocabulary knowledge than ILH. This suggests that the effectiveness of IL may vary on the basis of the characteristics of L2 learners, ranging from their L2 proficiency to the similarities between L1 and L2. However, one of the advantages of using ILH rather than TFA is that several previous studies have found evidence in support of the predictive power of ILH (see Hulstijn & Laufer, 2001; Keating, 2008; Kim, 2008; Nassaji & Hu, 2012). However, only a few recent empirical studies have examined the predictive power of TFA (see Hu and Nassaji, 2016; Khoshsima and Eskandari, 2017). While previous studies have explored the use of ILH in CAT instruction (see Alharbi, 2017; Laufer and Garsai, 2008a; Zhang, 2018), more research is needed in this field to better understand its applicability and predictive power.

In the eyes of Laufer and Hulstijn (2001), among all of the three components, it is the *search* component that might have the smallest effect. Furthermore, Kim (2008) suggests that it is the *evaluation* component that may appear to be the most influential factor for learning. Similarly, Tang and Treffers-Daller (2016) and Yanagisawa and Webb (2020) found the influence of the third component - *evaluation* - to be the strongest. They point out that the evaluation component contributed to the largest amount of learning; this was then followed by the first component, which is *need*. The second component, *search*, however, was the one that did not contribute to learning and did not result in any remarkably different learning outcomes. In their meta-analytic study, Huang et al. (2012) argue that the time spent on tasks had a significant impact on vocabulary learning. Their results reveal that spending more time on output tasks leads to more vocabulary gain in comparison to spending less time on tasks.

It should also be emphasised that the ILH has two requirements which are both of paramount importance (Laufer & Hulstijn, 2001). The first is that other factors are equal. This sets the scene for the argument that learning gains and the predictions of the ILH may not be consistent with each other in cases such as time spent on tasks (e.g., Bao, 2015), or frequency in which learners encounter or use target words (e.g., Folse, 2006). The second factor is that the ILH revolves around incidental vocabulary learning rather than deliberate vocabulary learning, where learners learn words intentionally. In other words, contrary to deliberate vocabulary learning where a broad range of strategies may be employed by students for learning to occur, incidental vocabulary learning takes place as a result of engagement with activities, but without purposeful intention to commit target vocabulary items to memory (Laufer & Hulstijn, 2001). Therefore, instead of learners' strategies, learning gains during incidental learning may reflect both the cognitive processes as well as resources involved in performing the task. However, not only does the cognitive ability of learners to acquire vocabulary need to be taken into consideration, but particular attention must also be paid to minimising overloading (Chandler & Sweller, 1991; Cooper, 1998).

2.4.2 Incidental Vocabulary Learning and Vocabulary Enhancement

A great deal of attention has been paid to reading with vocabulary enhancement, which could involve looking up target words in dictionaries (Hill & Laufer, 2003), looking at glosses (Hulstijn, 1992), or even highlighting (Doughty, 1991). Where multimedia glossing is concerned, a study by Rott et al. (2002) argued that although reading and using L1 glosses may initially provide better results than applying incidental learning alone, in the long term, the results might be similar to those of using incidental learning only. Studies have found mixed results in relation

to reading with vocabulary enhancement. The incidental vocabulary learning of advanced learners of French who were repeatedly exposed to new words, was positively impacted by the simultaneous use of vocabulary enhancement that included bilingual dictionaries and marginal glosses (Hulstijn, et al., 1996). Interestingly, Hulstijn et al. (1996) argued that repeated exposure could not be beneficial without such techniques, as in their absence, learners were found to either avoid the words or infer their meanings wrongly. Moreover, their study compared both techniques of using marginal L1 glosses and bilingual dictionaries. The results showed better retention with the group that used marginal glosses. The authors explained that this was because L2 learners rarely opened the dictionaries; however, when they did look up words in dictionaries their chances of recalling the meaning of words were higher than the marginal gloss learners. These findings can be seen in light of two different perspectives related to two theories that were previously discussed. The first is that the authors' explanation is in line with ILH (Laufer & Hulstijin, 2001) and underlines the value to the three components of ILH in tasks that require higher involvement for better learning. Secondly, these findings question the perspective of the cognitive load theory (Cooper, 1998), which posits the importance of avoiding external sources (i.e., dictionaries) to avert cognitive overload. Hulstijn et al (1996) have shown that the use of external sources helped bolster the incidental vocabulary learning of advanced learners of French.

2.4.3 Incidental Vocabulary Learning and CAT Instruction

The concept of contrastive analysis (CA) was established in the 1960s following the work of Lado (1957), who defined the CA of two languages as a process that allows for the prediction of problems, which could be faced by learners of the L2. CA can also help explain errors produced by L2 learners. Lado (1957) argues that CA enables teachers to make comparisons between the L1

and the L2 and helps predict and diagnose difficulties encountered by learners and provide them with appropriate materials. Lado (1957) argued that:

> individuals tend to transfer the forms and meaning, and the distribution of forms and the meanings of their native language and culture to the foreign language and culture both productively when attempting to speak the language and to act in the culture, and respectively when attempting to grasp and understand the language and the cultures as practised by the native. (p. 2)

This quote, in particular, and Lado's research in the field more generally, have made a significant contribution to research and his work has been cited in numerous empirical studies in the area of second language acquisition. Laufer and Girsai (2008a) further explored contrastive analysis and translation activities (CAT) explicitly, and argue that CAT does not solely refer to translations or bilingual glosses to the L2 word, "…but to the kind of instruction which leads to learners' understanding of the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system" (p. 697).

To test CAT instruction, Laufer and Girsai (2008a) conducted their study on words and collocations (e.g. *meet the expectations*, *laudable*) with an aim to analyse the results in light of the noticing hypothesis (Schmidt, 2001), the output hypothesis (Swain, 1995, 2005), and task-induced involvement load (Laufer & Hulstijn, 2001). The noticing hypothesis (Schmidt, 2001) focuses on learners' *attention* intentionally for successful language learning. The output hypothesis claims that "the art of producing language constitutes (speaking or writing), under certain circumstances, part of the process of second language learning" (Swain, 2005, p. 471). The task-induced involvement load hypothesis (Laufer & Hulstijn, 2001) proposes that vocabulary tasks are

effective when they induce higher learner involvement (see section 2.4 on ILH). The aim of Laufer and Girsai's (2008a) study was to find the effect of focused-on-form instructions. This was following their preliminary study (Laufer & Girsai, 2008b) on the effectiveness of contrastive focused-on-form instruction, which revealed that the L2 group that practised reading and translation outperformed (in retention of target items) the other group that only received reading in English.

Laufer and Girsai (2008a) chose not to include a control group, and instead recruited three groups, which included: meaning focused instruction (MFI), non-contrastive form-focused instruction (FFI), and contrastive analysis and translation (CAT) instruction. Each group received different interventions. The first group received two tasks, one a reading exercise with teacher's feedback on answers, and the other a pair/group discussion. The teacher did not focus on the target items and only English was used in the interaction. The second group also received two tasks which included meaning recognition of the target items in multiple-choice format, and a fill-in-the-gaps exercise with target items given in a word-list. They were asked to choose the correct definitions of the given collocations. In the second task, feedback was provided and answers were covered by the teacher. Explanations of the target items were offered in English only. The third group were asked to complete two translation exercises followed by explicit contrastive teaching.

The first exercise required participants to translate L2 to L1, while the second exercise required them to translate from L1 to L2. After the groups completed their tasks, the teacher went through the answers and provided feedback. Following that was a brief explicit instruction and clarifications were provided regarding what is unacceptable in the English lexical system. The three groups were tested on active recall by translating from Hebrew to English. A week after, they were tested on passive recall where it was required to translate English words and phrases to

Hebrew or to give English explanations. One week after, the same tests were given to the students in the same sequence. Test scores depended on writing the target item accurately in the Hebrew prompts, whereas in the Hebrew or English translations, scores depended on correct understanding of the semantic concept. Laufer and Girsai (2008a) found significantly different results in the recall (active and passive) of both form and meaning among the two groups that were receiving (CAT) instruction and non-contrastive form-focused instruction (FFI), as opposed to the group that was receiving meaning focused instruction (MFI). The CAT group outperformed the other groups due to the fact that the latter groups did not receive instruction on similarities and differences.

It is important to note that sustained engagement with vocabulary needs to be maximised. There is a general consensus that the more the learners engage with new vocabulary items, the more likely they are to learn them. Drawing on the involvement load hypothesis (Laufer & Hulstijn, 2001), Hulstijn and Laufer (2001) conducted a motivational and cognitive study involving the three components*: need*, *search*, and *evaluation*. They argue that retention is caused by greater engagement through evaluating. There were three types of engagement tasks: the first was to read a text and answer the given comprehension questions, the second was to read a text, fill in the blanks with the missing words from the provided word-list and answer the comprehension questions, and the third involved a composition task in letter-format and incorporating the target words. As the third task involved more engagement than the other two tasks, it resulted in the highest retention.

In light of the above, it may be argued that any tasks which result in more exposure and attention to lexical items have the potential to enhance learning outcomes. However, it needs to be mentioned that teachers should not focus solely on individual words, but must also focus on phrasal vocabulary and collocations, which are also a key component of the English lexicon. The use of

L1 in second language instruction should also be considered when needed. The current study will use the CAT method in teaching vocabulary, and will expand on the topic of semantically non-transparent words and collocations.

2.5 Intentional Learning: An Overview

Intentional learning, also known as deliberate (Lindstromberg & Eyckmans, 2019) or explicit and direct (Schmitt, 2000, 2008) learning, takes place when the learner is aware of the intention to learn the target word. It is normally done through activities that are completed with the aim of learning a particular vocabulary item, such as rote learning of sight words (e.g., words in flash cards), crossword puzzles, synonyms, and antonyms, amongst others. This is particularly true for beginners and intermediate level students who may use a variety of strategies, such as memorising and word rehearsing (Hulstijn, 2001).

It is important to note that when it comes to incidental and intentional vocabulary learning, there is a very fine line between them. Incidental learning requires learners to infer the meaning of words from context, while intentional learning involves learning vocabulary through a task, such as translation or memorisation (see section 1.2). Because both incidental and intentional vocabulary are complementary to each other, they can be integrated into a single L2 lesson so that learners can produce learned vocabulary items and retain them (Schmitt, 2008). The following section discusses activities in intentional vocabulary learning.

2.5.1 Intentional Learning Activities and FonF

Focus-on-form and focus-on-forms have long been established as the two types of form-focused instruction (FFI). The former is referred to as an approach that "overtly draws students' attention to linguistic elements as they rise incidentally in lessons whose overriding focus is on meaning or communication" (Long, 1991, pp. 45-46). The second is focus-on-forms (FonFs), which is an approach associated with typical and conventional methods that involve teaching separate lessons in a sequential order determined by the syllabus designer or writers (Long, 1988, 1991). Simply stated, in the former, language is used for communication, so students perceive themselves as language users who use the language to learn it. In the latter, however, they perceive themselves as learners of the language who learn the language to use it (Ellis, 2001).

Furthermore, according to proponents of FonF, FonF activities are expected to involve incidental learning and a communicative approach and authentic language tasks (Laufer, 2006), while FonFs' activities are "teaching and practising discrete lexical items in non-communicative, non-authentic language tasks" (Alharbi, 2017, p. 49). An example of FonFs is remembering target words. However, since both incidental and intentional learning are related and complementary (Schmitt, 2008), FonFs can be implemented in both learning approaches (Laufer, 2010). Similarly, FonF can also be used within instructional tasks that focus learners' attention on the target words in the context of both incidental and intentional learning (Laufer, 2005).

Having raised the importance of emphasising instruction in form and meaning, Schmitt (2008) argues that while the exposure method works better for advanced learners, intentional learning is essential for beginners. For example, explicit learning can be in the form of providing beginners with direct translations and form-meaning, which can enhance their learning and avoid overload in their WM capacity (Cooper, 1998).

In order for vocabulary instruction to be enhanced in intentional learning, certain considerations have to be taken into account. Since "maximising engagement is a key principle in vocabulary learning" (Schmitt, 2008, p. 342), vocabulary learning tasks should be designed very carefully. In addition, language learners should be frequently exposed to target vocabulary items. Therefore, it is essential to target words and "the first recyclings are particularly important" (Schmitt, 2008, p. 343). If recyclings are neglected, learned words will not be remembered and all the work put into acquiring these new words will become futile (Nation, 1990).

Last but not least, different aspects of words –that is, meaning-focused input, meaning-focused output, language-focus learning, and fluency development, should be carefully considered. Indeed, these aspects can be taught separately or even in an integrated fashion.

2.6 Noticing

Attention and noticing play a key role in L2 learning. To illustrate this, Schmidt (2010) draws attention to two personal experiences. While the first one is a case study of a Japanese man named Wes who picked up English in a natural environment without receiving formal classroom instruction, the second one was Schmidt's own experience of learning Portuguese. Although Wes was able to do business and deliver his thoughts sufficiently and fluently, he had grammar-related constraints mainly concerning syntax and morphology. Having observed him for three years, Schmidt noticed a "lack of aptitude and over reliance on implicit learning strategy" (Schmidt, 2010, p. 723).

However, in his own case of living in Brazil and attending language classes over the course of three months, he recognized that there were language features around him at all times which he

did not acquire until he consciously noticed them (Schmidt, 2010). Therefore, Schmidt (1990) argues that "the notion of consciousness is useful because it ties together such related concepts as attention, short term memory, control vs automatic processing, and serial vs. parallel processing" (p. 131).

It is important to note that in order for input to become intake, language learners should notice forms and meanings (Robinson, 1995, 2002; Schmidt, 1995, 2001). It has been argued that "input does not become intake for language learning unless it is noticed, that is, consciously registered" (Schmidt, 2012, p. 27). This is the definition of the Noticing Hypothesis which was later revised "consciousness as intention, consciousness as attention, and consciousness as awareness" (Schmidt, 2012, p. 30). However, noticing is a challenging task for second language learners, as it requires taking in meaning and form together (Batstone, 1996).

Therefore, it can be argued that more than one method is required to aid the noticing process. For example, highlighting linguistic features can be considered one method. In the quasi-experimental study conducted by Doughty (1991), the participants in the experimental group were exposed to computer-aided reading lessons with relative clauses highlighted. They appeared to outperform the control group where participants did not benefit from such an exposure. Another method is to promote noticing through interaction and feedback (Laufer & Hulstijn, 2001; Schmidt, 1990, 1994, 2001, 2012). Laufer and Girsai (2008a) encouraged noticing by giving three different groups different treatment conditions. The three groups received meaning focused instruction MFI, non-contrastive form-focused instruction FFI, and contrastive analysis and translation CAT. In the end, the FFI and CAT groups outperformed the other group in terms of recalling words and collocations. This is because CAT and FFI had to be intentionally focused to notice the target words in their tasks, whereas the incidental exposure provided by the task that

was given to the MFI did not encourage noticing, as they only read the material and answered the questions.

It is worth noting that emphasis is placed on the paramount importance of attention in not only learning, but also in knowledge retrieval. Logan et al. (1996) refer to the instance theory of automatization and argue that learning occurs as a result of attending. They even go as far as putting forward the attention theory as a branch derived from the instance theory. They argue that "attention determines what is taken out of memory at retrieval time as well as what gets into memory at encoding" (Logan et al., 1996, p. 620). As can be seen, attention extends beyond encoding, which is the first and most essential step in new memory, and learning to retrieve. Therefore, it may be argued that areas related to a stimulus are retrieved from memory only when attended to.

To conclude, the previous sections have discussed different pedagogical views on teaching L2 vocabulary. The following section will discuss pedagogical tools (i.e., use of bilingual corpora) to stimulate vocabulary knowledge within language learners as touched upon in some of the previous literature reviewed.

2.7 Bilingual Corpora in L2 Vocabulary Learning

The use of corpora in L2 teaching enables learners to construct their L2 knowledge independently by examining the linguistic data within the corpus. Corpora can provide learners with multiple examples of how target linguistic items are used in a sentence (Johns, 1994).

One of the first influential studies on the use of collocations in corpus-based L2 teaching was a study by Chi et al. (1994). Learners were intermediate and advanced learners of English with Mandarin as their L1. The researchers identified appropriate and inappropriate use of collocations i.e., de-lexical verbs. De-lexical verbs are "the tendency of certain commoner transitive verbs to carry particular nouns or adjectives which can in most cases themselves be transitive verbs" (Carter & McCarthy, 1988, p. 153). These include words such as: *get, make, do, have, hold, keep,* etc. The study found that the L2 learners used de-lexical verbs frequently but interchangeably. They also misused the de-lexical verbs in producing collocations. The researchers also stressed the importance of L1 use in the production of collocations, since L1 can influence transfer and phonology when using de-lexical verbs. While the results of the study are interesting, the study was on a small scale and relied only on non-native participants. As such, there was the absence of a detailed comparison between the collocations produced by non-native speakers of English, and those of native speakers, which places a limitation on the findings.

Several studies have reported the positive effects of the use of a corpus in L2 learning (see Alharbi, 2017; Boulton & Cobb, 2017; Chan & Liou, 2005). Corpora used as learning materials often make use of the feature known as key word in context (KWIC). This function allows learners to search for a vocabulary target item by typing it. This initiates a search, which generates multiple examples of how the word is used in context. The concordance lines give the learner various examples of the target items (e.g., frequency effect in Ellis, 2002). Additionally, the KWIC format

makes the concordance lines a salient learning aid, which enhances input (see Chapelle, 2003) and improves the chances that students will notice target items and be able to acquire them (Alharbi, 2017). Alharbi (2017) described the advantages of using KWIC as follows:

> A KWIC format denotes that several sentence examples with the target word are generated. The lines may comprise incomplete sentences and are organised one below the other for the purpose of centralising the intended word or grammatical point in the middle of each line. Through using this technique, the attention of the learners is attracted to the intended word or lexical item and its immediate context in different sentences. (p. 107)

This heavily draws on the noticing hypothesis (Schmidt, 2001), as engaging with a corpus requires greater learner involvement within the three components (*need, search, and evaluation*). According to Laufer and Hulstjin's (2001) involvement load hypothesis (ILH), this should enhance the learning of target lexical items (see Lee et al., 2017, discussed later in this section).

Additionally, Cobb (1997, 1999) focused on adult Omani L2 learners and investigated their vocabulary learning, comparing between the use of a corpus tool (experimental condition), and a traditional method involving the use of dictionaries and word-lists (control group). The study found that learners in the experimental group made greater vocabulary gains compared to the control group. After the intervention, the experimental group had greater definitional knowledge and greater ability to transfer word knowledge to another context. In the same vein, Chan and Liou's (2005) study on adult Taiwanese L2 learners also looked at the use of corpora in L2 vocabulary acquisition. Their study showed that the use of bilingual corpora tools resulted in significant improvement in participants' knowledge of non-congruent collocation.

It is important to note that there are also some drawbacks to using a corpus in L2 vocabulary learning. Boulton (2010) refers to such limitations with the umbrella term *barrier*. Limitations of the approach include the use of technology (e.g., concordance), the use of new material (e.g., when using the KWIC format) and learning approaches (e.g., inductive learning). It has been suggested that learners' inferences about language from contexts (such as concordance lines) may be inaccurate because, at times, they could retain incorrect inferences in their lexicons (Mondria, 2003). It is therefore important that concordance lines provided in corpora use are clear to learners and include sufficient contextual clues to allow learners to process target vocabulary items (Lee et al., 2019).

In a recent study by Lee et al. (2017) on vocabulary learning in digital reading environments, the authors explored the effects of two different vocabulary learning conditions and the clicking behaviour relating to them. The first condition presented learners with reading the concordance lines of target lexical items. Here learners were expected to find the meaning of the target item by reading the sentences. The second condition offered learners additional support by providing the definition of the target lexical items after they read the concordance lines. This allowed learners to check whether their meaning inference was correct. Initially, the EFL learners completed a meaning-recall vocabulary pre-test; after that, there was a two-week period in which different reading tasks were completed by the learners followed by an immediate post-test on meaning-recall vocabulary. These tests required learners to provide definitions of the target items by using L1 translations or writing the meaning in English. The findings of the study indicate that the scores of both conditions reveal better meaning-recall than the control group. According to Lee et al. (2017), this supports the Noticing Hypothesis (Schmidt, 2001) and ILH (Laufer & Hulstijn, 2001).

Moreover, the findings also reveal that meaning-recall in the second condition was more effective and resulted in better vocabulary gains compared to the first condition. The authors explain this finding with reference to previous empirical studies (e.g., Cobb et al, 2001) which suggest that if learners' inferences are followed by confirmation of meaning, this will result in more accuracy in meaning inferences (i.e., from concordance lines). Lee et al.'s (2017) findings also confirm the outcomes of Laufer's (1993) empirical study where having both the meaning and example sentences maximises vocabulary and comprehension knowledge. Interestingly, Lee et al. (2017) suggested that the optimal number of concordance lines for their meaning inference is three lines, and that there are complex interactions between: the target item and its context, the learner's clicking behaviours, the selected concordance lines, and the learner's proficiency background.

The above-mentioned studies have focused on the potential value of corpora in L2 instructed vocabulary learning, teaching definitions and transferring them to other contexts (e.g., Cobb, 1997, 1999), learning collocation (e.g., Chan and Liou's, 2005), and inferencing (Lee et al., 2017). It has been previously suggested that corpora have a role in developing linguistics and language education (e.g., Thorndike and Lorge's *Teacher's Word Book of 30,000 Words* (1944) or West's *General Service List* (1953)). In 1987, John Sinclair established a project called COBUILD which included an electronic corpus and lexicography. Sinclair observed the convergence between language teaching and corpora and stated the following: "it is my belief that a new understanding of the nature and structure of language will shortly be available as a result of the examination by computer of large collections of texts" (Sinclair, 1991, p. 489).

Furthermore, corpus linguistics has been commonly defined as both "… a means to explore actual patterns of language use and as a tool for developing materials for classroom language instruction" (Reppen & Simpson-Vlach, 2020, p. 91). This is mainly because a corpus has many

useful features. For example, it can include authentic texts that help language learners understand how language is used in real life, as opposed to the "unreal and sometimes stupid" examples in textbooks (Chambers, 2005, p. 120).

Another feature that could be included in corpora is having parallel translated lines in two or more languages, which is referred to as parallel corpora (Alotaibi, 2017). Alotaibi (2017) has developed an ongoing project called the Arabic English Parallel Corpus (AEPC) and has used it for her students to help with translation training and language teaching. The aim of AEPC is to develop a 10-million-word Arabic–English parallel corpus as the first phase. This is an advantage for learners as most Arabic–English corpora are small in size (around 1–3 million words only) (Alotaibi, 2017). Another important feature, which is mentioned repeatedly by Alotaibi (2017), is that AEPC has a user-friendly web interface, which is free of charge and "supports bilingual search queries and several filtering options" (p. 332) making it relatively less complicated to use compared to other parallel corpora (e.g. OPUS in Tiedemann (2012)) and with no cost barriers (e.g. Linguistic Data Consortiums, LDC (2023)). In addition, APEC includes highlighted features for the target words, while it also has audio options and aligns the texts for the two languages in a parallel display (Alotaibi, 2017). Indeed, by having aligned translated texts, this can help learners and translators to pinpoint the similarities and differences between expressions for both languages (Hunston, 2002).

Another feature of corpora is profusion. Due to the nature of computers, a wide range of language information can be stored in corpora, making them an important resource to language learners (Alharbi, 2017). In other words, a corpus offers "the opportunity to condense and intensify the process of learning through exposure to multiple examples of the same vocabulary item in context, and to promote awareness of collocational relationships" (Thurstun & Candlin, 1998, p.

270). Some could argue that dealing with such a wide collection of language data can be overwhelming and time-consuming for teachers. In reality, however, it can be made more manageable by language teachers, for example, using small genre-specific corpora, sub-corpora derived from large corpora, or language for specific purposes corpora (Braun, 2005). Furthermore, language teachers should take into account the goals of the lesson and select appropriate corpus-material.

Many studies have explored the positive impact of using parallel corpora (Reynold, 2015; Sangawa, 2014; Wong & Lee, 2016). Reynolds (2015), for example, conducted empirical research on 25 Taiwanese medical students to explore the usefulness of a web-based English/Chinese bilingual parallel corpus collocational concordance and how it can improve self-edits in writing and the production of collocation. Through a mixed-methods approach, the results indicated that using a bilingual parallel corpus improved learners in verb-noun collocation across three drafts of two different essays, i.e., descriptive writing and opinionated writing. These results were measured in the following ways: in the first draft, students received instruction to emphasise global mistakes, while in the second draft, students were instructed to look at local errors and were encouraged to review their work using a bilingual parallel corpus. There was a decrease in the number of collocations used from draft one to two. However, in draft three there were more insertions of collocations in both essays. These findings reveal that there was a significant improvement in accuracy between draft one and three, and between draft two and three. There was no significant difference between draft one and two, which suggests that the bilingual parallel corpus had contributed to this improvement. Concerning the qualitative analysis, the learners' feedback on the experience of using a bilingual corpus indicated different levels of acceptance and success. It is

also important to note that Reynold's (2015) study did not include a control group, and therefore, the results should be viewed with caution.

2.8 Rationale for Using a Corpus with CAT Instruction

The previous literature discussed the merits of using corpora. The researcher will conclude with a rationale for using a parallel corpus plus CAT instruction in the present study. By definition, CAT instruction focuses on L1 and L2 similarities and differences in the lexical system as a whole including vocabulary (Laufer & Girsai, 2008a). The parallel corpus can work as an operational tool to maximise CAT instruction. Through the parallel corpus, the learner can participate in explicit learning to raise awareness and undertake tasks to contrast the L1 and L2. In other words, the parallel corpus is a tool that can complement CAT instruction by providing a contrastive analysis and translation method, with the addition of positioning both L1 and L2 within an authentic and wide range of texts and phrases. It also encourages learners to notice these differences through highlighting the target items. Moreover, since the use of the AEPC corpus aligns with the use and purpose of CAT instruction, it is of interest to use AEPC as a learning tool in language teaching and explore its effectiveness and outcomes.

Another point to take into account is that in the real-world, the growth of computer-assisted learning has rapidly emerged and continues to grow in all fields and majors; and language classrooms should not be an exception. In fact, there is today a tradition of research and practice towards supporting the use of corpora in language teaching and learning (Boulton & Pérez-Paredes, 2014). Furthermore, Mahboob and Elyas (2014) expressed the view that traditional and often outdated teaching methods are the main factors that have led to poor learning outcomes

amongst students and a loss of interest in language classes. Therefore, it may be beneficial to use computer aids in contemporary language learning classrooms.

In order to enhance language learners' performance and motivation, and to encourage the integration of contemporary learning tools, the present study will explore the utilisation of a bilingual corpus within CAT instruction. The researcher understands that the bilingual corpus is not without its limitations as it is characterised by flaws that must be taken into account (see chapter 3.6.2 on limitations). These limitations include: mismatched results from searching for some target items from Arabic to English and vice versa. Also, because the target items are semantically non-transparent, some of the searches and translations can be misleading and confusing, particularly where searching from Arabic to English is concerned. Additionally, although some of the target items do have a translation from L2 to L1, when it is done the other way around, no results are generated. Finally, AEPC does not fully support English collocations which means that in order to show the results and translations of some collocations, articles or propositions that are part of the collocation need to be excluded from the search.

2.9 Chapter Summary

The following section will summarise the aims of this study. The review of literature identifies several research gaps within the area of L2 vocabulary learning. The present study seeks to explore and address these gaps.

1. Although there are a number of studies investigating both incidental and intentional learning, there is limited research on the development of vocabulary knowledge within incidental learning associated with explicit focus-on-form methods, particularly involving

CAT instruction. Thus, the present study will explore CAT instruction in two different conditions (in a bilingual corpus and traditional methods) and will explore the effectiveness of each condition.

2. The previous research discussed L2 vocabulary learning where the target items were words or collocations, yet semantically non-transparent linking adverbials both as words and collocations in a CAT setting have not been investigated. Given the difficulty that such adverbials have been shown to have for learners (Larson-Walker, 2017), it is important to explore whether CAT instruction can be beneficial.

3. Based on the literature presented, there are few empirical studies on CAT instruction, but even fewer studies exploring CAT instruction combined with corpora. Thus, the present study explores both these areas.

Having established in chapter One and chapter Two the rationale for using CAT instruction in a Saudi context, this research draws on the following perspectives: the noticing hypothesis (Schmidt, 1994) and the involvement load hypothesis (ILH) (Laufer & Hulstijin, 2001), as well as the influence of L1 transfer to L2 vocabulary acquisition. The researcher aims to address the existing gaps in knowledge by investigating the learning of L2 semantically non-transparent words and collocations through two different approaches: i) CAT instruction led by the teacher, and ii) CAT instruction using a parallel corpus. In this context, this study is driven by the following research questions (RQ) and hypotheses:

2.10 Research Questions:

RQ1: *What is the impact on short-term learning of semantically non-transparent words and collocations (form and meaning recall) among high-intermediate learners of English of:*

(a) a contrastive analysis and translation (CAT) instruction by the teacher?

(b) a combined approach using both a parallel corpus and direct teacher CAT instruction?

(c) how do the previous two approaches compare with an approach using neither CAT nor a parallel corpus?

RQ2: *What strategies do learners report using to understand semantically non-transparent words and collocations in a think-aloud task?*

RQ3: *What are the views of the learners on the teaching they have received? Do these views illuminate the learners' strategy use and how well they did in the tests?*

2.11 Research Hypotheses

Based on the literature and theories presented in this chapter, the researcher proposes five hypotheses associated with CAT instruction.

H1. The CAT+ Corpus and the CAT conditions will lead to receptive recall of a significantly larger number of semantically non-transparent words and collocations than the control group.

H2. The CAT+ Corpus condition will lead to receptive recall of a significantly larger number of semantically non-transparent words and collocations than the CAT condition and the control group.

H3. The CAT+ Corpus and the CAT conditions will lead to productive recall of a significantly larger number of semantically non-transparent words and collocations than the control group.

H4. The CAT+ Corpus condition will lead to the productive recall of a significantly larger number of semantically non-transparent words and collocations compared to the CAT conditions and the control group.

H5. Receptive knowledge will be expected to emerge earlier compared to productive knowledge.

The first hypothesis predicts that both treatment groups (CAT+ Corpus and CAT) will outperform the control group in receptive vocabulary knowledge because of the CAT method. The CAT method depends on a structured approach in using both L1 and L2 (Laufer & Girsai, 2008a). This hypothesis is based on previous empirical studies involving the positive effects of CAT teaching method (e.g., Zhang, 2018). The second hypothesis predicts the CAT+ Corpus condition will outperform the other two groups in receptive vocabulary learning because of the exposure to a) authentic language within the corpus (e.g. in Alharbi, 2017; Boulton & Cobb, 2017) and that b) previous research have found that intermediate and advanced students gain vocabulary from using corpus, c) according to the literature involving ILH (see Yanagisawa & Webb, 2021), the more evaluation there is, the more learning will occur. When learners are working with a corpus, evaluation can be considered higher because it requires them to identify the meaning of words used in an authentic context. For this kind of activity, learners need to spend more time on a task and connect to the new vocabulary (evaluation is further discussed in section 2.4.1). The third and the fourth hypotheses have the same justifications as the first and second hypothesis respectively but for productive knowledge. Finally, the fifth hypothesis looks into how the interventions work differently for both receptive and productive knowledge. Receptive knowledge usually precedes and exceeds productive knowledge which is explained in Nation (2001) (see chapter 2.1).

Chapter Three

Methodology

3.1 Research Questions

RQ1: *What is the impact on short-term learning of semantically non-transparent words and collocations (form and meaning recall) among high-intermediate learners of English of*

(a) a contrastive analysis and translation (CAT) instruction by the teacher?

(b) a combined approach using both a parallel corpus and direct teacher CAT instruction?

(c) how do the previous two approaches compare with an approach using neither CAT nor a parallel corpus?

RQ2: *What strategies do learners report using in each condition to understand semantically non-transparent words and collocations in a think-aloud task?*

RQ3: *What are the views of the learners on the teaching they received? Do these views illuminate the learners' strategy use and how well they did in the tests?*

3.2 Paradigm Rationale

When it comes to social sciences, the main paradigms are: post-positivism, critical theory, and interpretivism or constructivism (Guba,1990). Positivism argues that reality exists as a measurable entity that can be assessed through quantitative data. One could argue that vocabulary knowledge is something that can be measured, and thus a quantitative approach is appropriate in order to assess how effective an intervention targeting vocabulary is. Interpretivism, in contrast, claims that *reality* is subjective and unmeasurable. It is impossible to measure someone else's perceptions, ideas, or thoughts as an objective, external reality. Qualitative methods are more

appropriate as a result. Arguably, therefore, to gain a full picture of the impact of an intervention, using both qualitative and quantitative methods might be preferable, which is the position taken by the pragmatism approach. Creswell (2012) explains that the "quantitative data and results provide a general picture of the research problem; more analysis, specifically through qualitative data collection, is needed to refine, extend, or explain the general picture" (p.542).

The study adopts a pragmatic approach towards data collection (Teddlie & Tashakkori, 2009), employing a mixed-methods strategy comprising both quantitative and qualitative methods. Johnson and Onwuegbuzie (2004) argued that one should "choose the combination or mixture of methods and procedures that works best for answering your research questions" (p. 17). Similarly, Tashakkori and Teddlie (1998) also emphasise the importance of adopting a research method that can best answer the research questions. The rationale for adopting a mixed-methods approach is therefore informed by the research questions guiding the investigation. While, on the one hand, the study seeks to objectively measure whether a particular intervention has an effect on students' vocabulary learning, it is also concerned with understanding students' thought processes and views on the intervention methods, which can only be explored through a qualitative data collection method. As a philosophical movement, pragmatism came into existence in the United States of America in the late 19th century, and it was initially developed by Charles Sanders Peirce, William James, and John Dewey (Hammersley, 1989). It is a combination of two core elements which are referred to as the pragmatic method and theory of truth. It is important to note that instead of abstract first causes (surface levels), a commitment to end causes and practice outcomes (deep levels and results) is considered key to the pragmatic method. This is mainly because it "tries to interpret each notion by tracing its respective practical consequences" (James, 1995, p. 18).

Furthermore, it is important to note that pragmatism remains committed to real-life problem-solving by relying on useful knowledge that has been validated through the theory of truth. This is because problems in the real world can be referred to as the "dynamic interrelation between social meaning-generating rules and ensembles of individual disposition-factors" (Franke & Weber, 2012, p. 8). This contrasts with ontological realism which works on the assumption of a world that is independent of the observer (Friedrichs & Kratochwil, 2009). According to Friedrichs and Kratochwil (2009), pragmatism is intersubjectivity – that is, shared understanding that serves as a characteristic of the social world. As such, it can stand for epistemological instrumentalism and a consensus theory of knowledge evaluated by scholarly and external communities. However, they also argue that pragmatism is a "reflexive practice of discursive communities of scholars" (Friedrichs & Kratochwil, 2009, p. 711). This implies that theories of truth can be renegotiated according to new conditions and situations. In fact, Franke and Weber (2012) add that "our current expedient theories, philosophies, and truths…might one day become thought of as of little use, that is, as false" (p. 16).

The current study addressed the two research questions using a mixed-methods explanatory sequential design that began with a quantitative method followed by qualitative data collection. The reasons for using a mixed-methods approach were, firstly, to gather quantitative data in order to measure L2 students' improvements in vocabulary knowledge following a CAT intervention. Secondly, to use qualitative data to gain a deeper insight into the learning of participants receiving the CAT intervention. The following sections will explore both methods.

3.3 The Explanatory Design (A Quasi-Experimental Research Study)

One of the hallmarks of experimental research is randomization (Mackey & Gass, 2005). In such studies, the researcher randomly assigns participants to different treatment conditions for the experimental variable. Simply stated, the randomly assigned participants in the experimental group receive treatments. It should be mentioned that they can either receive just a post-test or both a pre-test and a post-test (see Table 3.1). Where the latter is used, data are collected both prior to and after the intervention. This enables the researcher to conduct statistical tests, which are able to explore trends and correlations more robustly than studies with just a post-test, and to determine whether any differences between the groups can be attributed to the intervention (Creswell, 2012).

Table 3.1

Difference Between Experimental and Quasi-Experimental Research (Creswell, 2012, p.310)

Types of Between-Group Designs

True Experimental Designs

Pre- and Posttest Design **Time** →

Random assignment	Control Group	Pretest	No Treatment	Posttest
Random assignment	Experimental Group	Pretest	Experimental Treatment	Posttest

Posttest-Only Design **Time** →

Random assignment	Control Group	No Treatment	Posttest
Random assignment	Experimental Group	Experimental Treatment	Posttest

Quasi-Experimental Designs

Pre- and Posttest Design **Time** →

Select Control Group	Pretest	No Treatment	Posttest
Select Experimental Group	Pretest	Experimental Treatment	Posttest

Posttest-Only Design **Time** →

Select Control Group	No Treatment	Posttest
Select Experimental Group	Experimental Treatment	Posttest

For the purpose of this study, a quasi-experimental research design was used. Unlike experiments, a quasi-experiment does not rely on random assignment, but rather, participants are assigned to groups based on non-random conditions. Creswell (2012) summarises not only the process, but also the differences between experimental and quasi-experimental research (see Table 3.1). Experimental research includes "(1) a pre- post-test design, (2) a treatment group and a control group, and (3) random assignment of study participants" while quasi-experimental studies "lack one or more of these design elements" (Silver-Pacuilla et al., 2011, p. 17). Furthermore,

many researchers in the education sector use intact groups for their experiments, as they can be less disruptive to the educational setting (Porte, 2002). It should be emphasised that although an intact group is not a feature of a typical experimental design "it may have the advantage of enhancing the face validity of certain types of classroom research" (Mackey & Gass, 2005, p. 143).

3.4 The Explanatory Design (Think-Aloud Interviews and Follow-Up Interviews)

Although a number of studies have revealed some of the positive effects of CAT instruction on incidental learning (Alharbi, 2017; Laufer & Girsai, 2008a), how L2 learners actually use contrastive information for recall of items while completing tasks, requires further qualitative investigation. Quantitative data are only able to show that different learning outcomes have occurred, yet without explaining how or why. In order to gain such insights, this study used think-aloud (TA) interviews to gain a more complete picture of students' learning of L2 semantically non-transparent words and collocations in CAT settings, compared to what can be gained from test data alone. Think-aloud interviews, also referred to as concurrent reports, are "those collected as subjects verbalise while performing the task in question..." whereas other types of interviews that are retrospective reports "… are collected when subjects verbalise some time after performing the task" (Bowles, 2010, p. 13).

Importantly, it is argued that thinking aloud can affect the validity of the research because it is unknown whether the act of speaking concurrently while completing a task is actually providing a true reflection of thoughts and ideas, or whether it just involves responding to an extra task and altering cognitive processes (Bowles, 2010; Ellis, 2001). Therefore, Bowles (2010) provides a series of recommendations to be implemented before and while using the think-aloud

method to ensure validity (see Figure 3.1). The implementation of these recommendations within the current study is further discussed in sections 3.11.2**.**

Figure 3.1

Recommendations for Think-Aloud Interviews (Bowles, 2010, p. 21)

Before the study

- Decide if think-alouds are appropriate to use with the type of language task you plan to use.
- Decide on the type and language of verbalization you want to elicit.
- Include an indication that voice samples will be recorded in the informed consent document.

During the study

- In the think-aloud instructions provided to participants:
  - Include a rationale for having participants think-aloud.
  - Provide general instructions to participants about how to think-aloud
- Provide learners with a warm-up task to let them practise thinking aloud before they move on to the experimental task.
- Record think-alouds using appropriate equipment to ensure the necessary level of detail.
- To ensure validity:
  - Verify that participants in think-aloud groups continue thinking aloud throughout the task. Remind them of this as necessary
  - Include a small control group that performs the same tasks without thinking aloud as a check on validity.

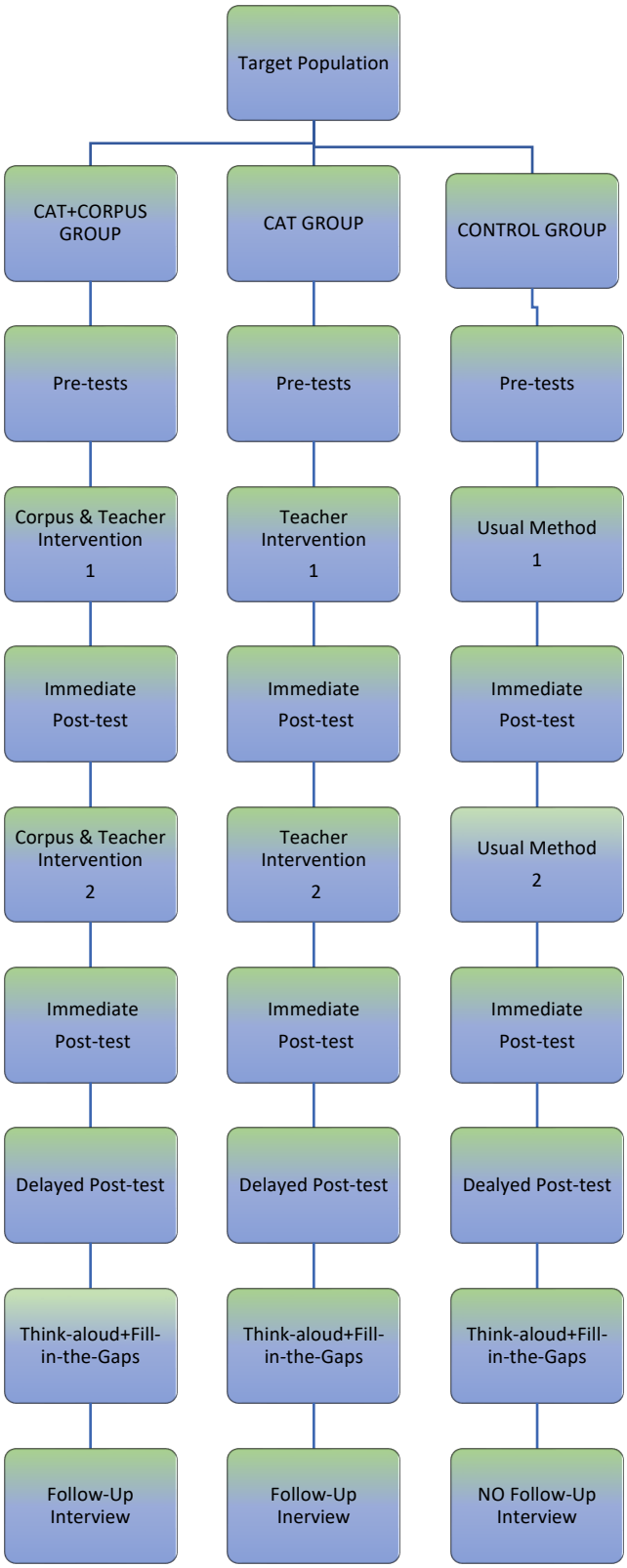
This study explored the effectiveness of CAT and CAT+Corpus teaching interventions on participants' learning of a set of 19 target items. Participants were divided into three different groups: a CAT group, a CAT+Corpus group and a control group. The main study was carried out over the course of one term, which lasted for six weeks; four weeks for the experimental study and one week for the qualitative interviews. Participants took a pre-test, which was then followed by two teaching interventions (see section 3.12). After the interventions, participants completed an immediate post-test allowing the researcher to assess pre- and post-test scores. This was followed by a delayed post-test (one week after the second post-test) and a subsequent fill-in-the-gaps task with TA interviews. Furthermore, in order to gain a better understanding of participants' views on the interventions they received, the current study included follow-up interviews, which took place after the TA interviews. Participants in the CAT and CAT+Corpus groups were asked about their perspectives on their experiences of the interventions. A pilot study was conducted, which provided a valuable opportunity to test the chosen methods. Importantly, the pilot brought to the fore many critical unforeseen elements that helped to shape and refine the final design process and data collection procedure of the main study. This is discussed in more detail later in this chapter.

For the quasi-experimental study, the original plan was to have three groups in each of two consecutive terms taught on the programme. This would have involved the same procedure and method for group allocations, but with a different set of participants, in order to obtain a large enough sample size. In other words, the intervention would be repeated with different participants across the two terms. However, there were particular challenges in recruiting participants because of the global pandemic, which resulted in a lockdown being imposed in March and April of 2020. This had an immense impact on the data collection phase for the second term. Therefore, the study had to be conducted during one term only, with three groups that included two treatment groups,

namely a CAT with bilingual corpus group (CAT+Corpus); a CAT with no corpus group; and a control group that received no treatment. Figure 3.2 outlines the research design adopted in the main study. It is important to note, however, that that even the data collection process for the second post-test and the delayed post-test in term one was affected by the pandemic, since the term ended in February 2020 (see section 3.11).

The study also originally planned to have a larger sample size, as suggested in Field (2018), who, in his statistical manual book, recommended a sample size of 30 participants per group. Laufer and Girsai (2008a) also had a slightly larger sample size compared to the present study, with groups of 23 and 26 participants. Zhang's (2018) study similarly utilised groups which ranged between 37 and 40 participants. In order to assess the adequacy of this sample size, a post-hoc power analysis was conducted, as reported in Chapter 4. However, because of practical reasons related to the Covid-19 pandemic and absence of the participants recruited, each group only included 18 or 19 participants.

Figure 3.2. *The Experimental Process and Method in Conducting the Study Design*.



The assessment methods and treatments adopted in this study bear a strong resemblance to those used by Laufer and Girsai (2008a). In this study, the assessment method for vocabulary, examined knowledge of meaning and form through the administering of pre-, post-, and delayed tests. The treatments included reading sessions, comprehension questions, and CAT intervention, with and without the use of a corpus. It should also be mentioned that the terms receptive knowledge (translating target items from L2 to L1) and productive knowledge (supplying the translation of target items from L1 to L2) are also used in the present study.

3.5 Participants

Initially, the original plan was to recruit 150 female native speakers of Arabic, aged between 18 and 21, in their first year at an English Language Institute (ELI) located in Jeddah in Saudi Arabia. The research was to take place across two consecutive terms, with 75 participants being recruited in each term. After gaining their full consent, the 75 participants would be equally split across each of the three research groups, with each group comprising 25 students per term, or 50 across both terms. However, as discussed above, in wake of the pandemic, adjustments were made to the number of participants, which resulted in 56 female participants being recruited during one term at ELI. There were 19 participants for the CAT group, 18 participants for the CAT+Corpus group, and 19 participants in the control group. These three groups included students who were at the high-mid proficiency level and were randomly selected.

A background questionnaire was administered to ascertain that participants had not lived in an English-speaking country. However, the majority, if not all, had studied some English during their primary and secondary school years at public schools, and very few in private schools.

3.5.1 Participants' Background in Learning English in Saudi Arabia

All participants started learning English from the age of nine. English is now a mandatory subject in the standard curriculum of public primary schools in Saudi Arabia. Students are expected to attend 90 minutes of English classes on a weekly basis. Starting from age 12, English classes increase to three hours per week. By the time Saudi students enter university age 18, they have received approximately nine years of English instruction. Upon gaining entry to KAU university, they take a standardised placement test through Cambridge Standards Exams to assess their English language proficiency. The result of the entry test helps to determine the level (beginner, low-intermediate, high-intermediate, and advanced) at which they need to be placed. At university, each level is designed to be completed in one term, which spreads over a course of approximately six weeks. At the end of each term, students take another Cambridge Standards Exam to assess their level. For example, low-intermediate students are assessed on the knowledge they gained from their current level. Upon successful completion of each level and exam, students move up through the system to the next level. Thus, the participants for this study were all equally qualified for their level entry. All participants had passed the Cambridge Examination to enter level 3 (high-intermediate), however students' individual scores were not shared with instructors, therefore this information was not used in the data analysis. There were, however, observable differences in the participants' overall competence in English, including vocabulary knowledge, as is the case with every group. This was determined through teacher observation of students' classroom contributions during the teaching sessions. For the purpose of this study, participants were chosen from the high-intermediate level, as by this stage, students are required to broaden their knowledge of collocations (e.g., collocations in linking adverbials). However, most of the participants had little exposure to reading outside the classroom. Out of the 56 participants in the current study,

only 26.7% from the three groups stated in their information sheet that they read English books, while 53% revealed that they read English online.

English instruction in the university in which the study is set combines explicit and implicit approaches. At each level, two textbooks are used to address the four language skills – that is, speaking, listening, reading, and writing. The teachers mainly use English as the medium of instruction in their classes; however, when encountering unknown words, some teachers may use other instructional methods, such as translation. This mainly depends on both the complexity of the word and the teacher's language background; there are some ELI teachers who cannot translate, as they do not speak the Arabic language.

3.6 Teaching Materials and Measures

3.6.1 Reading Materials

This section describes the process of selecting and modifying the reading materials, which were used in the teaching interventions. In the ELI courses, in order to cover all language skills, students are taught from two textbooks, one focusing on reading and writing and the other on speaking and listening. The textbooks are called *Unlock* (Westbrook et al., 2019) published by Cambridge and are designed specifically for English learners at university level. The reading materials that were used for the current study were taken from the students' reading and writing textbook. The book is level 3, which is equivalent to B1 in the Common European Framework of Reference. The reason for using the reading materials from these textbooks is because, overall, it is quite challenging to find reading texts with enough multi-unit words in them. Thus, in order to fulfil the requirements of this study, the reading tasks from the students' reading and writing textbook had to be modified (see Figures 3.3 & 3.4). The modification was done on two reading

passages only. The modification required inserting specific target items into the original extracts taken from the textbook.

After having two teachers, both native speakers of English, evaluate the modified reading tasks prior to them being used, 19 words and collocations were inserted into two passages; nine target items in the first passage and the remaining ten target items in the second passage. The passages were then printed out so that students did not need to use their textbooks. Figures 3.3 and 3.4 present the first reading passage used for the first reading session, and the second reading passage that was used in the second reading session, respectively. As can be seen, the text has some transitional words, but limited transitional collocations. The original transitional items in the book were replaced with semantically non-transparent linking adverbials that have the same meaning and serve the same purpose. For instance, *however* was replaced with *on the other hand*. Therefore, the target items did not interrupt or cause a breakdown in the meaning or flow of the passage.

Figure 3.3.

Passage 1 for reading session 1 (Westbrook et al., 2019)

Our changing planet

The Upsala glacier in Argentina used to be one of the biggest glaciers in South America. In 1928, it was covered in ice and snow but now the glacier is melting at an annual rate of 200 metres, so the area is covered in water. This is evidence of global warming.

In the last 100 years, the global temperature has gone up by around 0.75 degrees Celsius. This may not sound much but such a small increase is causing sea levels to rise and threatening the habitat of many species of plants and animals. An increase of two degrees Celsius in global temperatures could result in extinction for 30% of the world's land species.

The Northwest Passage is a sea route which runs along the northern coast of Canada between the Atlantic and the Pacific Oceans. In the past, it was often difficult to use, as the waters were frozen; however, increasing temperatures and the subsequent deglaciation have made it easier to travel down this route. The major issue is that this will lead to loss of habitat for the polar bears and other species that live in this area.

Sea levels in the UK have increased by around 10 cm in the last 100 years and experts predict that global sea levels could rise by up to 59 cm by the end of the century. Consequently, areas which were land a few hundred years ago are now submerged and many low-lying islands may be under water in the future.

As a result of the changing climate, the world's ecosystems are also changing faster than ever before. Over one-third of the world's mangrove forests and around 20% of the world's coral reefs have been destroyed in the last few decades. Forests are being cut down to provide land for food because the population is growing at such a rapid rate. Approximately a quarter of land on earth is now used agriculturally for growing food. As a result of the higher temperatures and higher levels of carbon dioxide in the atmosphere, plants are producing more pollen which could lead to more cases of asthma.

So what is causing climate change? The main cause of climate change is the huge amount of greenhouse gases such as methane and carbon dioxide (CO_2) in the atmosphere, but the reason for this is the world's population – you and me. As the population increases, more land is needed to provide food, and more energy is needed. Burning fossil fuels for heating, lighting, transport, electricity or manufacturing produces CO_2. Furthermore, humans breathe out CO_2 and trees 'breathe in' CO_2 and produce oxygen – so by cutting down trees, we are increasing the amount of CO_2 in the atmosphere and reducing the amount of oxygen. As a result of these activities, CO_2 levels are now at their highest for 800,000 years.

The biggest challenge we all face is to prevent further environmental disasters. We must do something before it is too late. We need to reduce the amount of CO_2 in the atmosphere. We need to stop burning fossil fuels and start using renewable energy. We can get enough energy from renewable fuels like solar energy, hydroelectric energy or wind power to be able to stop using fossil fuels completely.

Sign the petition to get governments to take action before it is too late!

Figure 3.4.

Passage 2 for Reading Session 2 (Westbrook et al., 2019)

Deforestation means the removal of a forest so that the land is converted for urban use or agriculture. The destruction of forests occurs for many reasons: trees are used as fuel or for construction, while cleared land is used as pasture for livestock and crop plantations. The main adverse effects of deforestation are aridity and damage to animal habitat, as well as climate change and erosion damage.

The main causes of deforestation are commercial agriculture by big business and subsistence farming by local people. In Indonesia, industrial logging is carried out to clear areas for the production of palm oil while in Brazil, large areas of the Amazon rainforest were cleared to grow soya and vegetable oil. Subsistence farmers clear an area big enough to graze cattle or grow crops by cutting down the trees and burning them. However, after two or three years, the land can no longer be used so the farmer moves to another piece of land. It takes around ten years for the piece of land to recover. However, in populated areas, the land cannot recover and this leads to heavy erosion because the layer of soil that protects the ground is removed during the crop-growing process. This can cause flooding problems in heavy rain.

One area affected by deforestation is the Amazon Basin in Brazil. The vast rainforests of the Amazon cover an area around 25 times the size of the UK and absorb an estimated 1.5 billion tonnes of carbon dioxide annually. They are thought to have helped keep global warming under control in recent years. However, in areas where deforestation has taken place, the increased temperatures result in less rain because there are not enough trees to provide water for clouds to form. If tropical forests dry out, more trees will die and there will be more logging and more fires. This will cause more emissions of carbon dioxide, making the rainforest contribute to global warming rather than help solve it.

Forest destruction is also having an effect on biodiversity. The growth in the world's population is causing the loss of habitats and damage to land where plants and animal species live, reducing biodiversity and leading to the extinction of many species. A decrease in biodiversity threatens entire ecosystems and destroys future sources of food and medicine.

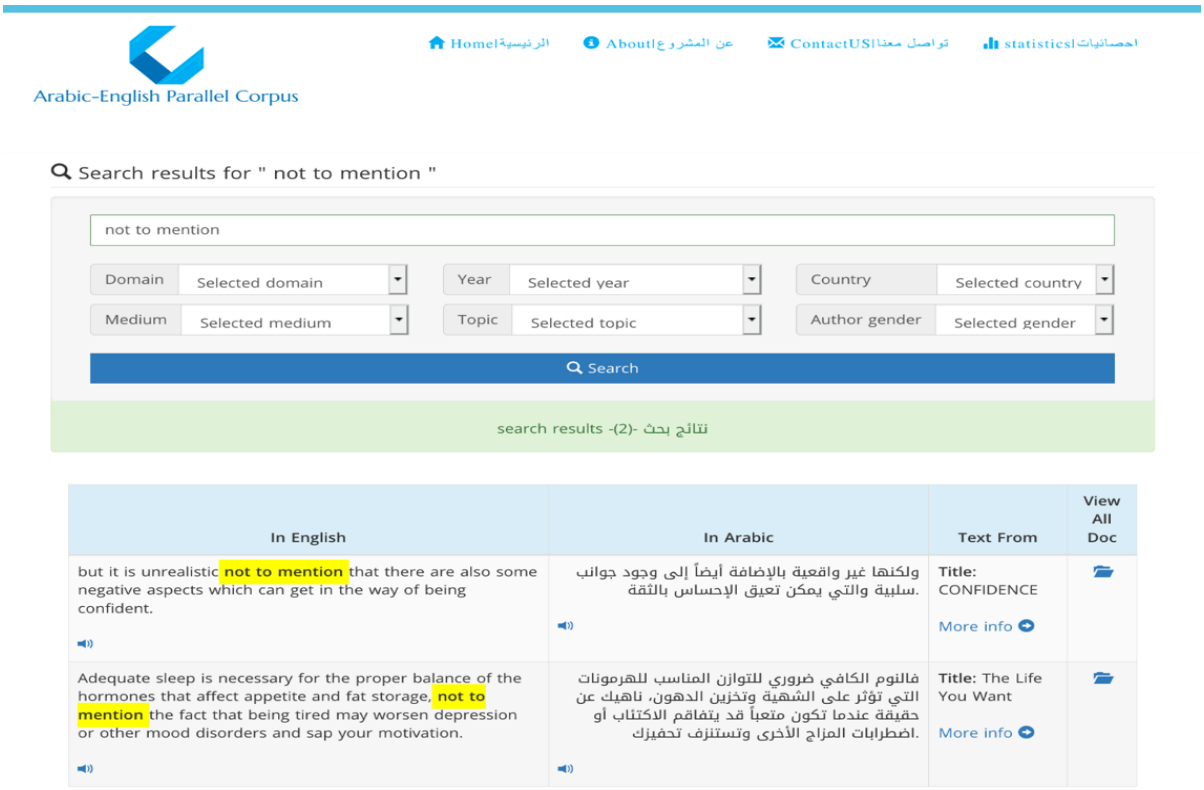
The damage caused by humans to the world's forests leads to changes in the natural environment and causes global warming. Governments should act to protect forests from illegal logging and plant more trees to absorb carbon dioxide. Deforestation on such a large scale is sure to have disastrous effects for the environment.

3.6.2 The Bilingual Corpus

The CAT+Corpus intervention involved providing students with a parallel corpus as a tool for learning the target collocations. It was therefore necessary to find a corpus to be used in one of the interventions. Unfortunately, there are not many options. For example, there is the *English-Arabic Parallel Corpus of United Nations Texts* (EAPCOUNT) developed by Salhi (2013), and the *English-Arabic Parallel Text Corpus* developed by Al-Ajmi (2003). Although the AEPC is still under construction as an ongoing project to help translation training and language teaching (Alotaibi, 2017), it was used as the corpus in this study (see Figure 3.5). This is because it is free of charge, is relatively more user-friendly than other corpora, has highlighted features for the target words, has audio options, aligns the texts for the two languages in a parallel display, and has sufficient semantically non-transparent transitional items as words and collocations.

Figure 3.5.

Arabic-English Parallel Corpus Showing Examples of a Collocation.



As mentioned earlier, the AEPC is still under construction, and it has some restrictions, particularly regarding the translations. The following explains some of the limitations concerning the target items in the study.

1) There are mismatched results from searching for some target items from Arabic to English and vice versa. For instance, *on the other hand* has its Arabic translation as *من وجهة نظر أخرى*; which means *from another perspective.* However, when it is done the other way around, there is only the literal English translation that denotes *viewpoint* (see Figures 3.6 and 3.7).

2) Since the target items are semantically non-transparent, some of the searches and

translations can be misleading and confusing when looking up a translation from Arabic to English. For example, when searching for the target item *thereupon*, the Arabic word عندئذ appears, which means *then*. Moreover, when the search is done the other way around, different translations are generated for the word *thereupon*.

3) Some of the target items have a translation from L2 to L1 only. For example, *the bottom line* has the Arabic translation الخلاصة, which means in *conclusion*; but, when it is done the other way around, there are no results at all.

4) Using the same example as in point 3, when searching for the collocation *the bottom line*, the searcher has to type *bottom line* without the article *the* to show the results and translations of the collocation *the bottom line*. This is because AEPC does not fully support English collocations.

5) Aside from the target items, there are some typographical errors in the concordance lines such as the word *does* instead of *dose* (see Figure 3.6).

To resolve these issues, the researcher printed out modified copies for L1 to L2 tasks from the AEPC. In fact, similar studies have used such a method previously (Alharbi, 2017). This will be discussed further in the Procedure section (section 3.11).

Figure 3.6.

AEPC Limitations with no Highlight of Arabic Target Items and English Typographical Error.

Figure 3.7.

AEPC Limitations with no Highlight of English Target Items.

3.6.3 Target Vocabulary Items

The 19 target vocabulary items used in this study are words and collocations (there were originally 21 items, (see section 3.9). They are made up of one, two, three, or four word units (see Table 3.2). In this study, most of the target items are linking adverbials e.g., *on the other hand*, with very few being nouns, such as the *outcome*, or the *aftermath*. Not only do all of the target items share the feature of having a non-literal translation, but they also have semantically non-transparent features. That is to say, single item words in compound words, and some collocations, may consist of two or three morphological elements that can have different meanings separately.

Some examples can be words such as *thereby*, *nevertheless*, *nonetheless*, and *whereas*. It should also be mentioned that other multi word units are idioms, for example, *on the other hand*, *in the aftermath*, *the bottom line*, *it boils down to*, *owing to*, *much less*, *in the long run*, *not to say*, *not to mention*, and *seeing that.* Table 3.2 illustrates the target items taught in intervention 1 (nine items) and intervention 2 (10 items).

Table 3.2

Target Items and their Introduction in the Study.

Intervention 1	*Intervention 2*
in the long run, let alone, thereupon, on the other hand, the outcome, the upshot, thus far, nevertheless, it boils down to.	*the bottom line, seeing that, thereby, owing to, henceforth, in so far as, the aftermath, not to mention, much less, nonetheless.*

Each target word and collocation were taught once in one teaching session. This contrasts with previous research that supports frequency of occurrence in teaching nonliteral collocations, and considers it as the main factor for L2 learning (e.g. Ramadhan, 2017). However, the present study is different in many to Ramadhan's study. In terms of L2 teaching methods, Ramadhan's study did not use CAT instruction and thus found that vocabulary with high frequency of occurrence was the main factor in collocation learning in L2. However, previous research (Alharbi, 2017; Laufer & Girsai, 2008a), which did not use frequency of occurrence in instruction, also had positive learning results.

These target items were selected based on certain justifications. Firstly, they had semantically non-transparent meanings; secondly, they did not have the same literal translations in Arabic, and finally, they included some idioms. Indeed, idioms can be challenging and even difficult for language learners. Although some parts of idiomatic phrases can have literal meanings,

there are no literal translations for them in their entirety. The only exception is with *not to mention* and *seeing that,* which can have literal and direct meanings, although how they are used in an Arabic context is different than an English context. For example, the collocation *not to mention* has a literal and figurative (indirect) meaning in the Arabic language, but the application of the figurative meaning differs. The figurative meaning in Arabic is not used in official speaking or writing, whereas in English, it is commonly used in unofficial and official speaking and writing. This can cause confusion as ESL learners can understand *not to mention* in the literal meaning. Similarly, the collocation *seeing that* has both similar literal and figurative meanings in the English language. In fact, *seeing that* is used as commonly in Arabic as it is in English. However, the researcher wanted to investigate the learning of *seeing that* as a figurative meaning by L2 learners receptively and productively. This was of particular interest to the researcher, especially because the target item *seeing that* denotes and connotes *see* in both literal and figurative meanings for English and Arabic.

Another important reason for choosing these specific target items, was that the researcher aimed to expand the students' vocabulary. Therefore, she searched for words and collocations that were common in English but uncommonly taught for L2 students in the English institution, or within their textbooks. A reason for choosing the linking adverbials among the target words and collocations, was that students were required to study them in high-intermediate English classes at the ELI. However, most of those in their textbook - *Unlock* (Westbrook et al, 2021) - if not all, have literal Arabic translations and include items that are congruent with meanings that can be easily guessed by L2 learners. Most of these linking items were repeated and were very common in students' previous textbooks, for example *however*, *furthermore*, *also*, *but*, etc. Thus, the target items in the reading materials used in this study replaced the original literal linking adverbials in

the course book materials. Second, the researcher aimed to help students connect their ideas and sentences in speaking and writing, and thus adverbials were selected among other parts of speech in the English language. Finally, the target items and their translations were included in the bilingual parallel corpus, which allowed participants to search for the 19 target items and see them in the text alongside their translations.

The chosen target words and collocations were reviewed by two highly experienced bilingual faculty members at the university. One of the reviewers had a background in Applied Linguistics and is currently an EFL instructor. They both agreed that these words and collocations can certainly cause confusion in everyday speech and writing, and thus including them as part of the study's teaching intervention would provide an important learning opportunity for the participants, hopefully expanding their vocabulary repertoire.

3.6.4 The Interventions

3.6.4.1 Experimental Group 1 (CAT +Corpus)

The 90-minute session (60 minutes for the intervention and 30 minutes for the immediate post-tests) began with 15 minutes of reading, which included 10 minutes of silent reading then five minutes of selected students reading out loud alternately. The passages were originally from the *Unlock* textbook (Westbrook et al., 2019). However, some modifications were made to the words and collocations in the textbook. After the participants finished reading, they had 15 minutes to answer the questions and complete the worksheets. This involved reading (see Figure 3.3) and discussing the passage's main ideas, supporting details, and some definitions (see Appendix A.2) in addition to answering the questions. Upon completion of the tasks, the researcher discussed the answers, but the passages and answers were not collected. The students were then asked to translate

the target items in bold in the passages using the bilingual corpus on their computers. They began with the receptive recall task, which required students to translate the target items from English into Arabic. This task lasted 15 minutes. The teacher asked the learners to read the authentic texts in the corpus, search for the target items in the corpus, and type in the Arabic translation next to them. There were times where the teacher redirected the students' attention to the correct sentences that included the correct target items, as some of the translations were inaccurate. The researcher also monitored the students and offered help if asked.

Once the participants had completed the translation task, the researcher provided accurate feedback and used the contrastive method briefly, but explicitly. This included explaining the similarities and differences between Arabic and English and clarifying the meanings. The aim was to draw the learners' attention to the target item within L1 and L2. For example *the bottom line* in Arabic has the literal meaning of a line, which is underneath, thus, the researcher helped students understand the actual meaning of the collocation *the bottom line* and how it differs from the Arabic translation. By using the combined approach of CAT instruction and a bilingual corpus, the researcher used translations provided by the bilingual corpus and helped learners distinguish between the semantically non-transparent target items and word-by-word translations.

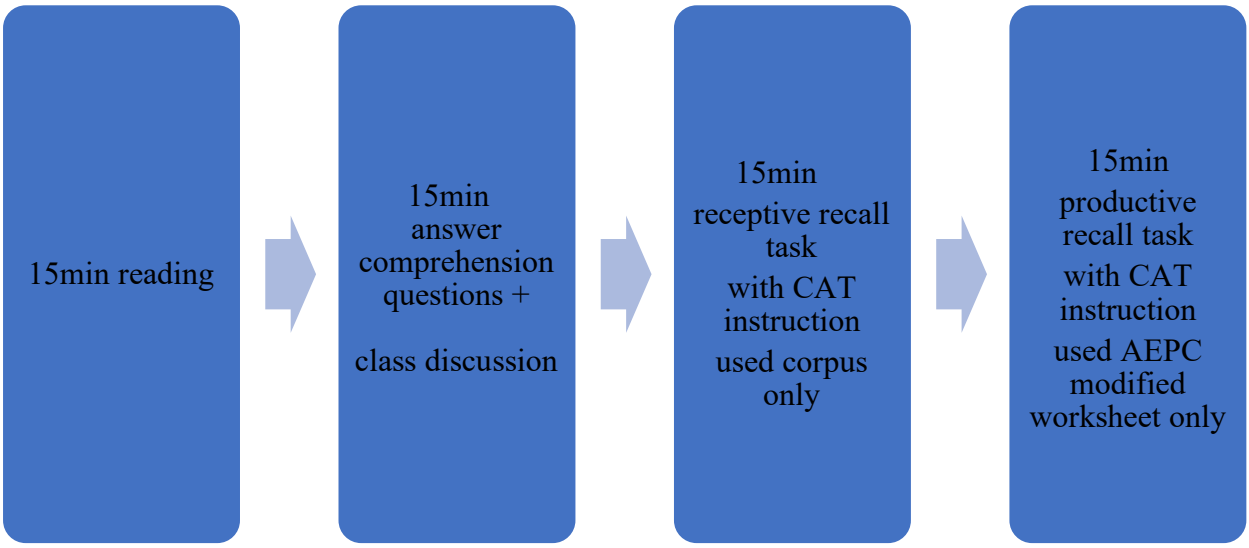
Upon completion of the receptive task, participants attended to the productive task. The responses to the receptive task were collected, in order for the next 15-minute-long task to be carried out without having access to their responses to the previous task. They were given a printed out and modified worksheet from the AEPC. These worksheets consisted of Arabic target items within sentences. Students were required to find the English translations of the target items. An example of this worksheet can be found in Appendix A.5.6.3.

Participants were free to ask the researcher for assistance if needed. The researcher

provided answers and feedback in English. Once the students had completed the translation tasks, the researcher used the contrastive analysis method with reference to the authentic texts in the AEPC. It should be mentioned that this task, too, took 15 minutes (see Figure 3.8).

Figure 3.8.

Intervention CAT+Corpus (60 min).

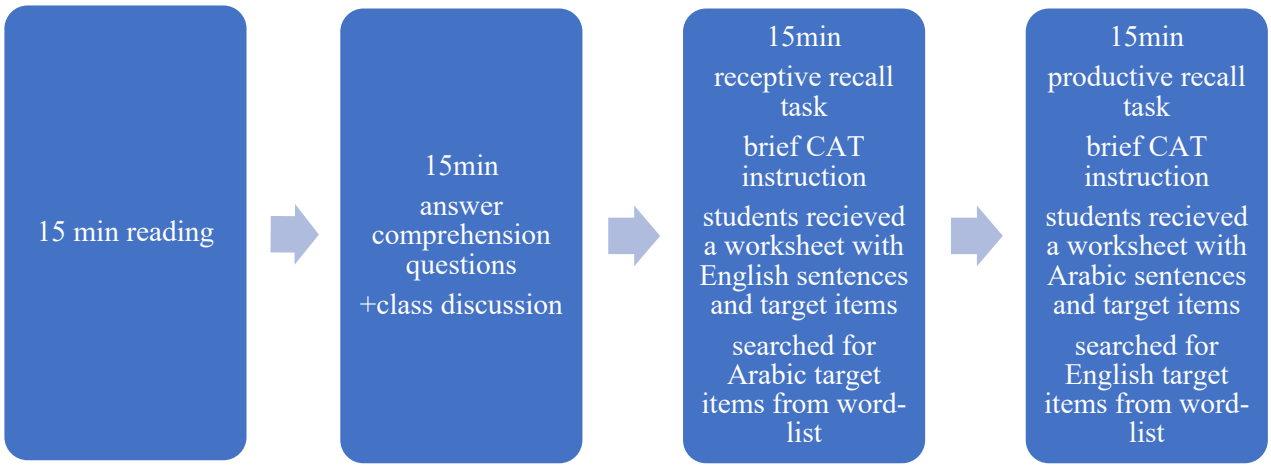


3.6.4.2 Experimental Group 2 (CAT)

This group went through the same first session (reading session treatment) as the CAT+Corpus group. Exactly the same amounts of time were allocated for their reading session, comprehension questions, and the two tasks. The latter, however, differed from those completed by the CAT+Control group. For the second session, the first task focused on translating from English into Arabic, whereas the focus of the second task was the other way around. Both of the tasks included some sentences together with a wordlist. With the aid of the teacher, the students matched vocabulary items from the word-list and wrote their translations next to them; they were also cautioned against providing word-by-word translation. The teacher monitored the class and answered any questions that the students had in English. Once the tasks had been completed, the researcher used the CAT method to explain the similarities and differences between the two languages and clarified the meanings in English. This was similar to the other experimental group, but without the corpus aid (see Figure 3.9).

Figure 3.9.

Intervention CAT (60 min).

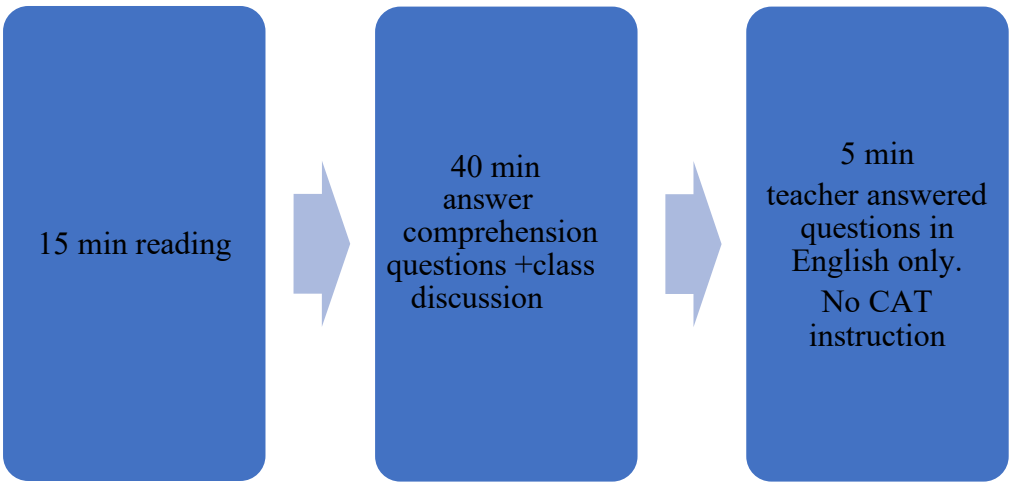


3.6.4.3 The Control Group (Traditional Lessons: No CAT, No Corpus)

The first session for the control group – that is, the reading session - was the same as for experimental groups 1 and 2. The second session was different once again. The students were exposed to the target items while reading the passages, but did not receive any translation or CAT instruction. In fact, more comprehension questions were given to the students in this group to ensure that they were exposed to the same amount of classroom instruction as the two experimental groups. The researcher monitored the students, answered their questions, and provided feedback in English. It should once again be emphasised that neither the contrastive method nor translation was used for the control group. Moreover, it is worth mentioning again that the researcher taught this group to ensure English was used without translations.

Figure 3.10.

No Intervention (60 min).



3.7 Instruments to Assess the Impact of the Interventions

Participants completed the tests (pre-test, post-test 1, post-test 2, and delayed post-tests) followed by TA interviews with a fill-in-the-gaps task to assess their learning of the target vocabulary items. Immediate follow-up interviews took place after conducting the TA interviews. All tests, that is the receptive test, productive test, and the fill-in-the-gaps task, were reviewed by two native English-speaking professors. The tests were:

1) Receptive tests (translating the target vocabulary from L2 to L1, see Appendix A.1).

2) Productive tests (translating target vocabulary from L1 to L2, see Appendix A.1).

3) Fill-in-the-gaps task (a test consisting of 19 exercises with a gap in each sentence; each exercise had three multiple choice options. The learner chose one correct answer to fill-in-the-gap (see Appendix A.3).

3.8 The Pilot Phase

3.8.1 Participants in the Pilot Study

Prior to undertaking the data collection, a pilot study was carried out in October 2019. Conducting a pilot study prior to carrying out the full data collection has been highly recommended. Gass and Mackey (2017) assert that:

> Pilot testing can often lead to revisions and fine-tuning of the protocol and can help to avoid costly and time-consuming problems during the data collection procedure. Careful pilot testing can also help to avoid the loss of valuable, potentially useful, and often irreplaceable data. (p. 52)

Two groups of high-intermediate English university learners were assigned to the pilot study. The first part of the pilot study was conducted towards the end of term one. Participants were in their first year at university and taking their first term in English. The first part of the pilot study was conducted with 11 students and included a pre-test, a fill-in-the-gaps test, one intervention using CAT+Corpus (using printed out modified copies for L1 to L2 tasks, and the digital version for L2 to L1 tasks), a post-test, a delayed post-test, and a fill-in-the-gaps test. All tests, excluding the fill-in-the-gaps, were for both the receptive and productive knowledges (see Appendix A.5 for the pilot study). The proficiency level of these participants was fairly homogeneous. In addition to the tests, one student was interviewed immediately after completing the tests. Initially, the plan was to interview two students, however due to time constraints, only one student was interviewed. The second part of the pilot study was conducted in term two with 28 participants and only included the pre-test for the receptive and productive knowledge.

3.8.2 Interviewing Participants in the Pilot Study

During the pilot phase, an interview was modelled on the principles of stimulated recall interview (SRI), although the main study adopted TA interviews. SRI is "a subset of retrospective reports that occur after task completion but include a video- or audio- recording to serve as a stimulus for the participant" (Bowles, 2010, p. 1). The aim of using SRI, which includes interviewing participants following test completion, was to better understand their thought processes and rationale for the strategies used to complete the tasks. In the pilot study, the participant completed a fill-in-the-gaps task, which was used as a prompt for the interview. However, after having used SRI in the pilot study, it became evident that the participant could not provide much detail on the strategies she used while completing the tests.

Furthermore, with an SRI method, the researcher would have had no guarantee that participants would commit to taking part in the interviews. A separate time would need to be allocated for participants to take part in the SRI, while with TA interviews, all the data can be collected at the same time. From a practical point of view, it would have been challenging to schedule SRIs (within the SRI recommended time frame) with all the participants in the main study.

3.8.3 Pilot Study Procedure

After receiving ethical approval (see Appendix D), the researcher immediately initiated the pilot study. The pilot study consisted of one intervention (involving one reading passage, follow-up comprehension questions, and receptive and productive tasks) given to the CAT+Corpus group, and one interview. One reason for choosing this structure was that the researcher could try out the combined approach of CAT+Corpus as it could provide more insight because of its comprehensiveness and inclusiveness, compared to CAT instruction only. Another reason was to

evaluate the allotted time for introducing the corpus (30-40 min) to the participants. Additionally, participants could provide feedback on the corpus used, which in itself could be helpful for the main study. Therefore, the researcher followed the protocol and piloted the study to allow time for revising and revaluating issues prior to the actual study. The purpose of the pilot study was firstly, to rehearse the intervention session and to see if the planned procedure was practical for the main study and, secondly, to check the validity and reliability (see sections 3.10 and 3.11) of the materials and instruments used (e.g., tests, reading passages, and interview questions). Thirdly, the aim was to understand the length of time required for teaching the intervention and for both the receptive and productive tests. Fourth, to identify any unexpected issues that might interfere with the study.

In order to ensure the reliability of the SRI, it was scheduled to take place 15-minutes after the fill-in-the-gaps task. The interview was conducted in a closed classroom and two audio recorders were used along with the fill-in the gaps task, which was the stimulus. An additional participant was supposed to take part in the SRI, although this did not take place due to time constraints. The aim of recruiting a second participant for the SRI was to gain greater insight into the employed strategies. The participants in the pilot study were students in an ELI course taught by a different teacher. At the time of the data collection, students were also approaching the examination period, so it was hard to recruit participants.

After conducting the pilot study, the SRI and the fill-in-the-gaps tests were revised and replaced with one TA interview and one fill-in-the-gaps task, which will be further discussed in section 3.8.5. This is because TA interviews involve collecting introspective verbal reports concurrently with L2 production, whereas SRI, as mentioned previously, is a retrospective report. Think-aloud (TA) interviews have the additional advantage of being able to capture students'

narration of their reasoning in real time while solving problems (Reinhart et al., 2022). Reflecting on the results of the SRI, which was conducted in the pilot study, the researcher felt that the participant was not able to describe all the thought processes, which took place while she was completing the fill-in-the-gaps task. This was probably due to the fact that the SRI was conducted after the participant had completed the task. Therefore, it was decided to adopt TA interviews rather than SRI for the main study.

As previously mentioned, due to time constraints and exams, the timeframe of the pilot study was affected. The pilot study did not have a gap between the pre-test and the intervention and the post-test; they all took place on the same day. The second fill-in the gaps test and the SRI were conducted two days after the intervention, but before the delayed post-test. However, the SRI should have originally taken place after the delayed post-test (see Table 3.3 and Figure 3.11). The delayed post-test was conducted eight days after the initial testing and intervention. Indeed, the reason behind this modified arrangement was once again due to the tight schedule and students' exams.

Table 3.3

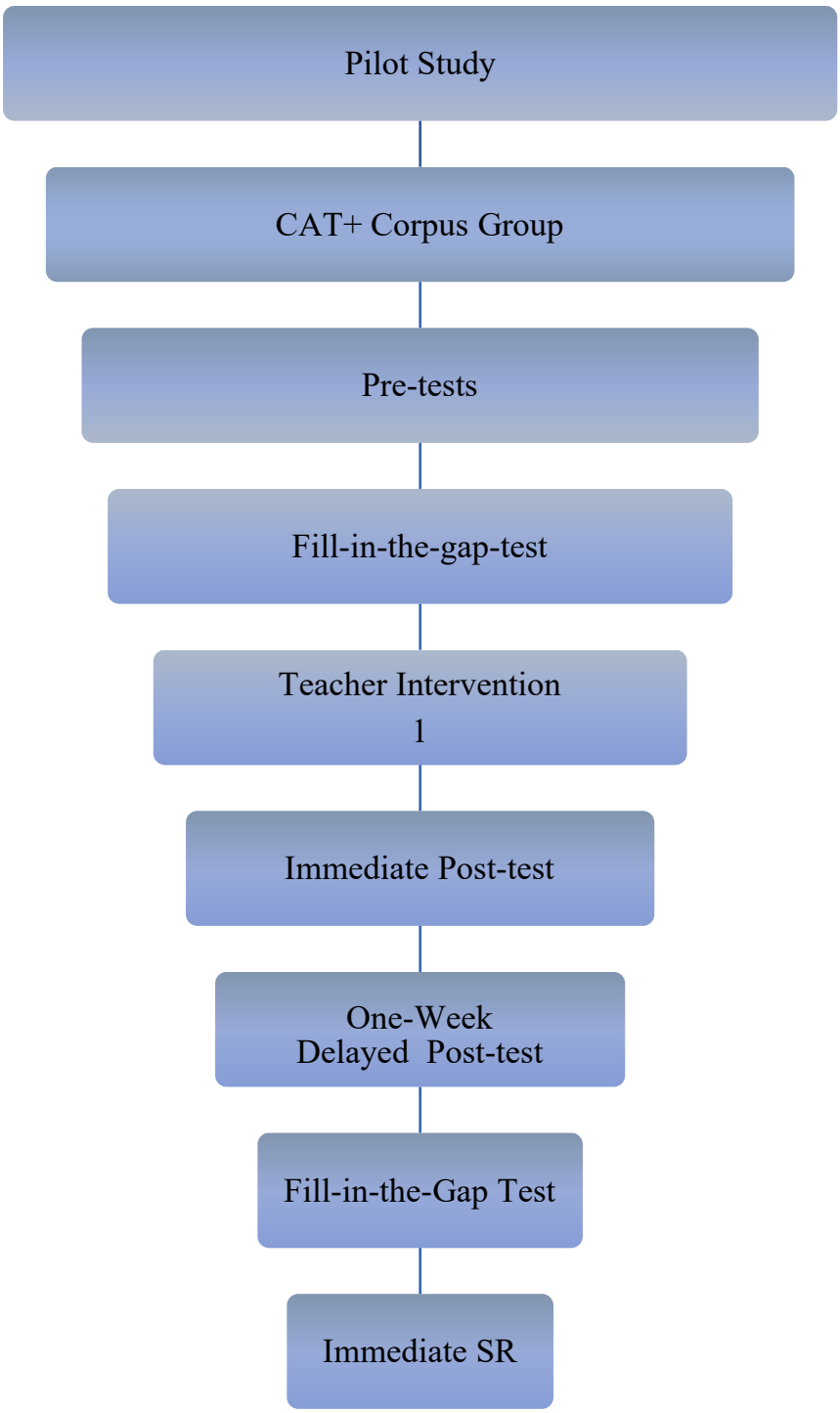
Procedure and Timeline for the Pilot Study.

Group/Term	Date	Procedure
Group 1/Term 1	October 6, 2019	Pre-test Fill in-the-gaps test Intervention Post-test
	October 9, 2019	Fil- in-the-gaps test Immediate SRI
	October 14, 2019	Delayed Post-test
Group 2/Term 2	October 30, 2019	Pre-test Fill-in-the-gaps test

The second part of the pilot study included 28 participants all from the same class. They completed pre-tests and one fill-in-the-gaps test. The tests were administered on the same day with a 15-minute gap between them (see Table 3.3 for timeline). The purpose of this second part of the pilot study was simply to explore if students in this level of proficiency were familiar with the target items, and to assess how well students could complete the tests. For example, the researcher observed that students found the sentences that included some words from the receptive tests to be more complex, since many students asked about the meaning. Thus, it was decided to change the words *barely* to *hardly*, and *obliged* to *required.* This adjustment was done to help the students read the sentences more easily and to understand the meanings of the words surrounding the target items.

Figure 3.11.

Pilot Study Outline.



3.8.4 Quantitative Analysis for the Pilot Study

Data from the pilot study were analysed to assess whether the procedures chosen would enable the researcher to address the study's first research question: *what is the impact on short and long-term learning of semantically non-transparent words and collocations (form and meaning--recall) among high-intermediate learners of English of a combined approach using both a parallel corpus and direct teacher instruction of CAT?* Although this is the research question that was used initially to guide the pilot study, in light of the findings from the pilot phase, this research question was later revised and updated. This is discussed in more detail in section 3.11.1.

After collecting the quantitative data from the 11 participants in term one, and from the 28 participants in term two, it was important to determine the marking scheme. Initially, it was planned that each correct answer would be marked as 1, and each wrong answer as 0. However, given their overall proficiency level, it was decided to award 1 for words which were recognisable but misspelt. For example, the productive test required students to translate from Arabic to English. If the word was misspelt, but the pronunciation of the word remained the same without altered meaning, then it would score 1. A score of 1, for instance, would be given to a student who wrote *seeing thet* instead of *seeing that.* An example of a wrong answer would be writing *the beether line* instead of *the bottom line*. Concerning the receptive test, it required students to translate from English to Arabic. If the students gave one or more answers with correct translations, then that scored 1. However, if there were a few answers and one was an incorrect translation, then the score would be 0. For example, if the student gave two correct translations for the target item *the bottom line*, that would count as correct. However, if the translation included a literal meaning for the

target item (e.g., the line that is at the bottom) then this would be counted as incorrect. If the participant did not give an answer to a question, that would also score a 0.

The receptive and productive tests were marked and revised by the researcher with a gap of one week in between to check the marking of the scores. For both tests, the marking procedure involved was identical. A rater, who was an ELI instructor, also revised the marking of the scores of the productive and the receptive tests for the CAT+Corpus group, which consisted of 11 students. An inter-rater reliability was calculated for both tests. The researcher and the rater agreed on 97.53% of the scoring for the productive test, and on 98.35% for the receptive test. The maximum score for each test depended on the test itself. For example, in the pre-test and the delayed post-test, the maximum score was 21. In the post-test, the maximum score was 11.

There are two major ways to interpret data statistically: descriptively and inferentially, both of which were utilised in this study. Regarding descriptive data, it is important to explore the frequency distribution or histogram by looking at the general tendency of the collected data without making any inferences. Additionally, it is important to present the centre of the distribution, which includes the mean, median, mode, as well as the minimum and maximum score measures of dispersion (e.g., standard deviation) (Field, 2018).

However, inferential analysis makes inferences and assumptions based on the data. Four assumptions were considered for implementing an appropriate statistical test, although not all these assumptions were relevant to every test. These assumptions were: "additivity and linearity; normality of something or other; homoscedasticity/homogeneity of variance; independence" (Field, 2018, p. 173). Since the 11 participants were randomly assigned and they were tested three times (pre-, post-, and delayed test) with the dependent variable (score 0 or 1), Friedman's test was used as an alternative to the one-way ANOVA. The Friedman's test, as other non-parametric tests,

can be used with a sample size as small as five, which made it suitable for analysing the pilot study data. Non-parametric tests are also appropriate where the population distribution is not precisely known (Dallmeier et al., 2013).

The following sections will discuss descriptive and the inferential data analysis conducted on the first group of 11 participants, starting with the test of productive recall and followed by the test of receptive recall.

The Productive Recall Test:

Table 3.4.

Descriptive Statistics on Productive Test Scores for the CAT+Corpus Group of 11 Participants in the Pilot Phase.

Tests	Mean	*SD*	Median	Range
Pre-test	.0	.0	.0	.0
Post-test	8.27	2.34	9	8
Delayed Post-test	4.90	2.46	5	7

Note. The maximum score for each productive test is10.

Table 3. presents descriptive statistics of the productive test scores for the CAT+Corpus group. The median for the pre-test, post-test, and the delayed test was (.0), (9), (5), respectively. The results showed that there was an improvement between the pre- and post-test. These results also showed that students had no prior knowledge of the target items, since they all scored 0 on the pre-test. However, following the intervention, participants were able to productively demonstrate that they had recalled most of the target items. Furthermore, in the delayed post-test, mean scores declined, but still demonstrated evidence of long-term recall of the target items.

The Friedman test revealed that there was a significant difference between the three tests $\chi 2(2) = 21.54$, $p < .001$. Therefore, further pairwise tests were carried out using Wilcoxon signed rank tests. A Bonferroni correction was applied, resulting in a significance level set at $p < .017$. There were significant differences between the pre-test and post-test ($z = -2.96$, $p = .003$), between the pre-test and delayed post-test ($z = -2.95$, $p = .003$), and between the post-test and delayed post-test ($z = -2.84$, $p = .004$).

Receptive Recall test

Table 3.5.

Descriptive Statistics on Receptive Test Scores for CAT+Corpus Group of 11 participants in the Pilot Phase.

Tests	Mean	*SD*	Median	Range
Pre-test	3.72	2.53	3	7
Post-test	6.91	2.34	7	8
Delayed Post-test	7.36	2.46	8	7

Note. The maximum score for each receptive test is 10.

Table 3.5 reports descriptive statistics of the receptive test scores for the CAT+Corpus group. The results indicate that while participants had some prior understanding of the target items, which can be seen by the mean score in the pre-test, scores improved significantly after the intervention. In fact, both the post-test and delayed post-test had a higher mean compared to the pre-test. Interestingly, the mean score for the delayed post-test was slightly higher compared to the immediate post-test. This seems to suggest that the intervention was effective in improving participants' recall of the target items both immediately after the treatment, and in the longer-term.

The Friedman test for the receptive scores indicated that there was a statistically significant difference between the three tests (pre-, post-, and delayed), ($\chi 2(2) = 15.84$, $p < .001$). Given these significant results, further pairwise tests were carried out using the Wilcoxon signed-rank test. A Bonferroni correction was applied, resulting in a significance level set at $p < .017$. There were significant differences between the pre-test and post-test ($z = -2.67$, $p = .008$), between the pre-test and delayed post-test ($z = -2.95$, $p = .003$), although there was no significant difference indicated between the post-test and the delayed post-test ($z = -1.29$, $p = .194$).

3.8.5 Qualitative Analysis of the Pilot Study

At the time the pilot study was conducted, the second research question was formulated differently. The original RQ2 was: *what learning strategies do learners employ within each approach*. In the main study, and in light of what was revealed in the pilot phase, RQ2 was revised to: *what strategies do learners report using to understand semantically non-transparent words and collocations in a think-aloud task?* The revised research question was more specific and was re-written to incorporate the think-aloud interview method, which was not used in the pilot study.

In order to address the second research question during the pilot phase, the researcher used the SRI method to gather data. Using SRIs requires researchers to follow certain guidelines. For example, it is recommended for the SRI to be carried out promptly after the task to avoid memory loss, or no later than three hours after the event (Gass & Mackey, 2017). Therefore, the researcher interviewed one student immediately after the fill-in-the-gaps test, which took place two days after the tests and the intervention. Furthermore, the researcher took into account the language background of the participant, using Arabic to conduct the SRI. Indeed, using the right language to conduct SRIs is of paramount importance to ensure all participants understand the interview

questions (Gass & Mackey, 2017). In addition, the researcher considered the structure and wording of the questions, as unclear questions can compromise answers; therefore, the questions were designed to be clear and appropriate for the specified time frame. Finally, the researcher used the fill-in-the-gaps test as a stimulus, which is key to gather sufficient data from participants (Gass & Mackey, 2017).

After conducting the interview, the first thing the researcher did was to transcribe it into both Arabic and English. The researcher read the transcripts thoroughly with no expectations in mind. When encountering common themes and patterns, the researcher would add notes. Accordingly, coding was applied following Saldaña's (2009) procedures, which was to apply initial coding and recoding with a "more attuned perspective" (p10). The following transcript is a small segment from the interview. The segment starts with the researcher asking the student to look at the fill-in-the-gap tests and recall examples of learning. The examples were: *not to mention* and *owing to.* The participant referred to the following two fill-in-the-gaps questions mentioned here:

1) The costs, _____ the risks, of moving the satellite to and from space would have been difficult.

a) not to mention
b) whereas
c) thereby

2)_____ the heavy traffic, he was late.

a) The upshot
b) Seeing that
c) Owing to

In the SRI, the participant described the way she recalled her thought processes while completing the tasks.

R: Can you give me another example as well?

S: [Student starts by reading from the fill-in-the-gaps test] Not to mention, the cost. [Student then explains] then there was *the risks*, which means *naheek an* (an Arabic translation for not to mention). I quickly understood I'm talking about something, *naheek an* the other details of the other things.

R: So again for you the Arabic was helpful?

S: Yes it was helpful.

R: Can you remember anything unique or special while you were doing the test?

S: Here [pointing and reading the example _____the heavy traffic, he was late]. In the first test, the answer was for sure *because* , I was thinking it was for sure *because*. But I did not ever expect *owing to,* I think I put *seeing that.* But now I feel *owing to* is very logic.

Indeed, the segment shows there are indications of potential learning. The participant was looking at her second fill-in-the-gaps test and was explaining to the researcher the difference between her answers. In the first fill-in-the-gaps test, which she completed prior to the intervention, she assumed that the correct answer would be *because*; however, the word *because* was not included among the options. She did not expect the collocation *owing to* to be correct. After the intervention, she was able to provide the correct answer and stated that *owing to* was the logical option.

Although the SRI was able to provide insights into participants' learning, the decision was taken after the pilot to use a think-aloud (TA) interview instead. It was decided that a TA interview would provide a more profound understanding of learning semantically non-transparent words and collocations for the following reasons. Initially, in terms of practicality and time constraints, the researcher could guarantee being able to conduct all interviews within the first 48 hours of taking the task as recommended by Gass and Mackey (2017) for SRI. In fact, however, SRI would require two different settings; one to take the fill-in-the-gaps task and the other to conduct the interview. As such, having them both within the first 48 hours might not be practical for participants'

schedules, and thus having a TA interview was more feasible in terms of conducting one interview in one setting, allowing better opportunities to collect data compared to the SRI. It is also important to note that the repetitiveness of different tests could affect students' learning. In fact, the participant mentioned that the tests got easier because of the repetition of the words.

Therefore, it was deemed important to reduce unnecessary exposure to the target items and to only administer the productive and receptive tests. In doing so, the fill-in-the-gaps would not be used as one of the tests, but as a task within the TA interview. Thus, the SRI was abandoned and a think-aloud interview was used instead to provide deep insight into the process of learning while performing the fill-in-the-gaps task.

3.8.6 Importance and Limitations of Pilot Study

Having conducted the pilot study, the researcher was able to revise and rehearse the procedure and the instruments for CAT instruction and bilingual corpus. The statistical data showed significant results for both productive and receptive tests, which indicated progress in learning under CAT instruction. However, results also suggested that learning was more effective for receptive knowledge, where scores remained high even in the delayed post-test. The delayed post-test for productive knowledge saw a decline in the overall mean score compared to the immediate post-test. Participants did express that a couple of words, that were not from the target items but part of the sentences, seemed difficult. The two words were *barely* and *obligated*, and thus they were changed to *hardly* and *required*, respectively. The reason for these modifications was to make the vocabulary in the sentences easy and understandable so that the learners could focus on learning the target items and not get distracted with unknown vocabulary. Regarding the corpus tutorial, the researcher estimated that the participants would take 30-40 minutes to familiarise themselves with it. However, participants quickly learned how to use the online tool

and commented on its ease of use. Therefore, 15 minutes were allocated to explaining the use of the corpus in the main study.

Regarding the receptive and the productive tests, two main issues came up in the pilot. In the productive test, the students had to translate from Arabic to English. The students seemed intimidated by the questions as they only had two English letters at the beginning of a word or phrase and they were required to complete it (see Appendix A.5). However, during the test, the researcher explained that it was alright if they did not know the words and asked that they do their best. Some students left them empty while others wrote words that were different to what was requested. To resolve this, the researcher specified in the Arabic instructions on the productive test that this task might be intimidating, but that it is important to complete the word or phrase and not give other words. The researcher made every effort to encourage participants to do their best and reminded them that the results would not affect their GPA at any point.

Furthermore, the allocated time was expanded for the pre-test and the delayed post-test as they both had 21 items, whereas the post-tests had almost half the number of items. Thus, instead of 20 minutes, the pre-tests and delayed post-tests in the main study were extended to 30 minutes, and the post-tests were left at 15 minutes. The researcher decided to increase the allotted time based on observations of the participants' experiences in pilot group 1 and pilot group 2, where the first group had 10 target words and completed the tests within 15 minutes, but the second group requested more time to complete the test, as 20 minutes were insufficient.

In the receptive test, participants in pilot group 1 and pilot group 2 took longer than anticipated. The allotted time was 20 minutes and the test requested full translations of sentences from English to Arabic. Because of the time constraints, many participants just wrote the meaning of the underlined word and not the whole sentence as required. Therefore, this may have affected

the results. Thus, the timing was modified in the main study. The pre-test, since it consisted of translating more sentences than the post-tests, was allotted 30 minutes for the receptive test and the productive test. Participants took a 15-minute break between the productive and receptive pre-tests, which was useful because it distracted the participants from relating the two tests to each other.

Is also important to note that the fill-in-the-gaps tasks might have affected the results of this study. This is because all participants in pilot group 1 were given a fill-in-the-gaps task, which took place after the intervention and the post-test, but before the delayed post-test. Based on the results of the fill-in-the-gaps tasks, the researcher selected an interviewee and used the tests as a stimulus.

However, it was pointless to have all the participants complete the fill-in-the-gaps task and only select one interviewee for the SRI. In fact, the researcher felt it gave unnecessary exposure to target items, which could have affected the results of the data. This was confirmed by the interviewee who said that seeing the target items over and over again helped in her learning. This can be confusing because she had also described learning from CAT instruction. This confusion can be avoided by using another interview method that provides less exposure to the target items and provides a deep understanding of learning. Hence, the SRI was revised and replaced with a TA method. The second research question was adjusted accordingly.

There were some limitations to the study, such as the small selection of bilingual corpora in Arabic. Accordingly, the researcher used the most user-friendly corpus; however, the corpus had limitations in the way the target items were translated (see section 3.6.2). Another limitation relates to the structured syllabus in the English Institute where the study was conducted, as it runs every six weeks for one term, unlike other departments in the university that have approximately

twelve weeks in one term. Indeed, this structure causes conflicts on some occasions with the English classes. For example, some students might skip classes to study for major exams in other subjects, and thus, the data collection is affected. Additionally, all the participants in this study were female, and this was due to cultural limitations and the tradition of gender segregation in Saudi universities. Moreover, because of cultural reasons, most female students are not comfortable with interviews and recordings. Therefore, the researcher explained the benefits of conducting the interviews and assured participants that no one else would have access to the recordings.

3.9 Reliability

Reliability refers to the consistency of a measurement. For the sake of research reliability in both the pilot and the main study, the researcher taught the target items both in CAT for the two treatment groups, and in the lessons for the control group. The main reason for this was to ensure that no incidents of translation would occur in the control group from the teacher nor from using any external devices such as smartphones, which was common for students to do. In fact, when the teacher taught students in the control group, the students who asked for clarification of target items or collocations only received explanations and definitions in English. The researcher also ensured that the control group did not receive computer corpora tasks nor CAT activities and that they read the passages and answered comprehension questions only.

The researcher conducted Cronbach's alpha reliability test on the 21 target items (see Tables 3.6 and 3.7). The Cronbach's alpha for the productive test was higher than .7, but it was below .7 for the receptive test. Other studies in L2 also had similar outcomes when following the reliability analysis. Dörnyei and Taguchi (2010) kept the original item wordings in one of their

questionnaires in their study even though the alpha was (.64). This was because in order for their study to be conducted on Asian participants, it was crucial to meet the same measurements and construct as another study conducted by Dörnyei et al. (2006) on Hungarians. In sum, under certain conditions, a slightly lower alpha is accepted as part of a process that seeks:

> to improve the overall reliability of a scale by identifying and then deleting items that reduce the scale's Cronbach Alpha. This is a stepwise process—we need to focus on one item at a time—and sometimes it requires several steps (i.e., the deletion of several items) before adequate reliability is achieved (at least 0.70). It can also happen that we reach a point when no further deletion would cause any improvements even though the scale is still not sufficiently reliable—in such cases we either exclude the scale from the analyses or we accept it as a less than perfect solution. (Dörnyei & Taguchi, 2010, p. 119)

Therefore, following Dörnyei and Taguchi's (2010) justification concerning reliability, item 20 was omitted for the main study analyses, which was the word *whereas,* thus increasing the Cronbach's alpha to .680.

Table 3.6.

Cronbach's Alpha for the Productive Tests for the Pilot Study.

Cronbach's Alpha	N of Items
.714	21

Table 3.7

Cronbach's Alpha for the Receptive Tests for the Pilot Study.

Cronbach's Alpha	N of Items
.664	21

After months of searching for appropriate target items and contacting the developer of the bilingual corpus, the researcher initially selected 21 words and collocations. However, after conducting the pilot study, the main study had a total of 19 words and collocations. This was mainly for two reasons. First, to increase reliability, item 20 was removed (see Table 3.8). Second, the collocation *not to say* was omitted from the analyses because it had the literal meaning in the bilingual corpus and did not align with the purpose of the present study. It is worth mentioning again that a main factor that played a big part in selecting the semantically non-transparent target items was their availability within the bilingual corpus (see section 3.6.2 for limitations of the bilingual corpus).

Table 3.8.

Pilot Study – Cronbach's Alpha if Item Deleted

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
x1	7.68	9.337	.436	.628
x2	7.61	9.877	.267	.650
x3	8.07	10.365	.166	.660
x4	7.46	9.888	.332	.643
x5	8.04	9.962	.303	.646
x6	8.00	10.296	.157	.662
x 7	7.93	10.513	.063	.673
x 8	8.21	10.915	-.013-	.668
x9	7.86	9.831	.275	.649
x10	8.00	10.296	.157	.662
x11	7.46	10.110	.245	.652
x12	8.00	10.000	.266	.650
x13	8.04	9.962	.303	.646
x14	8.14	10.646	.093	.665
x15	7.57	9.810	.302	.646
x16	7.46	9.517	.482	.627
x17	7.68	9.856	.261	.650
x18	7.93	9.402	.448	.628
x19	7.79	10.026	.202	.658
x20	8.00	10.815	-.026-	.680
x21	8.07	10.365	.166	.660

Note. x20 is the target item *whereas.*

3.10 Validity

Validity of a research instrument refers to "whether an instrument measures what it is set out to measure" (Field, 2018, p. 13). A common distinction is made between internal and external validity. Internal validity refers to the extent to which the findings can be interpreted with confidence, while external validity is concerned with the extent to which the results of the study can be generalised (Bryman, 2012). One aspect of internal validity, known as content validity, refers to evidence that a research instrument - such as a test - corresponds to the content of the construct it was designed to cover (Field, 2018). In this study, both the pre- and post-tests assessed all of the 19 target items which were included in the teaching interventions. Criterion validity, also an aspect of internal validity, is concerned with whether it is possible to establish that an instrument measures what it claims to measure by comparing it to objective criteria (Field, 2018). The addition of a control group to the study also had the aim of increasing internal validity. Without a control group, it is hard to determine if a change (e.g. in test scores) has resulted from the treatment.

Furthermore, to ensure validity of the TA interviews, the researcher reminded the participants to think-aloud while completing the fill-in-the-gaps task, whenever it was required. There were also red circles next to each question as a reminder to think-aloud. Participants should not remain silent for more than one minute during the interview. It was important to remind students to describe their thought processes while completing the task, since it is normal for test-takers to remain silent when concentrating on test items. The researcher also included participants from the control group to perform the same fill-in-the-gaps tasks.

Ecological validity is one aspect of external validity where researchers examine whether research results represent what happens in everyday life, and therefore can be generalised to real-life settings. To improve the ecological validity of the results obtained from the pilot study and the

main study, the reading materials and tests were taken from the students' English course books. Although a few adjustments and modifications were made to fit the study's design, these were reviewed by two native English speakers. The reviewers were asked to look at the teaching intervention materials and the tests before these were administered to the participants. They were asked to review the items in terms of appropriateness and clarity. Feedback received from the reviewers focused primarily on wording and grammatical accuracy. The reviewers also confirmed that the materials and tasks were appropriate, thus strengthening the face validity of the instruments.

Furthermore, efforts were also made to prevent cheating in order to ensure that the test scores would be reliable. During the tests, the researcher invigilated the students and arranged their seating. The two interventions were also taught by the researcher, who was also a full-time instructor at the ELI. This was done to ensure consistency in the delivery of the intervention across the three different groups. Furthermore, it was previously observed that students have a tendency to take their full-time instructors more seriously than other staff members, who may take over a teaching intervention for data collection purposes only. Thus, the researcher taught all three groups to ensure participants would perform to the best of their ability on the tests.

3.11 Main Study Procedure

3.11.1 Quantitative Data Collection Procedure

The experimental design was implemented through several steps (see Table 3.9). In order to accurately assess participants' previous knowledge of all 19 of the planned target items at the outset, everyone across the three groups took the pre-test in week one. This was to ensure that participants were not already familiar with these words and collocations and to evaluate their existing receptive and productive knowledge of the target items (form and meaning). There was a

15-minute irrelevant task in between the productive recall tests and receptive recall tests and each of these tests took 30 minutes.

During the second week, there was one 90-minute intervention for the two treatment groups only. They were taught nine items (see Table 3.2). However, the same amount of time was allocated to the control group through traditional lessons. The teaching sessions included a reading passage along with related comprehension worksheets and in-text questions. None of the comprehension questions were related to the target items. It was important to ensure that participants across the three groups were given an equal amount of time to complete the tasks, and thus 30 minutes were allocated for everyone, regardless of the teaching method used, to read the passage and do the tasks.

The two treatment groups and the control group were assessed through an immediate post-test. The post-test (productive and receptive test) took 15 minutes each and included a 5-minute irrelevant task in between. This cycle was then repeated for the remaining 10 target items in week three (see Table 3.2). In week four, the participants completed the delayed post-tests, assessing knowledge of all 19 items. The questions in all the immediate post-tests and delayed post-tests were identical to those of the pre-test. The only difference was that all of the 19 target items were included in the pre-tests and delayed post-tests, while the immediate post-tests contained nine or 10 of them. However, only the data from the pre-test and the first post-test was used for the quasi-experimental study because many students were absent during the second post-test and the delayed post-test.

The administered tests assessed receptive and productive skills. The receptive tests involved translating the sentences with the target vocabulary from L2 to L1. The target items were underlined in each L2 sentence. The translation would be written under the English sentence. By

doing so, the learner demonstrated receptive recall. The sentence was considered correct if the overall sentence had the same meaning as the L2 original sentence; and if the target vocabulary was translated correctly. The researcher specifically told the students to provide a translation of the underlined words within the sentence. Furthermore, the productive tests involved translating an underlined Arabic word or collocation, presented within a sentence, into English. The first few letters from each English target item were provided. Participants were required to complete these words and collocations.

One could argue that providing a few letters can trigger participants to recall form but not meaning. However, the main reason for providing a few letters was to ensure students wrote the exact target item and to avoid other English translations. Rather than expect a full recall of the target items, the researcher used the first letter approach to clearly measure the students' knowledge and, as far as possible, determine accurately if they knew the target item or not. Furthermore, the productive test is a type of assessment that is not new to English proficiency tests. In fact, the Duolingo English Test, an English measurement test, consists of the C-Test that requires test takers to complete words with missing letters. The task includes a paragraph with complete first and last sentences, whereas words in the intervening sentences are *damaged* by removing the second half of the word. Test takers rely heavily on understanding the context and the meaning (LaFlair & Settles, 2019). Both the C-Test and the current productive test have similarities in that they both rely on understanding the context and the meanings of the provided sentences, and both require responding to tasks of completion. However, English is used in C-Tests whereas Arabic and English are used for the current productive test. It is vital to note that for every test, learners first completed the productive test before the receptive test. If done the

other way around, the participants could be exposed to the target items, which could compromise the accuracy of the productive test scores.

The aim of the tests was to evaluate the effectiveness of CAT and CAT+ Corpus interventions on the participants' learning of the target items. Having a separate test for receptive and productive knowledge allowed the researcher to compare results and observe any differences between performance in both skills. Because the tests were developed by the researcher herself, the sentences used in the tests were modified to improve comprehensibility and checked for cultural appropriateness (for example, Arabic names were used). There is, therefore, a potential subjectivity within the tests, since these were designed specifically for the participants in the study. However, the English sentences used in the tests were taken from several websites such as the British Council, and therefore were examples of authentic language use. An advantage of the modifications following the pilot study is that this ensured that the instruments would not include unclear wordings. Furthermore, while the researcher developed the test items, these were reviewed by other experts in the field.

Regarding the scores of the tests, in the productive test, participants were awarded a mark of 1 for writing the correct target item in the fill-in-the blank test, even if the spelling was incorrect (as long as the term was recognisable). A mark of 0 was awarded for an incorrect word. As for the receptive test, if the participants wrote the correct translation, or an acceptable alternative with a similar meaning, they would be awarded a mark of 1, however if they wrote an incorrect translation or multiple answers including an incorrect translation, they would receive a mark of 0. While the approach towards scoring the test has potential subjectivity, since the researcher determined how marks would be awarded, it takes into account the different possible ways of translating the target items into the L1. It further does not reduce participant scores based on minor spelling mistakes

made in the productive test. In this way, the subjective nature of the scoring method has the advantage of gaining a better picture of the participants' actual knowledge of the target items. Some computerised scoring methods, for example, would only consider an answer to be correct if it was spelt completely accurately. Furthermore, it was decided to award only marks of 0 and 1 in order to facilitate the statistical analysis of the scores.

Also, to help familiarise those in the experimental group with the bilingual corpus, a 15-minute briefing session was organised for them to not only introduce them to corpora in general, but also to provide them with a thorough explanation of the usages of the AEPC in language learning. This helped the participants learn how to take the necessary steps when it comes to searching for a target item and finding it highlighted along with its Arabic translation. In other words, this enabled them to compare and contrast the Arabic and English words with each other. However, neither the second experimental group (CAT) nor the control group received such training, as the latter received no treatment, and corpora were not included in the teaching sessions of the former.

Furthermore, the first question for the quantitative data was modified. Initially, the research question aimed to explore the effects of CAT instruction both on short-term and long-term recall. However, because not enough participants took the delayed post-test, there was insufficient data to examine long-term effects of the treatment. The first research question was therefore revised to only take into account implications of CAT instruction for short-term learning.

Table 3.9.

Timeline for the Three Groups.

Term One (6 weeks)

WEEK 1 (Day 1)
- Pre-tests (19 items)

WEEK 2 (After 5 days)
- Teaching sessions for the three groups
- Immediate post-tests (10 items)

WEEK 3 (After 5 days)
- Teaching sessions for the three groups
- Immediate post-tests (9 items)

WEEK 4 (After 5 days)
- Delayed post-tests (19 items)

WEEK 5
- TA interview and fill-in-the-gaps task (19 items)
- Immediate follow-up interviews (for treatment groups)

3.11.2 Think-Aloud Interviews and Fill-in-the-Gaps Tasks

After the delayed post-tests, four participants from each of the three groups completed a fill-in-the-gaps test while being interviewed. The fill-in-the-gaps was an English only task. It consisted of 19 exercises comprising sentences with a gap. Each sentence had three multiple choice options. Each option contained possible words to be used to complete the sentence. The learner had to choose one correct answer to fill in the gap. The participants were asked to think-aloud during the process of selecting the answers to the fill-in-the-gaps exercise. The reason for this was

to gain a deeper understanding into the thought processes of each group – that is, CAT+Corpus, CAT, and control groups.

The criteria for selecting the interviewees went through several adjustments. At first, the researcher aimed to select, from each of the three groups, four participants who progressed differently between both their pre-tests and post-tests. In other words, the researcher originally intended to select four participants/interviewees based on their scores: 1) pre-test high vs. post-test low; 2) pre-test low vs. post-test high; 3) pre-test and post-test low; and 4) pre-test and post-test high. However, this criterion was adjusted and the researcher selected the interviewees based on their scores on the receptive delayed post-test only. The reason for this adjustment is that the receptive delayed post-test consisted of all 19 vocabulary items that were included in the fill-in-the-gaps test. As such, it was the most suitable test that reflected the students' learning preceding the fill-in-the-gaps test (see section 4.7.4). The students' scores were subdivided into four different groups: advanced, high-mid, low-mid, and low level and four participants from each group agreed to take part in the TA interviews. These four groups are also referred to as 'proficiency levels' throughout the thesis, however they do not refer to the participants' overall competence level in English, but rather are based on students' scores in the delayed receptive post-test.

For the TA interview, the researcher followed the recommendations by Bowles (2010), and used the task i.e. fill-in-the-gaps. The fill-in-the-gaps task was composed of 19 multiple choice questions and each question contained the target items. The researcher clearly informed the participants to use English or Arabic in the TA interview. During the interview, the researcher explained that it was important that the participants voice their thoughts and ideas while performing the task, as this would help the researcher gather information about their learning. The researcher also administered general instructions on how to think-aloud and provided warm-up

tasks. After the warm-up, the participants had a chance to ask questions. Two appropriate recording devices were used to ensure all details were audio-recorded (with participants' consent).

According to Bowles (2010), the rationale behind having a control group was that if scores on tests are statistically similar in the silent and think-aloud groups, it can be concluded that verbalization and the thinking aloud did not substantially work. However, the current research interviewed participants from three different groups including the control group. This was because the participants in the control group did not receive any CAT instruction, and thus their input and think-aloud reflections on completing the fill-in-the-gaps task was as important as those participants who formed the experimental groups and received CAT instruction. However, the pre-, post- and delayed tests were statistically measured between the three groups and thus they could stand as Bowles' (2010) recommendations.

3.11.3 Follow-Up Interviews

Upon completion of the delayed test, the researcher was eager to understand the students' perspective of their experience with CAT instruction and the utilised parallel corpus. When looking at the participants' test results, one of the questions that came to mind was whether it might be beneficial to ask students about their views on the teaching interventions that they had received. This reflection led to the formulation of a further research question, which emerged prior to collecting data from the TA interview. The additional research question was: *what are the views of the learners on the teaching they had received? Do these views illuminate the learners' strategy use and how well they did in the tests?*

After the TA interviews, the researcher conducted a follow-up interview with the treatment groups only. This took place immediately after students had completed the fill-in-the-gaps task. The interviews focused on students' views and perceptions of the intervention they received. A

total of seven students were interviewed, four from the CAT+Corpus group, and three from the CAT group. The interviews were very brief and only lasted approximately five minutes. The researcher conducted the interviews in Arabic and students replied in Arabic but used examples in English when referring to the target items. The interviews were audio-recorded and later transcribed and translated into English.

3.12 Ethics

Ethical approval was granted by the Institute of Education's Research Ethics Committee. Consent forms were given to the participants both in Arabic and English (see Appendix D), in which the purpose of serving as a requirement for a PhD research study is clearly stated. Informed consent was required because the research involved collecting data from human participants. Furthermore, because the current study (both the pilot and the main study) was conducted at a female university campus located in Saudi Arabia, it was important that participants fully understood the need for responses to be audio-recorded. Saudi Arabia in general is known for its conservative environment, and thus there is a tendency among students to object to audio recordings. Participants who accepted to take part in the think-aloud interviews, were interviewed individually and an audio recorder was used to record their answers.

Before conducting the study, the researcher provided the head of department with the information sheet and a consent form. These were then provided also to the participants. Both the information sheets and the consent forms were given to the participants in English and Arabic. The information sheet for the head of department as well as the one given to the participants included the following information: all details of the study, the type of tests and time allocated for the tasks, why they were selected for the study, what would happen if they participated, the risks and benefits

of taking part, what would happen to the data provided, while they were also reminded of their right to withdraw from the study at any point.

After the participants understood all the arrangements described in the information sheet, the consent form was given to them. The consent form required their signature and requested that they tick the following options: accessing their placement standardized scores, taking part in the study, taking part in an interview, and recording the interviews. In addition to that, both the information sheets and the consent forms had the name of the researcher as well as the names of the supervisors for any inquiries or complaints.

In terms of the lessons and interventions, the materials that were used were related to the course book. All three groups (the two experimental groups and the control group) could easily access their course books and find the lessons and the comprehension questions. Additionally, the study took place within the usual teaching schedule and usual site, while questions and assessments were related to the curriculum, interviews were related to the students' learning and assessments, identities were anonymised, and the confidentiality of the participants and the institute was to remain protected.

Within the experimental groups, students who did not wish to participate remained in class during the reading sessions and tasks, but they did not take any tests related to the study. Instead, they received comprehension questions related to their books, which were in paper format. They were not interviewed and their personal information and data were not used.

It is important to note that the researcher also paid close attention to the control group of the main study who just read the passages and answered comprehension questions on them. The group did not receive computer corpora tasks nor CAT activities. However, they did receive

vocabulary pre-,post, and delayed tests as well as comprehension questions and tasks related to their course book.

Furthermore, the researcher treated each participant respectfully, fairly, sensitively, and with dignity and freedom from prejudice. The researcher did not attempt to harm, embarrass, or shame the participants (see Appendix D for ethical approval). Furthermore, issues of power dynamics between the researcher and participants were also taken into consideration. The importance of establishing trustworthiness in qualitative research has been previously acknowledged (see Merriam & Tisdell, 2016). Trustworthiness refers to the authenticity and truthfulness of the participants' responses. In order to improve the trustworthiness of participants' responses, participants were reassured that taking part in the study would not affect their grades and that they had the right to opt out or withdraw at any time. Additionally, before commencing the data collection, the purpose of the current research study was fully explained and participants were informed of how their responses would help make a contribution to the research field of English Language Teaching. Seidman (2013) further emphasises the importance of engaging in reciprocity when conducting interviews to ensure that both the researcher and participants benefit from the study. As such, participants in this study had the opportunity to become aware of a different approach towards vocabulary learning (i.e., CAT instruction), and the Think-aloud interviews also provided an opportunity for participants to reflect on their learning process.

Chapter Four

Findings

4.1 Introduction

This chapter presents the results of the quantitative data collected from the main study, followed by the qualitative results, with the aim of addressing the key research questions guiding the investigation.

The study comprised a data collection period of four weeks for collecting the quantitative data and one week for the qualitative data. Both quantitative and the qualitative data were collected from three different classes from EFL learners within the English Language Institute. The procedure required the participants to complete different tasks, one to test productive knowledge of the test items and the other to test receptive knowledge.

4.2 Restating the Research Questions

1) What is the impact on short term learning of semantically non-transparent words and collocations (form and meaning recall) among high-intermediate learners of English of

(a) a contrastive analysis and translation (CAT) instruction by the teacher?

(b) a combined approach using both a parallel corpus and direct teacher CAT instruction?

(c) how do the previous two approaches compare with an approach using neither CAT nor a parallel corpus?

RQ2: *What strategies do learners report using to understand semantically non-transparent words and collocations in a think-aloud task?*

RQ3: *What are the views of the learners on the teaching they have received? Do these views illuminate the learners' strategy use and how well they did in the tests?*

4.3 Participants and Baseline

The data set was collected from 56 participants who were allocated into three groups: CAT, CAT+Corpus, and control group without intervention. There were 19 participants in the CAT group, 18 participants in the CAT+Corpus group, and 19 participants in the control group. The analysis mainly investigated the change from pre- to post-test in productive and receptive knowledge across the three groups (full information on scoring is presented in chapter three, section 3.5)

Before the intervention session for the three groups, the researcher randomly selected the participants to be placed into the three different groups in ELI level three (there are four levels in the ELI program). The participants were placed in this level based on their scores from the Cambridge placement test. However, there were some participants who missed the deadline for the entrance exam. These students were required to take all the previous levels starting from level one and go through Cambridge tests to pass to the next level. Conversely, students who do not pass the Cambridge test stay at their current level.

4.4 Data Analysis

The following will address the first research question: *what is the impact on short term learning of semantically non-transparent words and collocations (form and meaning recall) among high-intermediate learners of English of (a) contrastive analysis and translation (CAT) instruction by the teacher (b) a combined approach using both a parallel corpus and direct teacher*

instruction of CAT? (c) How did the previous two approaches compare with an approach using neither CAT nor a parallel corpus?

The data analyses for both productive and receptive tests consisted of first computing descriptive statistics in the form of indicators of central tendencies such as mean (*M*), also indicators of the data variability by the standard deviations (*SD*) and followed by inferential statistics, two-way mixed ANOVA.

A two-way mixed ANOVA is a type of repeated-measures ANOVA which tests the mean scores to show the difference among the groups that have been split on two independent variables (Larson-Hall, 2010). In two-way mixed ANOVA, these two independent variables are called two factors, and one factor is a within- participants factor (which includes time e.g., pre- and post-tests) and the other factor is a between-participants factor (includes independent groups such a gender: male/female) (Laerd Statistics, 2015; Field, 2018).

For the current study, a two-way mixed ANOVA was carried out with Time (pre, post intervention) as a within participants factor, and Group (CAT, CAT+Corpus, control) as a between participants factor in each (productive and receptive) analysis. The two-way mixed ANOVAs were carried out after checking the following assumptions, which underlie the use of this type of ANOVA:

1) There should be one dependent variable (at the continuous level), one between-participants factor that should involve at least two unrelated categories and one within-participants factor that should consist of at least two related categories (Larson-Hall, 2010);

2) The "dependent variable should be approximately normally distributed for each cell of the design" (Laerd Statistics, 2015, p. 7);
3) There should be the equal variances across samples of each group (Field, 2018).
4) There should be sphericity in which it requires not just that the variance of the different conditions that the repeated measures factor are equal, but also that the covariations between the different levels of the conditions are all equal. Therefore, in cases where there are only two levels of the repeated measures variable, as in the case with the current study, sphericity is not a concern (Field, 2018).

4.4.1 Exploring the Assumptions

To explore the first assumption, the dependent variable was the scores for the productive task for the first ANOVA and the scores for the receptive task for the second ANOVA. In order to explore the second assumption, the data were screened to establish whether they were approximately normally distributed (Laerd Statistics, 2015). To check normality in the productive data, Shapiro Wilks test was used (see Appendix B Table 2a), as this is the most appropriate test of normality for small samples (Field, 2018). The data showed non-normality among the three groups at pre-test ($p < .05$). Kurtosis and Skewness were also examined, confirming departures from normality: CAT group skew: $z = 6.79$; kurtosis: $z = 5.07$; CAT+Corpus group skew: $z = 3.26$; kurtosis: $z = 3.52$ and the control group skew: $z = 6.79$; kurtosis: $z = 5.07$. Similarly, the data suggested non-normality among the three groups at post-test ($p < .05$); and the results revealed that the values for Kurtosis and Skewness for the CAT group skew: $z = 0.85$; kurtosis: $z = -1.68$; the CAT+Corpus group skew: $z = 1.09$; kurtosis: $z = -0.86$, and the Control group skew: $z = 3.45$; kurtosis: $z = 3.64$. This suggests, as also is shown clearly in Figure 4.1 below, that there were floor

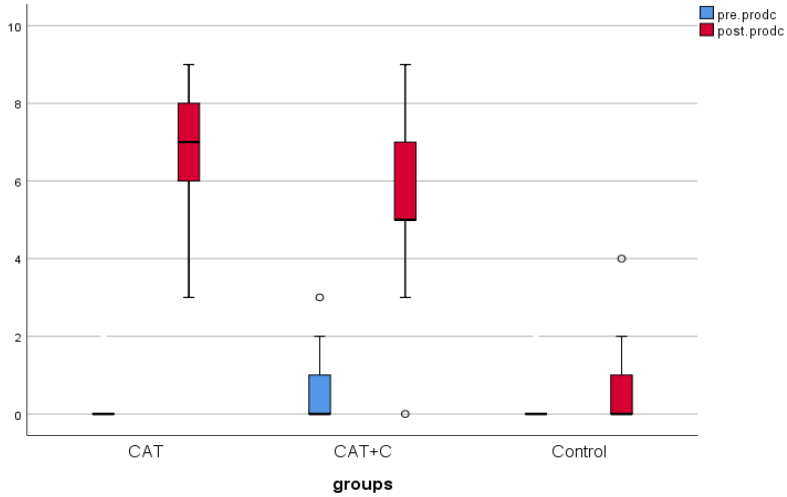
effects at pre-test for the productive task, and for the control group there was also some evidence of floor effects (positive skew) at post-test.

To check normality for the receptive test scores, the Shapiro Wilks test was used (see Appendix C Table 2a). Normality analysis revealed that there was significant non-normality, particularly in the post-test data across all groups, indicating the presence of ceiling effects, which can also be seen in Figure 4.2. The Kurtosis and Skewness at pre-test for the CAT group was found to be skew: z = -1.36 and kurtosis: z = -0.02, and for the CAT+Corpus group, skew: z = -1.29 and kurtosis: z = 0.30, and the control group skew: z = -1.11 and kurtosis: z = 0.59. However, the results at post-test showed the CAT group skew: z = -0.75, kurtosis: z = -1.74, the CAT+Corpus group skew: z = -0.31, kurtosis: z = -1.69, and the control group skew: z = -0.68, kurtosis: z = -1.69, which indicates there was a negative skew.

For the productive test scores, Figure 4.1 illustrates the box-plot generated for each group at each test time point. This shows that there were few cases of outliers identified by SPSS, however it was decided to keep these outliers especially since they are within expectable range and they were not data entry errors. Furthermore, excluding the outliers, which are not errors, could have potential implications on further data analysis (Laerd Statistics, 2015). Therefore, all cases were included in the analysis. For the receptive tests, Figure 4.2 illustrates the results for each group at each test time point. As can be seen from the boxplot, no outliers were identified.

Figure 4.1

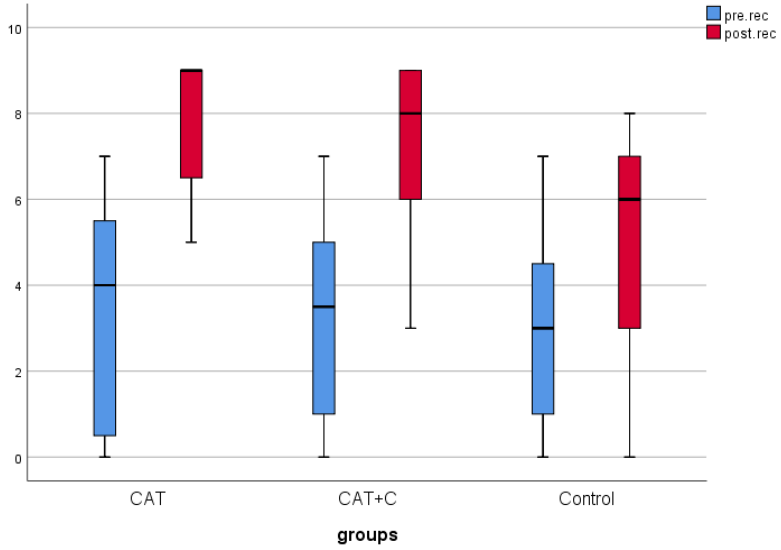
Boxplot of Productive Pre and Post-Test Scores for Intervention and Control Groups.



Note. The abbreviation of CAT+C is for CAT+Corpus, Pre. Pr is for pre-test productive, and Post. Pr for post-test productive.

Figure 4.2

Boxplot of Receptive Pre and Post-Test Scores for Intervention and Control Groups.



Note. The abbreviation of CAT+C is for CAT+Corpus, Pre. Rec. is for pre-test receptive, and Post. Rec. for is post-test receptive.

A further test was conducted to test for the third assumption, i.e., that the variance of all groups should be homogenous. The homogeneity in variances between the groups was checked using the Levene's test (see Appendix B Table 6 for Productive Test, see Appendix C Table 6 for Receptive Test). For the productive test, both pre- and post-test *p* values for the productive knowledge were greater than .05, indicating that the variances were not significantly different from each other, which means the assumption of homogeneity was met.

The Levene's test results for the receptive test data also showed that variances for the pre-test were homogenous, i.e., with *p* values greater than .05, but this was not the case for the post-test as the *p* values were less than .05 (see Appendix C Table 6).

With regard to the fourth assumption, in both the productive and receptive tests, sphericity was not a concern because there were only two levels of the repeated measures factor (Time), as stated above.

It was decided to analyse the data for both the productive and receptive tasks using ANOVA, in spite of the issues noted above, because of the advantages of a factorial design, where both main effects and their interaction could be tested within the same analysis, and because ANOVA is relatively robust to violations of assumptions. However, because of the violations, non-parametric tests were also carried out to validate findings, and are reported in the text below.

The next section will explore the descriptive statistics of the productive test followed by the two-way mixed ANOVA. Following that, the chapter will cover the descriptive statistics of the receptive test followed by the two-way mixed ANOVA.

4.5 Results from the Productive Test

4.5.1 Descriptive Statistics

The descriptive statistics for the productive task (see Table 4.1) suggest that there was a large improvement between pre- and post-productive test scores for the CAT and the CAT+Corpus group. The mean values for both intervention groups indicated vocabulary gain between the pre and post-test, while improvement within the control group was negligible. These data were then analysed using ANOVA as reported below.

Table 4.1

Descriptive Statistics for Productive Pre- and Post-Test Scores

Groups (N)		Pre-Productive Score	Post-Productive Score
CAT (19)	***M***	0.21	6.68
	SD	0.54	1.67
	Range	2.0	6.0
	Min	0.00	3.0
	Max	2.0	9.0
CAT+C (18)	***M***	0.50	5.61
	SD	0.86	2.25
	Range	3.0	9.0
	Min	0.00	0.00
	Max	2.0	9.0
Control (19)	***M***	0.21	0.63
	SD	0.54	1.12
	Range	2.0	4.0
	Min	0.00	0.00
	Max	2.0	4.0

Note. The abbreviation: CAT refers to the CAT only group; CAT+C describes the CAT+Corpus group.

4.5.2 Mixed ANOVA Results for the Productive Task

As shown in Table 4.2 and graphically illustrated in Figure 4.3, results of a Mixed ANOVA showed that for the productive scores there was a significant effect of time, with a large effect size (see Table 4.2). There was also a significant effect of group with a large effect size (see Table 4.2). The time by group interaction was significant with a large effect size (see Table 4.2). This aligns with Cohen's (1988) scale for partial eta squared, small ($\eta p^2 = 0.01$), medium ($\eta^2 = 0.06$), and large ($\eta p^2 = 0.14$).

Table 4.2

Results of the Time by Group ANOVA Carried out on Productive Scores

Source		*df*	*F*	*p*	Effect size (ηp^2)	Observed Power
Time	Sphericity Assumed	1	318.502	0.00	0.857	1.00
Group	Sphericity Assumed	2	50.682	0.00	0.657	1.00
Time * Group	Sphericity Assumed	2	67.84	0.00	0.719	1.00

Given the significant interaction, further one-way ANOVAs were then carried out to investigate the simple effect of group, first at pre-test, and then at post-test. The homogeneity in variances between groups was checked using Levene's test. As shown in Appendix B Table 6, it was found that the three groups' variances were homogeneous in terms of pre-test and post-test (though the results at pre-test were very close to significance ($p = .51$)). The results of the ANOVA showed that at pre-test there was no significant difference in the productive scores of the three groups ($F(2) = 1.189$, $p = .313$, $\eta p^2 = .04$). However, at post-test, there was a significant difference

between the three groups with large effect size ($F(2) = 65.852$, $p < .001$, $\eta p^2 = .704$) (see Appendix B Table 8).

An additional analysis (one way ANOVA) was conducted, comparing the CAT and the CAT +Corpus group only in order to ensure that there was sufficient power to detect a difference between the two intervention groups given the small sample size. The observed power was .46, indicating extremely low power.

Post-hoc tests were then carried out on post-test data (see Appendix B, Table 9) with adjusted *p*-values for multiple comparisons. These showed the mean difference of the post-test between CAT and CAT+Corpus was 1.07, ($p = .065$), which was not significant with a small effect size of $d = .54$; while the mean difference between the CAT and control groups was 6.05, which was significant ($p < .001$) with a very large effect size of $d = 4.26$. The CAT+Corpus scores were also higher than the control group, with a difference of 4.98, which was highly significant ($p < .001$) with a very large effect size of $d = 2.83$. This aligns with Plonsky and Oswald's (2014) scale for between-subject effects where $d = .40$ (small), $d = .70$ (medium), $d = 1.00$ (large).

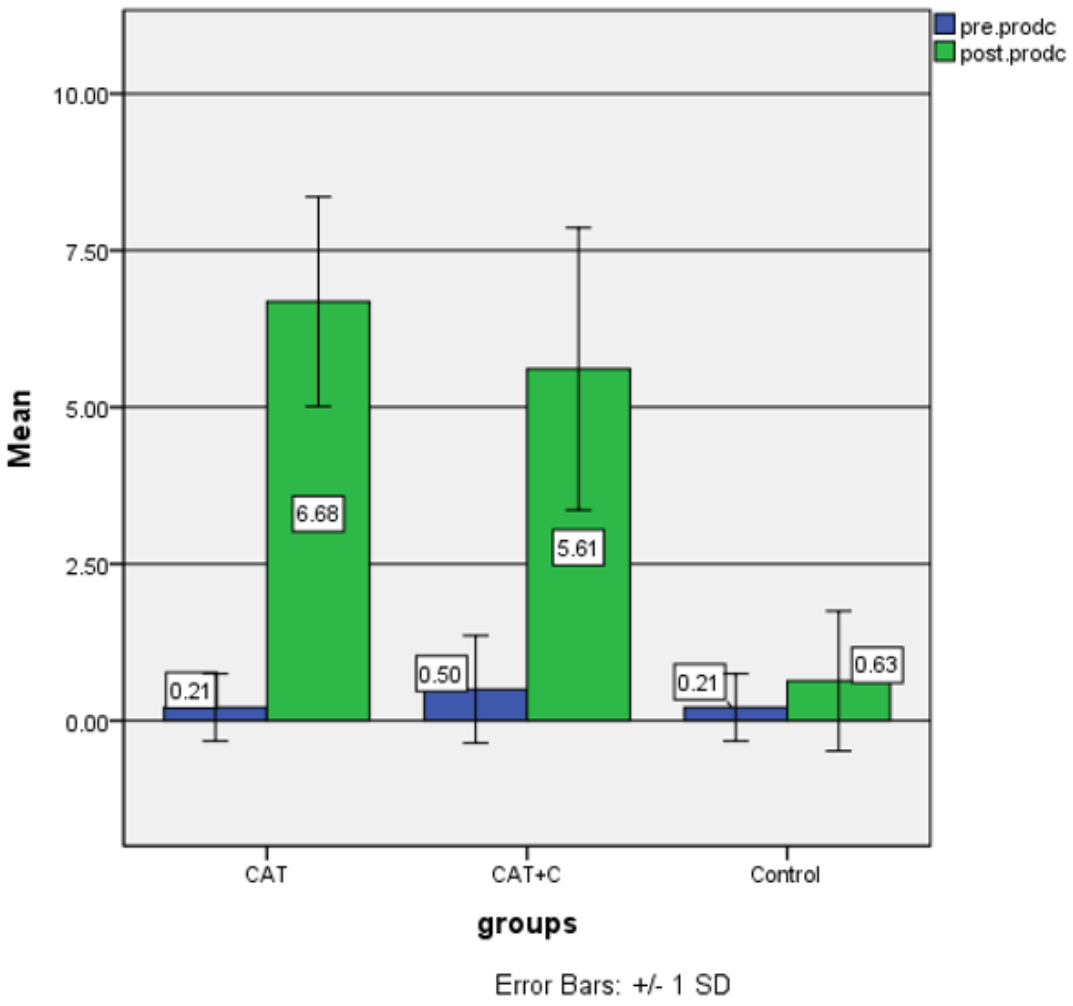
Pairwise tests were then carried out to see if there was simple effect of time in each of the three groups (see Appendix B Table 10). The CAT group's mean scores increased from 0.21 at pre-test to 6.68 at post-test with a very large effect size ($t(18) = -16.486$, $p < .001$, $d = 3.78$). For the CAT+Corpus group mean scores increased from 0.50 to 5.61 with large effect size ($t(17) = -10.01$, $p < .001$, $d = 2.36$). In contrast, the mean scores of the control group in the pre-test and post-test were 0.21 and 0.63, respectively, and did not increase significantly from pre- to post-test and had a very small effect size, ($t(18) = -1.909$, $p = .072$, $d = .45$). These effect sizes follow Plonsky and Oswald's (2014) scale for within-subjects where $d = .60$ is small, $d = 1.00$ is medium, $d = 1.40$ is large.

As outlined earlier, non-parametric tests were also run to validate the findings of the parametric tests. Kruskall Wallis tests were first conducted, looking at the Effect of Group, at a) Time 1 and b) Time 2. At Time 1, the three groups did not differ from each other significantly, $H(2) = 2.237$, $p = .327$, but did at Time 2, $H(2) = 34.806$, $p < .001$. Post-hoc Mann-Whitney U tests were then conducted, with a Bonferroni correction that reduced the alpha level to .017. These revealed significant differences between the scores of the CAT+Corpus group ($Md = 5$, $n = 18$) and the control group ($Md = 0$, $n = 19$), $U = 14$, $z = 4.81$, $p < .001$, $r = .6$ (large effect size), as well as between the scores of the CAT group ($Md = 7$, $n = 19$) and the control group ($Md = 0$, $n = 19$), $U = 2$, $z = 5.25$, $p < .001$, $r = .7$ (large effect size). There were however no significant differences between the CAT and the CAT+Corpus group, $U = 113.5$, $z = 1.77$, $p = .076$, $r = .2$ (small effect size). This aligns with Plonsky and Oswald's (2014) benchmark. These results of the three groups are the same as the results obtained in the parametric tests.

Next, the effect of Time was examined for each group separately, using a Wilcoxon Signed Rank Test. For the CAT group, there was a statistically significant increase from pre-test ($Md = 0$) to post-test ($Md = 7$), $z = 3.84$, $p < .001$, with a large effect size ($r = .6$). Similar findings emerged for the CAT+Corpus group where there was also a statistically significant increase from pre-test ($Md = 0$) to post-test ($Md = 5$), $z = 3.63$, $p < .001$, with a large effect size ($r = .6$). For the control group, however, the increase from pre-test ($Md = 0$) to post-test ($Md = 0$) was not statistically significant, $z = 1.87$, $p = .062$, with a small effect size ($r = .3$) (using Plonsky & Oswald's (2014) interpretation of effect size). These results of effect of Time were also as the same as the results obtained in the parametric test.

Figure 4.3

Productive Pre-Test and Post-Test Mean Scores Between Groups with Standard Deviation Error Bars.



4.6 Results from the Receptive Tests

4.6.1 Descriptive Statistics

The following section reports the analysis of receptive test scores for the three groups: CAT, CAT+Corpus, and control group starting with the descriptive statistics and followed by the results of the Mixed ANOVA.

Table 4.3 illustrates that there was an improvement between pre- and post-test. Mean scores for the receptive test indicated vocabulary gain for all three groups– that is, the CAT group, CAT+Corpus, and control group.

Table 4.3

Descriptive Statistics for Receptive Pre- and Post-test Scores

Groups (N)		Pre-Receptive Score	Post-Receptive Score
CAT (19)	***M***	3.32	7.84
	SD	2.60	1.50
	Range	7.00	4.00
	Min	0.00	5.00
	Max	7.00	9.00
CAT+C (18)	***M***	3.28	7.22
	SD	2.37	1.93
	Range	7.00	6.00
	Min	0.00	3.00
	Max	7.00	9.00
Control (19)	***M***	3.16	5.21
	SD	2.29	2.80
	Range	7.00	8.00
	Min	0.00	0.00
	Max	7.00	8.00

4.6.2 Mixed ANOVA Results for the Receptive Skills

As shown in Table 4.4 and graphically illustrated in Figure 4.4, findings indicate there was a significant effect of time, with a large effect size (see Table 4.4). There was nonsignificant effect of group (see Table 4.4). The time by group interaction was significant with a large effect size as

seen in Table 4.4. This aligns with Cohen's (1988) scale for partial eta squared, small ($\eta p^2 = 0.01$), medium ($\eta p^2 = 0.06$), and large ($\eta p^2 = 0.14$).

Table 4.4

Results of the Time by Group ANOVA Carried out on Receptive Scores.

Source		*df*	*F*	*p*	Effect size (ηp^2)	Observed Power
Time	Sphericity Assumed	1	161.04	<.001	.752	1.00
Group	Sphericity Assumed	2	2.405	> .05	.083	.46
Time * Group	Sphericity Assumed	2	7.41	<.001	.219	.93

Given the significant interaction, further one-way ANOVAs were then carried out to investigate the simple effect of group, first at pre-test, and then at post-test. The homogeneity in variances between groups was checked using Levene's test. It was found that the three groups' variances were homogeneous in terms of pre-test ($p = .819$) but not for the post-test ($p = .012$) (see Appendix C Table 6).

Therefore, one way ANOVA was used to compare the three groups in the pre-test while a Welch-F test was used for the post-test scores. The one-way ANOVA tests showed no significant difference in the receptive pre-test scores of the three groups ($F(2) = .022$, $p = .987$, $\eta p^2 = .001$) indicating a very small effect size. Regarding the receptive post-test, the Welch F-test results showed a significant difference between groups with a ($F(2, 33.43) = 6.39$, $p = .004$, $\eta p^2 = .227$) indicating a large effect size (Cohen, 1988) (see Appendix C Table 8 & 10).

Post-hoc tests with adjusted *p*-values for multiple comparisons showed the mean test scores increased from pre-test to post-test for all three groups. These tests showed the mean difference of the post-test between CAT and CAT+Corpus was 0.62, (p = .384), which is insignificant with a very small effect size (d = .36); while the mean difference between CAT and the control group was 2.63, which was significant with a large effect size (d = 1.17), p < .001. Furthermore, the difference between the CAT+Corpus scores and the control group was 2.01, p = .006, which was also significant with medium effect size (d = .72) (Plonsky & Oswald, 2014) (see Appendix C, Table 9).

In order to check there was sufficient power to detect a difference between the two intervention groups in view of the small sample size, a further set of analysis (one way ANOVA) was run, comparing the CAT and the CAT +Corpus group only. The observed power was .19, namely very low.

Pairwise tests were then carried out to see if there was simple effect of Time in each of the three groups (see Appendix C Table 10). The CAT group's mean scores increased from 3.32 at pre-test to 7.84 at post-test with a large effect size (t (18) = -10.096, p < .001, d = 2.31). For the CAT+Corpus group mean scores increased from 3.28 to 7.22 with a large effect size (t (17) = -8.697, p < .001, d = 2.05). The mean scores of the Control group in the pre-test and post-test were 3.16 and 5.21, respectively, and increased significantly from pre-test to post-test with a small effect size (t (18) = -3.896, p = .001, d = 0.89) (Plonsky & Oswald's, 2014).

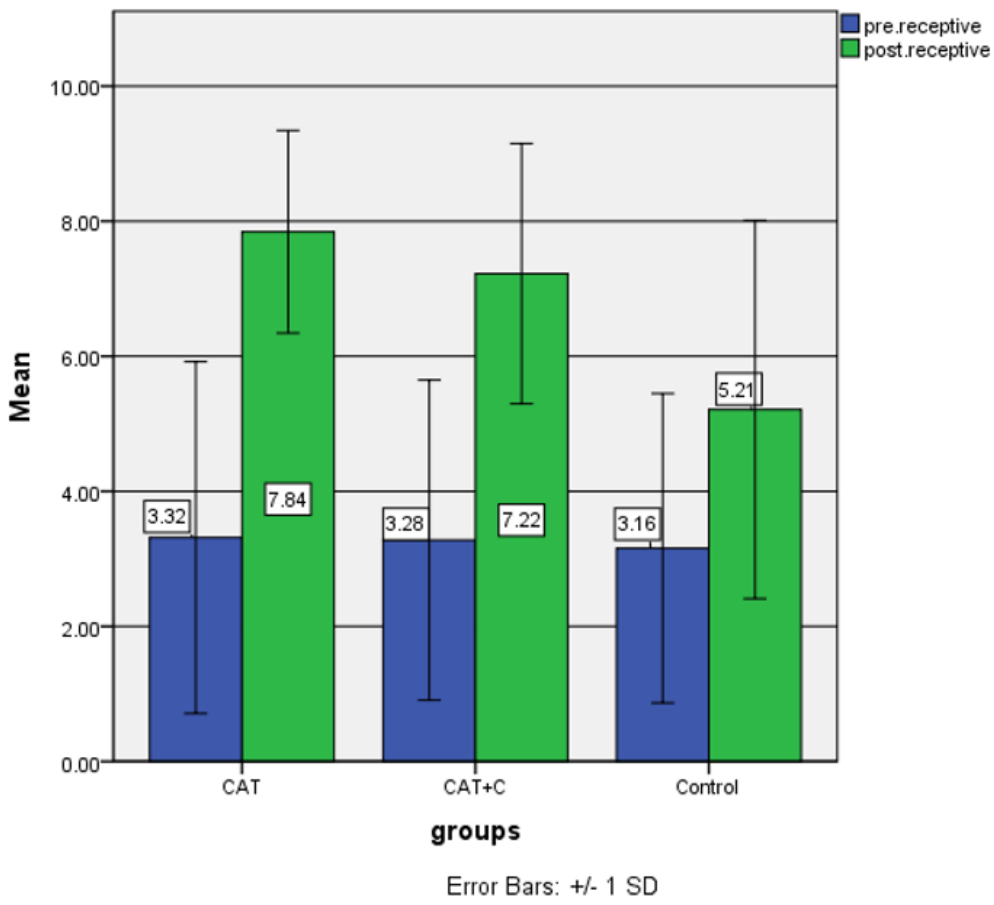
As outlined earlier, non-parametric tests were also run for the receptive knowledge to validate the findings of the parametric tests. Kruskall Wallis tests were first conducted, looking at the Effect of Group, at a) Time 1 and b) Time 2. At Time 1, the three groups did not differ from each other significantly, H(2) = 0.022, p = .989, but did at Time 2, H(2) = 12.406, p = .002

Post-hoc Mann-Whitney U tests were then conducted, with a Bonferroni correction that reduced the alpha level to .017. These revealed significant differences between the scores of the CAT+Corpus group (Md = 8, n =18) and the control group (Md = 6, n = 19), U = 90.5, z = 2.5 p = .013, r = .3 (medium effect size), as well as between the scores of the CAT group (Md = 9, n = 19) and the control group (Md = 6, n = 19), U = 69.5, z = 3.3, p = .001, r = .4 (medium effect size). There were however no significant differences between the CAT and the CAT+Corpus group, U = 136, z = 1.12, p = .264, r = .1 (small effect size) (Plonsky & Oswald, 2014). These results of the three groups are consistent with the results obtained in the parametric tests.

Next, the effect of Time was examined for each group separately, using a Wilcoxon Signed Rank Test. For the CAT group, there was a statistically significant increase from pre-test (Md = 4) to post-test (Md = 9), z = 3.83, p < .001, with a large effect size (r = .6). Similar findings emerged for the CAT+Corpus group where there was also a statistically significant increase from pre-test (Md = 3.5) to post-test (Md = 8), z = 3.74, p < .001, with a large effect size (r = .6). For the control group, there was also an increase from pre-test (Md = 3) to post-test (Md = 6) and this was statistically significant, z = 3.03, p = .002 with medium effect size (r = .5) (Plonsky & Oswald, 2014). These results of effect of Time were consistent with the results obtained in the parametric tests.

Figure 4.4

Receptive Pre-Test and Post-Test Mean Scores Between Groups with Standard Deviation Error Bars.



4.7 Procedure for Qualitative Analysis

One lengthy process that occurred during the data collection phase in the present study is the development of the code structure. It is important to note that where developing the code structure is concerned, substantial diversity can be observed, so much so that whether or not coding should be more inductive or deductive has been the subject of heated debate (Glaser, 1992; Heath & Cowley, 2004; Mohajan & Mohajan, 2023).

It should be mentioned that for the purposes of this study, both a bottom-up and top-down approach to coding was adopted. In light of Miles and Huberman's (1994) suggestions, the coding was conducted initially through a top-down approach while maintaining flexibility in adding, deleting, and reshuffling the codes as needed based on the data. For example, theories drawn from the literature review were considered, such as translation and L1 transfer, noticing hypothesis (Schmidt, 2010), incidental and intentional learning, and involvement load hypothesis (ILH) (Laufer & Hulstijn, 2001), although the researcher was mindful to remain flexible and open to refining the process based on the patterns that were emerging from the data. Therefore, as the analysis began with a framework in place to guide the coding and allow for easier navigation of the data, it can be argued that it was more of a deductive rather than an inductive approach.

It should also be emphasised that in relation to the process of coding, this study analysed all the think-aloud (TA) transcriptions by thoroughly dissecting and organising the raw data into manageable patterns. These patterns were then labelled into codes, and the codes were grouped into categories. In addition, within the process itself, and as new concepts would consistently emerge, they were compared and contrasted to the existing codes. At that point, they would then either be coded and incorporated into the existing categories, or classified as a new group. The following sections will undergo a more detailed and comprehensive discussion of the procedures involved in the qualitative analysis. The process of participant selection will be revised first, followed by an overview of the preparation, transcription, and analysis of the data.

4.7.1 Selecting Participants

For the TA interview, 12 participants were selected; however, one participant withdrew. These participants were selected according to their receptive delayed post-test score. This is because the receptive delayed post-test was the last test taken by the students before the TA, and it had variable scores that ranged from low to advanced scores. The range of the scores helped the me group the students into advanced, high-mid, low-mid, and low-proficiency levels. I compiled the scores of each group and compared them. Then the scores were sequentially selected and categorised according to the highest scores, the lowest scores, and the average scores. Then the average scores for each group were also divided into high-mid and low-mid scores. Accordingly, I started to categorise students into different proficiency levels. The first selected group was the advanced (score of 16-19), the second group was the high-mid (score of 13-15), the third group was the low-mid (score of 10 -12), and the last group was the low (score of 9 and below). Importantly, however, this categorization was not clear cut especially as some participants could not attend the interview. Accordingly, I attempted to find the closest scores to the initial categorization. Also, it is worth mentioning that one of the participants in the CAT only group withdrew her interview after being interviewed for personal reasons.

It is vital to note that although the productive delayed post-test was also the last test along with the receptive delayed post-test, the productive delayed post-test could not act as a measurement through which to select participants for two reasons. Firstly, most of the students received very low scores so it was difficult to differentiate between proficiency levels, while secondly; the test required the participants to read in Arabic and translate in English, which is different than the fill-in-the-gaps task that required them to read in English. After studying the receptive delayed post-test scores, I selected students who completed the intervention and

experiment, which included all tests (pre-test, post-test, second post-test, and delayed post-test in both productive and receptive knowledge). I then checked the students' confirmation of consent as there were many students who did not consent to take part in interviews.

During week five, I and the participants conducted the interview in an empty classroom and sat around a table. I explained the guidelines of the TA interview and did a rehearsal that involved mathematics. The students were then given the opportunity to practise simultaneously talking loudly and solving mathematical questions. After that, I gave the participants the fill-in-the-gaps task and asked them to do the same protocol of talking aloud while solving the questions. I also clarified that it was up to the students to use the language they preferred within the TA interview. I conducted a follow-up interview, which was a second stage of the same interview, asking the two treatment groups about their experience using CAT or CAT+Corpus. Thus, the views were taken from the two treatment groups; and the control group were not asked about their experience like the other groups since they did not have any intervention.

4.7.2 Data Transcription and Preparation

It is of vital importance to mention that all of the participants used both L1 and L2 in the TA interviews. I fully translated and transcribed the interviews in English; however, in order to ensure the accuracy of the transcription process, all of the transcripts were carefully reviewed and checked for mistakes by two researchers: another researcher in the same field and myself. This quality check provided assurance that the transcripts were consistent (the original transcript and the translated version) with the interview audio recordings to the highest degree possible. The transcriptions were then reviewed and analysed in great detail, multiple times, in order to extract the most important themes and common patterns. The next step was to perform the data coding

process, which involved an intercoder for reliability and validity purposes. The breakdown of the data coding process is presented in the following section.

4.7.2.1 First-Stage of Coding. The first stage of coding aimed to produce the initial list of codes extracted from the interview data. One participant was chosen randomly, and the process started with her transcript. I read through the whole transcript carefully. The data were arranged so that the content was presented on a line-by-line basis, to make the coding process more straightforward and less complicated to navigate. More specifically, the transcript was divided based on the phrases and sentences uttered by the participant. For example, whenever there was a pause of three seconds or more, a new sentence, or new idea, a new line was started. Next, I categorised each phrase and sentence by allocating either a code or codes to a line or group of lines, and the codes were written clearly in the margins. The coding structure was developed through careful analysis of the transcripts. In the TA interviews, participants described their thinking processes when deciding which target item to choose. For example, one student used the words *I think* before selecting her answer. This was initially identified as an example of the strategy of guessing. When analysing other transcripts, I noticed the use of the words *I predict* and *maybe*. I then coded these as evidence in support of the strategy of guessing. Another example of how the coding was developed can be seen in the strategy of translation. When analysing the transcripts, I noticed that many participants translated the sentences and the target items into Arabic. This was coded as an example in support of the strategy of translation.

Following that, I recorded the participants' verbal reaction to each question. For instance, if a participant focused on one part of the question, such as a collocation or a phrase, she would sometimes use one or more strategies to address it. She would then focus on another part of the

question and either use the same strategy or another strategy to work out the answer. In other words, participants would often compartmentalise the questions or divide them up to make them more manageable, and thus the number of strategies used for each question varied between participants and was a reflection of this division. For example, the phrase *I think* could be seen twice in one question, although stated in two different instances, and thus recorded as two occurrences. By following this thorough method, I was able to deeply analyse the transcript and capture the exact numbers of strategies used.

For each transcript, the process outlined above was followed for all the codes. The only two codes that were treated and counted differently were *monitoring against sense* and *translation* (check definition of codes in section 4.7.3.2). This is because both of these strategies depended on the question itself. For example, in the strategy *monitoring against sense,* participants would not break the question into sections, but instead would try and use the three multiple choice options against the provided sentence to see if it made sense. Also, for *translation*, there was great reliance on this strategy as it was used in all of the questions for most of the participants. For example, L1 would be used between words, for thinking, for transferring from one question to the other, and for clarification. Sometimes participants would even use the word "okay," which is used regularly as a common word in Arabic. Therefore, to have a more reliable count, I linked the question numbers to the strategies of *translation* and *monitoring against sense*. This meant that, for example, if a student used translation while completing question 1, then that would be counted as one incident of using translation, regardless of how many Arabic words were used. The same coding approach was used to count incidents of *monitoring against sense*. Once the first transcript was fully coded, this was used to construct an initial codebook to help by serving as a guide to facilitate the coding of the rest of the (TA) transcripts. Subsequently, after then reviewing four

more transcripts from other participants, some new strategies could be observed repeatedly in different places (such strategies include: stating the knowledge of a collocation, or recalling a word). Strategies were repeated either: within the same transcript, in the transcripts of the same group, and across the three different groups.

4.7.2.2 Second-Stage of Coding: Revising the Codebook. The main aim of the second stage in coding the transcribed data was to revise and finalise the initial codebook produced in stage one. To do so, I recruited a subject matter expert from a similar research discipline who is a native Arabic speaker who also speaks fluent English to code the same random samples of the five transcripts. To ensure satisfactory reliability and of course validity, the aims of the study as well as research questions were made clear to her, so that I would eventually be able to proceed with coding the remaining data on my own. It is important to mention that both the reviewer and I identified the same coding categories; however, we disagreed on some of the labels given to the codes. At this stage, the reviewer and I decided to postpone finalising the names of the codes and agreed that these would become clearer as the coding process continued.

4.7.2.3 Third-Stage of Revising the Codebook. Once the new version of the codebook and coding method were reviewed, I then coded the remaining six transcripts using the same coding approach, and thus a clear pattern of identified codes was gradually emerging (for example, some patterns of codes were not used in the control group whereas they were used in the treatment groups). These six transcripts along with the codes were sent to the same reviewer. There was an agreement of 94.3% in relation to the remaining coding. Thus, in order for a 100% agreement rate to be reached, I then held a discussion with the reviewer to clarify coding issues and solve all disagreements.

4.7.2.4 Final-Stage of Revising the Codebook. At this stage, the established and finalised codes were then reviewed overall by a second reviewer who was a native Arabic-speaking professional from another discipline with an excellent command of the English language. To ensure reliability and validity, the main aims of the study and research questions were again made clear to her. The reviewer understood the coding scheme and acknowledged that. Afterwards, I counted the codes for each participant, then for each of the treatment and control groups, and lastly for each English proficiency group (see Table 4.5). By quantifying the data in this way, it gave me as a researcher a better numerical overview of how and when participants were using different strategies. The following section will address the research questions specifically and discuss the findings in more depth.

4.7.3 Reporting Strategies

The second research question was the following: *what strategies do learners report using to understand semantically non-transparent words and collocations in a think-aloud task?*. This question was addressed by investigating the patterns of strategy use emerging from the data using both quantitative and qualitative analysis: However, first, it is important to start by defining the strategies of language learning and use.

4.7.3.1 Strategies for Language Learning and Use. According to Cohen (2011), there are specific strategies related to learning and using a second language. He explains second language *learning* strategies as "…conscious or semi-conscious thoughts and behaviours" whereas second language *use* strategies "are strategies for using the material that has already been learned"

(p. 12). This section will outline and explain the strategies used by participants in this study within the think-aloud (TA) interview when answering the fill-in-the-gaps-task. They included:

1. *Guessing*: participants answer the fill-in-the-gaps task by guessing and using utterances such as "I think," "I feel," "I guess," "I predict," and "maybe".
2. *Monitoring against Sense:* participants check the accuracy of their interpretation and try to make sense of the text within the fill-in-the-gaps task.
3. *Use of L1 and Translation:* participants verbally used both languages, Arabic and English, when answering the fill-in-the-gaps task.
4. *Elimination:* participants attempt to solve the questions in the fill-in-the-gaps task correctly by excluding and omitting the words or collocations from the choices provided. For example, a participant would report: "I will exclude on the other hand and the outcome".
5. *Elaboration*: a) participants identify a word or a collocation and assess their comprehension by reporting that they know the meaning of the word or collocation. For example: the participants would state: "I know the meaning of outcome*"*.

   b) participants identify a word or a collocation and assess their comprehension by reporting they do NOT know the meaning of the word or collocation. For example, the participants would state: "I do not know the meaning of henceforth*"*.
6. *Self-Correction:* participants verbally re-evaluate their answers.
7. *Recall:* incidents in which participants remember learning the target words and

collocations in the intervention.

Among the above-mentioned strategies, the ones that were used the most were *use of L1* and *translation* followed by *guessing*. *Recall* was the strategy that was used the least. Table 4.5 below presents all incidents counted for each strategy for each participant.

Table 4.5
Instances of Strategies Used Among all Students.

Strategies	Student CAT+C Adv	High-mid	Low-mid	Low	Total	Student CAT Adv	High-mid	Low-mid (withdrew)	Low	Total	Control Group Adv	High-mid	Low-mid	Low	Total
Guessing	16	9	3	10	38	4	3	0	9	16	0	6	14	10	30
Self-correction	2	6	0	3	11	3	0	0	0	3	0	1	0	0	1
Elimination	0	11	4	7	22	5	0	0	2	7	17	0	5	0	22
Elaboration on knowing meaning	1	2	0	3	6	6	0	0	2	8	0	0	0	0	14
Elaboration NOT knowing meaning	1	15	1	2	19	4	1	0	5	10	0	1	0	1	2
Monitor against sense	11	20*	15	15	61	11	4	0	3	18	19	7	6	7	39
Recall	1	1	0	0	2	1	1	0	2	4	0	0	0	0	0
Use of L1/ Translation	19	19	19	19	76	19	18	0	19	56	19	19	19	13	70

Note. *This one student had 20 instances because she attempted to answer one question using the strategy of monitoring against sense, but then skipped it and returned to it after answering the remaining questions.

4.7.3.2 Strategies Used in the TA Interviews and the Fill-in-the-Gaps Task.

4.7.3.2.1 L1 Use and Translation. All the students used L1 for thinking and solving during TA and the fill-in-the-gaps task. In fact, some of them wrote the Arabic translations of some English words down on paper. Others would say half of the sentence in English and complete it in Arabic. Some even would insert an Arabic article for an English word. It was obvious they felt comfortable using their first language. This is demonstrated clearly in Table 4.5. There were only seven out of 202 strategy incidents where translation was not used. Thus, the norm was the reliance on L1 in solving the fill-in-the-gaps task. The following is an excerpt from a fill-in-the-gaps task completed by a low-mid student in the CAT+Corpus group showing the insertion of an L1 article and using L1 to solve the question:

Fill-in-the-gap task question 3: There is not much we can do, _____they've already made their decision. a) in the long run b) the outcome c) seeing that

CAT+Corpus Low-mid Student:

There is not much we can do in the long run they already made their decision the outcome. They are already seeing that.
In the long run المدى البعيد ماتجي مع الجملة
The outcome النتيجة ما تجي مع الجملة لانهم
Already ال اختاروا decision
Seeing that هي الاصح المدى البعيد

English Translation:

There is not much we can do in the long run they already made their decision the outcome. They are already seeing that.
In the long run in the long run, does not work with the sentence
The outcome the outcome, does not work in the sentence because they
Already chose the decision
Seeing that is the most correct

The student initially answered by trying out all three multiple choice options in English and putting them in the gap. Then she translated two of the options. At the same time, she used L1 and used another strategy - monitoring against sense (this will be discussed further below). When she came to the last multiple choice option, she used reasoning in L1 and used the word لأنهم which means *because they*. She also replaced the English article *the* with the Arabic article ال. Interestingly, this Arabic article was inserted with the English noun *decision*. She then inserted the translated word اختاروا/meaning *they chose* which was not in the question. She then concluded by stating in L1 that the most accurate answer was *seeing that*. As this excerpt illustrates, translation was used for the purpose of: 1) solving the answers by translating the collocations; 2) substituting articles from L2 to L1; 3) using translation alongside another strategy; and 4) applying reason in her thinking when addressing the question and thus, reaching a conclusion.

4.7.3.2.2 Guessing. This strategy was one of the most frequently used in the TA interviews. Words such as "I predict," "I think," and "maybe" would be used by the students from the advanced to the low-proficiency level. There was one student from the low-proficiency of the control group who actually stated clearly that her answers were "all predictions and assumptions". As illustrated in table 4.5, CAT+Corpus students used the guessing strategy in many instances; they used it in 38 occurrences and more than the other two groups. At times, guessing was used correctly and other times it was inaccurate. The following is an excerpt from a CAT+Corpus advanced student; the only one among all participants to answer all the fill-in-the-gaps questions correctly. She used the guessing strategy the most out of the whole sample of participants (16 times), and showed her ability to use it productively to answer all the questions in the task correctly:

Fill-in-the-gap task question 1:
The costs, _____ the risks, of moving the satellite to and from space would have been difficult.
a) not to mention
b) owing to
c) thereby

Advanced CAT+Corpus student's words:

السعر the costs
the risks اااه اه
moving في صعوبة شويه في الكلمات
satellite aah to and from space would have been difficult
طيب ان الأسعار و risks يعني مخاطر لي آه الحركه حركه الستلايت من الفضاء يمكن ان تكون صعبه
او يمكن ان تكون فيها عده مخاطر، therebe (mispronounced for the thereby) ولا whereas ولا not to mention
therebe (mispronounced) انه آه I think اوكي
طيب خلينا نقول not to mention يعني التكاليف غير المخاطر لانتقال الستلايت من الفضاء

English Translation:

The costs,
risks, aaaah moving, there are difficulties in words, satellite
aah to and from space would have been difficult
Okay the costs and the risks for me, moving moving the satellite from space could be difficulties
Or a few risks. Not to mention or whereas or thereby
Okay I think thereby
Let us say not to mention, meaning the costs not to mention the risks to move the satellite from space

Here the CAT+Corpus advanced student found the question to be difficult. She tried to figure out which answer was the most accurate. She then used the strategy of guessing and stated "I think thereby". But interestingly, she did not use *thereby*. She answered correctly but after guessing the wrong answer. This could suggest that although she used this strategy indirectly, it led her to answer correctly. This could also suggest that learners of different proficiency levels use the strategy differently, and that lower-proficiency students cannot use it as well. Table 4.5 shows the number of times students of each proficiency level used guessing. While lower-proficiency students relied on the guessing strategy more than students of higher-proficiency levels, those of a higher-proficiency have the potential to guess the correct answer more frequently.

4.7.3.2.3 Monitoring Against Sense. I found that this strategy was used frequently by all the students across the three groups. I also found that this strategy involved the students trying to answer the main question by comparing one of the multiple-choice options with the meaning of the question. The count of this strategy was different from the other strategies (except for use of L1 and translation). Using monitoring against sense depended mainly on selecting one of the multiple-choice options and checking if it made more sense than the other answers provided. In other words, the maximum instances of monitoring against sense should be 19 instances since the task includes 19 questions only. This tells us that when using this strategy, the participants looked at the entire question and mainly relied on the sentence provided as well as the multiple-choice options. It is different from any other strategy (except for use of L1 and translation) that relies on the word or collocation itself.

The following example is from a low-proficiency level student in the control group. The excerpt confirms that multiple strategies can be used within one question. The strategies that were used here were the strategies that were used the most among the three groups: monitoring against sense, use of L1 and translation, and guessing:

Fill-in-the-gaps task question 7: I promise not to lie to you _____ a) in so far b) henceforth c) much less

Control group low-proficiency level Student:

I promise not to lie to you
In so far as
Much less
Hencyforth (mispronunciation for henceforth)ايش يعني Hencyforth (mispronunciation)?
I promise not to lie to you in so far as
انا ماكذب عليك على المدى البعيد
اقل من
اقرب حاجة احس In so far as ممكن

English Translation:

> I promise not to lie to you
> In so far as
> Much less
> Henceworth what does it mean henceforth?
> I promise not to lie to you
> I will not lie to you in the long run
> Less than
> Maybe in so far as, I feel it is the closest thing

This low-proficiency level student from the control group used three different strategies to try to solve the question, despite answering incorrectly in the end by choosing the answer *in so far as*. She tried monitoring against sense, translation, and guessing; but there was an absence of the other strategies. Her reading was good as she read the full question followed by the three multiple-choice options, but she hesitated a little with the word *henceforth*. She was trying to pronounce *henceforth* and know the meaning. She even stated in Arabic "what does it mean?*"*. When she came to the phrase *much less* she translated it incorrectly as "less than". In the end, she chose what she felt was correct: *in so far as.* Although she relied on the guessing strategy, her answer was incorrect. This could be because she relied on the guessing strategy alone when trying to work out the meaning of the target item. This can suggest that lower-proficiency students use strategies less effectively than higher-proficiency level students. Also, in comparison to the other groups, the control students were relying on fewer strategies in their attempts to address the questions (see Table 4.5).

4.7.3.2.4 Elimination. As stated earlier, the strategy of elimination was to omit and exclude one or more of the multiple-choice options that seemed irrelevant. This strategy was not used as much as the previous three strategies: translation, guessing, and monitoring against sense. I found that elimination occurred 22 times across the CAT+Corpus and the control group. In contrast, elimination occurred only seven times in the CAT group, which is three times less than the first

two groups. This big difference was mostly attributed to two participants who used this strategy of elimination the most: the advanced student in the control group followed by the high-mid proficiency level student in the CAT+Corpus group. In fact, the advanced student in the control group only used this strategy of elimination with translation in the fill-in-the-gaps task. She did not use any other strategy, which could suggest she preferred this strategy in general, or that she was not exposed to the intervention which could have affected her use of strategies. In contrast, the high-mid proficiency level student in the CAT+Corpus group used this strategy many times along with all the other strategies to answer the fill-in-the-gaps task (see Table 4.5). The following shows the strategy of elimination used by the advanced student in the control group:

Fill-in-the-gaps task question 18: It all _____ to money in the end. a) the bottom line b) let alone c) boils down

Advanced-proficiency level student in the control group:

It all boils down to money in the end.
The bottom line لا
let alone نستبعدها كمان
It boils down to money in the end.تصير

English Translation:

It all boils down to money in the end.
The bottom line , no
let alone, also exclude it
It becomes, it boils down to money in the end.

As we can see from the transcript, the advanced student read the question and reviewed the multiple-choice options thoroughly in English. She only used L1 to eliminate two of the multiple -choice options and to confirm that the answer was *it boils down*, which was correct. It is worth

noting that before answering this question, the student sighed and said out loud “here are new words and terminologies”. She sounded tired after solving the previous questions, which suggests that she was putting a lot of effort into answering the question correctly. This could also point to the fact that for this particular student, using elimination and L1 were sufficient enough to solve the fill-in-the-gaps task, especially as these were the only two strategies she used.

4.7.3.2.5 Elaboration on Own Knowledge. When using this strategy, participants identified words and collocations and assessed their own knowledge of them by stating phrases like: “I know the meaning of this word” or “if I only knew the meaning of this particular word”. There was a total of 14 occurrences in which participants across the CAT+Corpus and the CAT groups elaborated on knowing the definitions of the words and collocations. I found that this strategy was scarce and was not reflected in the control group at all. This could suggest that participants in the control group were not confident in their answers, whereas the other treatment groups felt more assured since the193tilized193zediarised with the target words and collocations through the interventions.

Furthermore, there was a total of 31 instances where all three groups elaborated on *not* knowing the definitions of the words and collocations. Interestingly, there were 29 instances in which participants in the CAT+Corpus and the CAT 193tilizedtilised this strategy in comparison to the control group that had only two instances. This could also indicate the lack of confidence that those in the control group felt regarding the target words and collocations.

The following excerpt is from the low-proficiency student in the CAT +Corpus group expressing that she knows the meanings of two target words and does not know the meaning of the third target word:

Fill-in-the-gaps task question 1:
The costs, _____ the risks, of moving the satellite to and from space would have been difficult.
a) not to mention
b) owing to
c) thereby

Low-proficiency level student in the CAT +Corpus group:

انا مااعرف بالضبط هذه ايش يعني، اعرف هذه واعرف هذه فأتوقع انه
Thereby

English Translation:

I do not know exactly what does this mean, but I know this one and this one
So I predict it is thereby.

This example shows that the student had difficulty reading this question and reading the three multiple-choice options. She then reported that she did not know exactly what *thereby* meant but knew the other two multiple-choice options, although she did not provide a definition or translation. She also answered incorrectly, and relied on L1 in elaborating. This suggests that lower-proficiency level students can often identify words and state that they know them, yet still answer incorrectly. This example also reveals that lower-proficiency students can be familiar with how a word or collocation appears, but not know its meaning or definition. Moreover, it is clear to see from this excerpt that L1 is often used in parallel to the strategy of elaboration.

The below excerpt is from the advanced-proficiency level student in the CAT+ Corpus group:

4) Fill-in-the-gaps question 4:
She forgot to tie her shoes, _____ tripping and falling down the stairs.
a) thereby
b) in the long run
c) seeing that

Advanced student in the CAT+Corpus group:

She forgot to tie her shoes aaaaah therebe (mispronunciation),
in the long run, whereis (mispronunciation)
therebe (mispronunciation) ahh tipping falling down the stairs.
In the long run يعني في المدى البعيد.
Whereis (mispronunciation) معناها الي اعرفه but
عشان كذا بما انها ما قدرت تربط حقتها حقها خيوط الشوز فهي بالتالي وقعت على الدرج

English Translation:

She forgot to tie her shoes aaaaah thereby,
in the long run, whereas
thereby ahh tipping falling down the stairs.
In the long run
Whereas what I know it means but.
Because she was not able to tie her shoe laces thereby she fell on the stairs.

As demonstrated with the lower-proficiency student above, she only identified and reported that she knows the meaning of the word, but here the advanced student provided an English synonym for *whereas,* and relied on L1 to elaborate that she does know the meaning of it. She then translated the complete question with the correct answer *thereby*. This example suggests that higher-proficiency level students can elaborate on knowledge better. Indeed, the more competent student appeared to be more confident in answering the question and was able to draw on both her vocabulary knowledge of L1 and L2.

4.7.3.2.6 Self-Correction. There was a total of 15 self-correction occurrences. The CAT+Corpus group had a total of 11 occurrences; the CAT had three occurrences; and the control group had only one occurrence. The fact the CAT+Corpus group used self-correction the most may have been because of exposure to the intervention. The following excerpt is from the advanced student in the CAT group. She corrected herself and actually stated “I’m wrong” after realising she selected an incorrect definition for one of the target items.

Fill-in-the-gaps task question 18:
It all _____ to money in the end.

a) the bottom line
b) let alone
c) boils down

Advanced-student in the CAT group:

It all to money in the end
It all boils down لانه اقرب شيء
هو كل معناه يتلخص
it boils down او bottom line انا غلطانه
معناه يتلخص
هذه علاطول حأستبعدها بس حأبقى محتارة بين اثنين
a and c
احس ارتب وتجي كيدة It boils down احس بس في تركيبة الجملة

English Translation:

It all to money in the end
It all boils down is the closest thing.
Everything means sums up to.
I'm wrong, bottom line or it boils down mean sums up
This will be immediately eliminated.
But two will remain, a and c.
But in constructing the sentence I feel it boils down.
I feel it's more organised and makes sense

Here, the advanced student in the CAT group used different strategies to solve this question. She used translation (see excerpt); elimination by saying this will be "immediately eliminated"; guessing by stating "I feel"; monitoring against sense when commenting that an item "makes sense"; and also self-correcting when stating *"*I'm wrong, bottom line or it boils down means sums up". She initially glanced at the three multiple choice options where the first choice – *the bottom line* - and the third choice - *boils down* - had a similar meaning, but didn't really give much thought to the second multiple choice option *let alone*. Conversely, it could be that this participant could be that she actually thought that the three multiple choice options had the same meaning. In any case, she corrected it and answered with the right collocation. This could suggest

that although any student can be confused or potentially make errors, the higher-proficiency students have the ability to self-correct their answers.

4.7.3.2.7 Recall. The strategy of recalling the target words was used the least with a total of six occurrences across the three groups. Recall occurred twice within the CAT+Corpus, four times within the CAT, and there were no occurrences of recall in the control group. The following excerpt is from the advanced-proficiency level student in the CAT group remembering the target word but not the meaning:

Fill-in-the-gaps task question 7: I promise not to lie to you _____ a) in so far b) henceforth c) much less

Advanced-proficiency level CAT student:

I promise not to lie to you مني عارفة معانها لكن اذا حسب انا ماتذكرانه
In so far as, I promise not to lie to you in so far as
تركب على الجملة ممكن تكون صح
مع انني مني عارفة ولا متأكدة من معانبها نرجع لها بعدين كمان

English Translation:

I promise not to to lie to you. I do not know the meaning but as I recall that.…
In so far as, I promise not to lie to you in so far as
It makes sense with the sentence. Maybe it's correct
Although I do not know and I am not sure of the meaning.
I'll get back to it later.

The advanced-proficiency level student in the CAT group used several strategies to solve this question. She used translation, monitoring against sense, elaborated on not knowing the meaning, and recall. She translated from L2 to L1 while trying to recall the meaning of *in so far*

as to help her answer the question. However, she could not rely on translation totally as she started to use monitoring against sense and the guessing strategy. In the end, she elaborated that she did not know the meaning and even skipped the question. After completing the rest of the questions, she returned to answer this question and chose the answer *in so far as*, although it was incorrect. It was possible that during translation when she stated "as I recall," she could not actually recall the meaning of *in so far as* in the L1, and accordingly answered incorrectly. The following excerpt is taken from a low-proficiency level student in the CAT group. The student started to make sense of the first part of the question and then tried to understand what was required.

Fill-in-the-gaps question 3

There is not much we can do, _____they've already made their decision.
a) in the long run
b) the outcome
c) seeing that

Low-proficiency level student in CAT group:

There is not much we can do okay مو مره نقدر نسويها
In the long run
ايوه على المدى البعيد او شيء زي كيده يدل على طويل الامد فترة طويله
Outcome
معناها بحاول اتذكرها... اتذكر اني اخذتها بس ماافتكر معناها
احس الجواب الصح
Outcome

English Translation:

There is not much we can do okay, we cannot really do it.
In the long run.
Yes in the long run or something like this indicates long term, long time.
Outcome
The meaning, I'm trying to remember it…I remember I took it but I do not remember the meaning
I feel the correct answer is outcome

This low-proficiency level student initially translated the first multiple-choice option *in the long run* and explained it in L1. Then she looked at the second multiple-choice option *the outcome*. Here she tried to remember the meaning of the word and even stated in Arabic that she was trying to remember it. She remembered that she had learnt it, but she did not recall the meaning. She then used the strategy of guessing and stated "I feel the correct answer is outcome". The above discussion raises a very interesting point as it presents two students from the same group (CAT) with different proficiency levels (advanced- and low- proficiency), and although they both used the same strategy neither one recalled the correct meaning. This can suggest that the intervention that was given to this group did not support or enhance their recall strategy.

The above section explored the strategies that were used in general among the CAT+Corpus, CAT, and the control groups. However, to understand these findings from an alternative angle, it is also important to look more closely at these strategy uses in relation to each proficiency level, as this may point to new insights and unique explanations. These patterns and associations will be discussed below.

4.7.3.3 Strategies Used Across Each Proficiency Level. The following sub-sections will elaborate on strategies used among each proficiency level (advanced, high-mid, low-mid, and low).

4.7.3.3.1 Advanced Level Students' Use of Strategies.

Table 4.6

Strategies Used Among All Advanced Students

Advanced Students	Student CAT+ Corpus	Student CAT	Student Control
Guessing	16	4	0
Self-correction	2	3	0
Elimination	0	5	17
Elaboration knowing meaning	1	6	0
Elaboration not knowing the meaning	1	4	0
Monitoring against sense	11	11	19
Recall	1	1	0
Translation	19	19	19

Table 4.6 shows the strategies used among the three advanced students. Clearly, the control student did not have a wide range of strategies to use, for example there were no instances of guessing, self-correction, elaborating on knowing a meaning of collocations or phrases, elaborating on *not* knowing the meanings, and recall. In fact, she depended only on three strategies which were heavily used in comparison to the students in the other groups. She had 17 instances of elimination and 19 incidents each of both monitoring against sense, and translation. However, participants from both treatment groups used a wide range of strategies. The only strategy that was not used by the advanced student in CAT+Corpus was the strategy of elimination. This could suggest that she relied on strategies that worked sufficiently for her - translation and guessing - and thus used less of the other strategies, and did not need to use elimination at all.

4.7.3.3.2 High-Mid Level Students' Use of Strategies.

Table 4.7

Strategies Used Among all High-Mid Students

High-mid Students	Student CAT+ Corpus	Student CAT	Student Control
Guessing	9	3	6
Self-correction	6	0	1
Elimination	11	0	0
Elaboration knowing meaning	2	0	0
Elaboration not knowing the meaning	15	1	1
Monitoring against sense	20	4	7
Recall	1	1	0
Translation	19	18	19

According to Table 4.7, the student in the CAT+Corpus group used a much wider variety of strategies than the other two groups; she used eight strategies whereas the students in both the CAT and the control group used only five strategies. Furthermore, the student in the CAT+Corpus group took 31 minutes to answer the questions in the fill-in-the-gaps task in comparison to other students in the CAT group who took 20 minutes, and the student in the control group who took 18 minutes. This could be because the student in the CAT+Corpus group felt comfortable and took her time to answer thoroughly, leading to her using a wider range of strategies. Also, it can be suggested that the intervention that was given to the CAT+Corpus group gave them an advantage over the other groups and in this case the student in the CAT+Corpus group used more strategies and felt comfortable while answering the task.

Among all the strategies used by the high-mid proficiency students, the strategy of translation and L1 was used the most as illustrated in table 4.7. There was one instance in one

question where the student in the CAT group did not rely on L1 to answer the question. It was the first question of the task:

Fill-in-the-gaps task question 1: The costs, _____ the risks, of moving the satellite to and from space would have been difficult. a) not to mention b) owing to c) thereby

To elaborate more, this high-mid student in the CAT group skipped this question and returned to it towards the end along with other skipped questions. However, this student did not translate in both times. When comparing the answer to this question with other answers from her transcript, it is clear that she did not know the translations of the three collocations included in the multiple-choice options. She would translate other words, except for these specific words and collocations. She also answered incorrectly in her paper and selected *thereby* as her answer to this first question (the correct answer was *not to mention*). This could suggest that students are more likely to translate when they know the translation; and in cases that they do not know the translation, they will try to solve the question in English.

Similar to the advanced- level student in the control group, the high-mid level student in the control group shared the same pattern of using fewer strategies and had no instances of recall or elaborating on knowing the meaning.

4.7.3.3.3 Low-mid Level Students' Use of Strategies.

Table 4.8

Strategies Used Among all Mid-Low Students

Mid-low Proficiency Students	Student CAT+ Corpus	Student Control
Guessing	3	14
Self-correction	0	0
Elimination	4	5
Elaboration knowing meaning	0	0
Elaboration not knowing meaning	1	0
Monitoring against sense	15	6
Recall	0	0
Translation	19	19

Table 4.8 shows that both the CAT+Corpus and the control students were similar in relation to the frequency by which they used strategies in the fill-in-the-gaps task, with the exception of using guessing and monitoring against sense. The strategy of guessing was used much less by the CAT+Corpus student (only in three incidents); almost four times less than the control student (see Table 4.8). Conversely, the strategy of monitoring against sense was used in many occurrences by the CAT+Corpus student, whereas it was used much less by the control student. This could suggest that the use of a corpus helped develop certain skills for the CAT+Corpus student allowing them to find the meaning of the target items by looking at the context instead of guessing the target item. Furthermore, when it came to the test scores, it was the low-mid proficiency level student in the CAT +Corpus group who scored 14/19 compared to the low-mid proficiency level student in the

control group who scored 9/19. This suggests that alongside frequency of strategy use, it is also important to consider how effectively strategies are used.

Table 4.9

Strategies Used Among All Low Level Students

Low-Proficiency Students	Student CAT+ Corpus	Student CAT	Student control
Guessing	10	9	10
Self -correction	3	0	0
Elimination	7	2	0
Elaboration knowing meaning	3	2	0
Elaboration not knowing meaning	2	5	1
Monitoring against sense	15	3	7
Recall	0	2	0
Translation	19	19	13

The data shown in Table 4.9 above point to three main findings. First, students from both CAT groups were more likely to use more strategies to answer the fill-in-the-gaps questions, whereas the control student used only four out of eight strategies. Second, the CAT+Corpus student showed a relatively high use of elimination and monitoring against sense, indicating that the use of a corpus might have helped students develop more skills compared to the control student. Third, the low-proficiency control student used translation in 13 questions in comparison to the other students in the treatment groups who used translation in all 19 questions; and she answered correctly in all of these remaining six questions that did not include translation. These questions were the last questions in the task, and she quickly went through them. This could suggest she got tired from the task and from translation, or that she knew the answers very well and felt competent enough to answer them quickly without translation.

4.7.4 Students' Perceptions of the Teaching they Received

The following section will address the third research question: *what are the views of the learners on the teaching they had received? Do these views illuminate the learners' strategy use and how well they did in the tests?*

To answer this research question, the views of learners towards the intervention they received in the follow-up interviews, will be discussed. As mentioned earlier, the views are from the two treatment groups as the control group was not asked about their experience since they were not exposed to any intervention. Students' views from both groups suggested that the interventions they received were useful. For example, the use of the L1 and the bilingual corpus were mentioned by students in the CAT+Corpus group, and students in the CAT group identified the use of L1 as beneficial. The following section will elaborate further on students' views on the interventions and test scores.

Firstly, the results obtained from the receptive delayed post-test and fill-in-the-gaps task are reported here in order to illustrate how well students performed in these tasks. This is followed by insights from the follow-up interviews, where students talked about their views of the intervention and the strategies they used to complete the tasks. As mentioned earlier, the receptive delayed post-test, which served as a guideline to select the participants, was the last test taken by the participants before the fill-in-the-gaps task. This was because the delayed receptive test consisted of all 19 target items and it required use of receptive knowledge, which was similar to the fill-in-the-gaps task. While results from both tasks were reviewed by me, these were not compared since the tasks are different. Table 4.9 illustrates students' scores from the delayed receptive post-test and the fill-in-gaps task.

Table 4.9

Scores for the Receptive Test and Fill-in-the-Gaps task.

Participant	Receptive delayed test score	Fill-in-the-gaps task score
Adv. stud. CAT+Corpus	19	19
High-mid stud. CAT+ Corpus	13	11
Low-mid stud. CAT+Corpus	10	14
Low stud. CAT+Corpus	5	11
Adv. stud. CAT	18	11
High-mid stud. CAT	11	10
Low-mid stud. CAT	9	9
Low stud. CAT	4	7
Adv. stud. Control group	17	12
High-mid stud. Control group	15	9
Low-mid stud. Control group	13	9
Low stud. Control group	7	12
Total test score	19	19

The data shown in Table 4.9 suggest there are no clear patterns for the scores of the three groups except for the fact that one advanced student in the CAT+Corpus scored all answers correctly in the receptive delayed post-test and the fill-in-the-gaps task. The other two advanced students who scored highly in the receptive delayed post-test, scored no higher than the low-mid proficiency students in the fill-in-the-gaps task. However, looking at the scores of other students, it can be noted that three out of four students from the CAT+Corpus group achieved a higher score on the fill-in-the gaps task or achieved the same score as their receptive delayed post-test score, with the exception of the high-mid proficiency level student who scored similar to a low-

proficiency level student in the same group. In fact, the scores from the fill-in-the-gaps task suggest that students of all proficiency levels from the CAT+Corpus group performed better than the other two groups, for the most part. Students' comments from this group further indicate that the intervention was useful for their learning.

One could question the exceptional score of the CAT+Corpus advanced student who achieved full marks in both tests, and wonder if she knew the target words and collocations beforehand. However, her receptive pre-test score showed that she scored 13 out of 19 which indicated she did indeed learn through the intervention and was able to successfully achieve a full score in both the receptive delayed post-test and the fill-in-the-gaps task. In her interview, the advanced student confirmed that she did indeed learn through CAT+Corpus. In one of her responses, she stated:

> The corpus helped me to learn words with preposition. If a preposition would come after a word, it would change the meaning. I feel this is a weak point in me. The corpus can help me in the future when prepositions follow words… because the meanings change accordingly.

This excerpt illustrates how this student learned to look at context and see words preceding and proceeding the target items to understand meaning. In fact, she exclaimed that this was a weakness in her skill set and mentioned how the corpus was helpful in enabling her to understand the meaning. Moreover, she also stated that there had been some issues with some target words and collocations during the fill-in-the-gaps task: "there isn't something difficult…hmmm but I get confused between nevertheless and thereby. I also took time with henceforth". Other students in other levels within the CAT+Corpus group made similar comments to the advanced level student. They confirmed that using the corpus had helped them in different ways. For example, the low-

proficiency level student commented that the corpus was useful in that she "could see the meaning of the word and other things." Also, the high-mid students shared similar perspectives on the corpus stating:

> it really added to me because we always memorise words, and sometimes we put it in a sentence and it does not work….but to read it twice is very helpful because you understand the context of the word and how to use it.

The comments of the students from the CAT+Corpus group during the follow-up interviews reflected their test scores in the fill-in-the-gap tasks, since they were generally higher than those of the two other groups. Furthermore, the four participants in the CAT+Corpus group seemed confident when referring to the intervention and the fact that it helped them with learning the words and collocations.

As mentioned earlier, Table 4.9 revealed some inconsistent findings across students in the CAT and control groups. In both groups, there were some students who scored lower than expected in relation to their proficiency level, whereas there were others who scored what would be expected of their level. However, the scores of the low-proficiency level students in the CAT and the control groups showed improvement on their previous scores in the receptive delayed post-test.

Students in the treatment groups mentioned the use of L1, yet it was evident that all three groups used L1 the most when completing the tasks. However, one low-proficiency student in the control group was a special case. This student did not use Arabic in six out of 19 questions within the fill-in-the-gaps tasks. Interestingly, she answered all six of these questions correctly. This can indicate that perhaps she found these target items easier than the others or that she did not need L1

anymore at her learning stage. Another possible reason might be that she was among the students that did not receive any intervention nor was exposed to L1 in class and thus, it could be that she did not consider answering all questions using her L1. This could be a possibility, although students were clearly informed to use either English or Arabic in the TA interview. However, because there was only one case, her reasons for using L2 fluently and correctly for the six questions cannot be ascertained. This, however, was not the case with the other students. For example, other students in the CAT group used L1 in completing the fill-in-the-gaps tasks and expressed relief towards having L1 when learning L2 words and collocations. For instance, a low-proficiency level student in the CAT group used L1 and translation for all questions. Towards the end of the fill-in-the-gaps task, the low-proficiency level student in the CAT group was asked about her experience with the intervention (using Arabic and English). She stated:

> The experience was really good…I liked the idea of seeing Arabic and English, because I'm seeing my language and the language that I'm learning and this is good…when it's all in English there are words I don't know. And if there is no Arabic translation I feel I don't want to finish the class…but to see Arabic and English makes it easier and I can remember it better.

One important point that was expressed by students in the two treatment groups and could have affected the tests scores and the students' learning, was the usefulness of repetition. Five out of seven students expressed the view that seeing the words and collocations was useful for their L2 learning. The advanced student in the CAT group stated an Arabic idiom commenting on this: "repetition teaches excellence". Another high-mid student in the CAT+Corpus group stated: "I did not need to study the words, I learned them through repetition". Based on these comments,

repetition could also have impacted the students' learning progress. However, it is unclear if the students meant repetition from the tests, from the intervention itself (i.e., repeating the target items from English to Arabic and Arabic to English), or both the tests and the intervention together. Reports by the control students could have illuminated this point since they did not receive any intervention. This suggests that the control groups should have been included in the follow-up interviews to gain more insight into this issue. This is a limitation of this study, which is discussed further in the Conclusion.

In the interviews, participants also commented on their learning of the target items. Participants from the two treatment groups reported some of their experiences with specific target items. For example, the high-mid student from the CAT group stated that the intervention helped her understand the meaning of the collocation *on the other hand.* She even commented on how she understood that *the upshot* and *the outcome* are synonyms. At the end of the interview, she commented as follows:

> The exercises engraved the information (referring to the target items), because I did not study them nor did I review them. I can solve them because I understood them without memorisation.

These target items were learnt better than others. The learning outcome was also reflected in most of the participants' post-test scores in receptive and productive knowledge. For example, most participants demonstrated understanding of the target item *on the other hand,* and this is reflected in the test scores (see Table 4.10 for pre and post-tests). This could be because the students' L1 has equivalent translations for this target item. However, there were target items that were more challenging such as *nevertheless* and *thereby* which was reported by the advanced

student from the CAT+Corpus group. She stated that she got confused between these two target items. Furthermore, there was one target item, *thereupon,* that students found particularly difficult (see Table 4.11 for pre and post-tests) and it was the least learnt among the other target items. The reason for this could be that there is no equivalent L1 translation for *thereupon.* However, the difficulty of this target item was not reported in the interviews among the treatment groups.

Table 4.10

Correct Scores for the Target Item 'On the Other Hand'

	CAT+Corpus 18 participants	CAT 19 participants	Control 19participants
Productive Pre-test	5	0	1
Productive Post-test	17	17	6
Receptive Pre-test	13	12	7
Receptive Post-test	17	19	10

Table 4.11

Correct Scores for the Target Item 'Thereupon'

	CAT+Corpus 18 participants	CAT 19 participants	Control 19 participants
Productive Pre-test	0	0	0
Productive Post-test	6	5	0
Receptive Pre-test	0	1	2
Receptive Post-test	10	13	6

In summary, this chapter has discussed findings of both the quantitative and qualitative data obtained from the three groups. In response to RQ1, the quantitative findings showed that the intervention was beneficial for both treatment groups (CAT and CAT+Corpus) and resulted in improved understanding of the target items in both productive and receptive tests. RQ2 investigated the patterns of strategy use emerging from the data using both quantitative and qualitative analysis. The use of eight strategies emerged in total, but three of them were used the most. These were: using the L1, guessing, and monitoring against sense. Amongst these, use of the L1 was the most frequently used strategy. This was also established in the follow-up interviews. In response to RQ3, data from the follow-up interviews confirmed that the interventions were useful. Again, use of the L1 emerged as the most frequently used strategy and students felt that it was necessary for learning the L2. In the interviews, students also commented on their perceptions of the target items in which some were described as easier to learn than others. These findings are further discussed in chapter Five.

Chapter Five

Discussion

5.1 Introduction

The research questions for the present study are as follows:

RQ1: *What is the impact on short-term learning of semantically non-transparent words and collocations (form and meaning recall) among high-intermediate learners of English of:*

(a) a contrastive analysis and translation (CAT) instruction by the teacher?

(b) a combined approach using both a parallel corpus and direct teacher CAT instruction?

c) how did the previous two approaches compare with an approach using neither CAT nor a parallel corpus?

RQ2: *What strategies do learners report using to understand semantically non-transparent words and collocations in a think-aloud task?*

RQ3: *What are the views of the learners on the teaching they have received? Do these views illuminate the learners' strategy use and how well they did in the tests?*

While research question 1 provides the main insights that form the backbone of this study, the outcomes of research questions 2 and 3 are also of significant importance as they provide a different lens through which to view the findings of this research. Indeed, the latter two questions give space for participants to voice their perspectives which paves the way for a clearer picture of their use of strategies and their unique thought processes following their exposure to CAT instruction. To the best of the researcher's knowledge, few if any studies have used think-aloud

(TA) interviews to explore the learning of complex lexicons (e.g., semantic non-transparent words and collocations) and to understand students' perspectives following CAT intervention. Thus, the aim of the current study was to carry out a quasi-experimental study to add more insight as well as depth to the findings. This chapter will discuss the quantitative and qualitative findings presented in Chapter 4, with reference to the theories and empirical studies referred to in the Literature Review.

Using a quasi-experimental design that involved 56 female first-year Saudi EFL learners, this study investigated the impact of CAT instruction on their short-term learning of semantically non-transparent words and collocations, which include linking adverbials and nouns. The researcher purposefully selected semantically non-transparent words and collocations that are not clear – that is, vague in meaning - the learning of which requires not only language development over time, but also movement away from the literal meaning of the constituents (Aronoff, 1976). Thus, L2 learners often find it challenging to learn semantically non-transparent words and collocations such as linking adverbials and nouns.

Since the meanings of these semantically non-transparent words and collocations were generally unclear and unguessable to some extent, it was decided for the purpose of this study to adopt a CAT instruction approach (Laufer & Girsai, 2008a) to teach learners the usage and meaning of these target items. CAT allows the instructor to draw attention to similarities and differences between the L1 and L2, which might lead to gains in the students' learning. Students were allocated to three groups, each of which was exposed to 19 target items in two blocks of teaching, which required all three groups to read the same texts. The first block of teaching included the following items: *in the long run, let alone, thereupon, on the other hand, the outcome, the upshot, thus far, nevertheless,* and *it boils down.* The second block included the following: *the*

bottom line, seeing that, thereby, owing to, henceforth, in so far as, the aftermath, not to mention, much less, nonetheless.

In each block, following the reading of a text in English, the two intervention groups experienced either combined instruction involving CAT with the use of a parallel English/Arabic corpus (the CAT+Corpus group) or CAT instruction with the use of a vocabulary exercise in paper-format (the CAT group). In both interventions, the meanings of the target words and collocations were briefly explained using both Arabic and English. The control group students, however, simply read the text and answered comprehension questions without any attention being brought to the vocabulary. The researcher would remain available to briefly answer any questions they had concerning the meaning of particular words. However, clarification was provided only in English, which was unlike the intervention groups where both Arabic and English languages were used. It is important to note that across all groups, the passages they read contained the target words and collocations.

There were some other key differences between the three groups. After reading the passage and answering the related comprehension questions, both treatment groups had different tasks. The CAT group was required to complete two tasks. In the first task, learners had to match the L1 target items from the word-box with L2 sentences and write these L1 items next to the L2 sentence containing the target lexical item. Conversely, the second task required learners to match the L2 target items in the word-box with L1 sentences and write each L2 target item next to the correct L1 sentence and target item. These L1 sentences had the same meaning as the L2 sentences in the previous task. The researcher discussed each target item together with the class by providing a brief clarification on the target items in order to raise learners' awareness of the similarities and

differences between the L2 target items and their L1 translations. Overall feedback was provided once students completed the tasks.

The CAT+Corpus group were also required to complete two tasks. In the first task, learners were required to search the target items in English by typing them in the corpus search box. Parallel lines of English and Arabic sentences appeared. Students saw both L1 and L2 authentic texts and were asked to read these texts in the corpus. They were also required to write the Arabic translation next to the target items in bold that were in the reading passages. There were times where the teacher had to redirect the students' attention to the sentences that included the correct target items, as some of the translations of the target items were inaccurate. In the second task, they were given a printed out and modified worksheet from the corpus (AEPC) that included Arabic target items highlighted within sentences. Learners were required to find the English translations of the target items (see Appendix A.5.6.3.). Similar to the CAT group, the researcher discussed each target item with the class and provided the same clarification on the similarities and differences between the L2 target items and their L1 translations. Overall feedback was again given to students at the end of the tasks.

The following sections in this chapter will include a brief reminder of the origins of CAT instruction in the research literature, and will then go on to discuss receptive and productive knowledge. Following this, the role of involvement load and the role of incidental learning in vocabulary acquisition will be covered, and lastly the effect of L1 on L2 vocabulary learning and the use of L2 strategies in L2 vocabulary will form the remainder of the chapter.

5.2 CAT and CAT+Corpus Instruction

The notion of *contrastive analysis* (CA), discussed in greater detail in chapter two, became mainstream in the 1960s. Lado (1957) defines the contrastive analysis of two languages as a

procedure which enables one to predict problems encountered by L2 learners or helps explain errors made by them. CA allows the teacher to make comparisons between the learners' L1 and L2 and predict difficulties which learners may encounter when learning linguistic patterns (Lado, 1957). CAT instruction is based on the premise that improved vocabulary learning comes from an approach which includes contrastive analysis and translation. This type of instruction guides learners to notice the differences and the similarities of both L1 and L2 with regards to individual words and the vocabulary system (Laufer & Girsai, 2008a).

5.2.1 Receptive and Productive Knowledge in CAT Instruction

The test of receptive vocabulary knowledge required learners to translate the semantic non-transparent words and collocations from L2 to L1, while the test of productive vocabulary knowledge required learners to translate the target items from L1 to L2. The receptive test thus focused on meaning recall, while the productive test focused on form recall. Meaning recall refers to being able to retrieve the meaning of a word, for example by writing a translation of the term in the L1, while form recall refers to being able to retrieve and produce the word in its correct form in L2. However, while the productive test can tell us whether learners were able to retrieve and produce the correct word in the L2, because the test included the first two to three letters of the target items, the results cannot be compared to productive tasks such as composition writing.

As reported in chapter four, an analysis of receptive and productive scores using a 2 x 2 mixed ANOVA showed that the CAT group outperformed the other two groups (CAT+Corpus and control) both in the receptive and productive post-tests. There were differences in how far the treatment groups outperformed the control group across the two areas of receptive and productive vocabulary learning, which are discussed in greater detail later in this section.

In the receptive test, all three groups performed significantly in the post-tests than the pre-test. However, for the productive post-tests there was no evidence of vocabulary gain in the control group, while there was a significant improvement in the post-test scores of both the CAT and CAT+Corpus groups. Scores for the CAT group were better but not significantly higher than those for the CAT+Corpus group in both the receptive and productive tests. Results indicate the impact of CAT and CAT+Corpus methods of vocabulary instruction on the short-term recall of 9 target items. Descriptive statistics for the productive skills indicate that learners in the CAT and CAT+Corpus groups showed vocabulary gains. The mean scores for the pre-test increased from .21 to 6.68 for the CAT group, and from .50 to 5.61 for the CAT+Corpus group. The control group, on the other hand, did not score significantly higher. Mean scores for the pre-test were .21 and .63 for the post-test. Moreover, inferential statistics for the productive skills showed both treatment groups' scores increased significantly with a large effect size. The findings also show that though the sample size was small, the observed power was low, making it difficult to detect an effect for the two intervention groups. Further research using CAT and CAT+Corpus could determine if it would be meaningful to have a larger sample size in order to obtain similar results.

The findings of the productive test suggest that the intervention for the CAT and CAT+Corpus groups, which comprised both incidental and intentional learning, led to improvement in the learners' productive knowledge of the target items. The CAT and CAT+Corpus groups experienced incidental learning through the reading comprehension task but also intentional learning through aspects of the CAT and CAT+Corpus tasks. It has been previously suggested that while incidental learning may play an important role in developing learners' receptive knowledge, it may be insufficient to develop productive vocabulary knowledge. Schmitt (2008) argued that "intentional vocabulary learning almost always leads to greater and

faster gains, with a better chance of retention and of reaching productive levels of mastery than incidental vocabulary learning" (p. 341). Furthermore, participants in the CAT and CAT+Corpus groups engaged in learning intervention activities that included productive tasks. Webb et al. (2020) suggest that learning vocabulary items in such tasks, for example where learners are presented with the meaning of words and need to produce their equivalent forms in the L2, leads to higher gains in productive vocabulary knowledge compared to receptive learning tasks. This helps explain the difference in the productive post-test scores observed in the two intervention groups and the control group.

However, descriptive statistics for the receptive skills revealed an interesting finding. They indicated that while mean scores for the CAT and CAT+Corpus groups increased in the post-test (with an increase in mean scores from 3.32 to 7.84 for the CAT group, and from 3.28 to 7.22 in the CAT+Corpus group), there was also an improvement in scores for the control group (with an increase in mean scores from 3.16 to 5.21) suggesting that incidental learning may have played a role here. Furthermore, the descriptive statistics show that the *SD* for the control group increased, while the *SD* for the CAT and CAT+Corpus group decreased in the post-test. This seems to suggest that, in contrast to the two treatment groups, the control group's post-test scores were more spread out, indicating inconsistent learning. Furthermore, inferential statistics for the receptive skills revealed that both treatment groups' scores increased significantly with a large effect size whereas the control group's scores increased significantly with a medium effect size. Similar to the productive test, there was low power and small effect size between the two treatment groups.

These findings indicate that CAT instruction benefited both treatment groups in both receptive and productive knowledge in the learning of semantic non-transparent words and collocations. The control group findings show that in the receptive and productive post-tests, they scored lower than the two treatment groups. This confirms previous empirical evidence in both reading (Laufer & Girsai, 2008a; Sonbul & Schmitt, 2010) and listening contexts (Zhang, 2018), and in corpus use (Alharbi, 2017) indicating that in order to improve learners' L2 vocabulary learning, incidental vocabulary learning needs to be enhanced by some kind of lexical focus on form (Laufer, 2005; Schmitt, 2008).

Importantly, the control group still made significant pre to post-test vocabulary gains in the receptive test despite the absence of any kind of CAT instruction. The meaning of the target items had to be inferred from the text, and vocabulary was explained briefly in English only when it was requested by the students, meaning that learning was largely incidental. The instructor did not focus on raising awareness and drawing attention to the similarities and differences between L1 and L2 for each target item as was the case with the treatment groups. Instead, the focus was on reading the passage and answering comprehension questions only.

These results of the productive and receptive post-tests have implications for the debate around the role of incidental versus intentional learning in vocabulary development in second language learning. As Hulstijn (2003) points out, one view of vocabulary learning suggests that in order for learners to acquire new vocabulary, they need to be involved in intentional learning. This intentional learning involves a deliberate acquisition of thousands of words, including their meaning, pronunciation and spelling. The other view suggests that language learners can also acquire new vocabulary through incidental learning. This can take place through taking part in learning activities such as reading or listening tasks, which focus on the meaning conveyed rather

than the form of language (Hulstijn, 2003). The results in the present study suggest that both the intentional learning aspects of the CAT and CAT + Corpus group tasks and incidental vocabulary learning, which took place in the reading comprehension task, led to an improvement in vocabulary knowledge in the treatment groups. Laufer (2005) argued that vocabulary learning, which focuses on meaning, can also include a focus on form (FonF). This could include drawing the learners' attention to the form of target items during a meaning-centred task. Scholars have previously acknowledged that meaning-centred instruction alone is not sufficient to support second language learners' grammatical competence, therefore exposing the need for a focus on form that draws attention to linguistic elements during a communicative activity (see Ellis 2001; Long 1991).

When it comes to the three groups, all participants had encountered the target items three times before the receptive post-test: in the initial pre-test and twice during the learning intervention, which focused on both receptive and productive knowledge. However, in the case of the control group specifically, improvements in receptive knowledge can be attributed to incidental learning, since there was no intentional learning of the words within the learning intervention. It has been previously suggested that repetition and frequency of exposure play an important role in incidental learning (Horst et.al, 1998; Webb, 2007). The control group scores suggest that incidental learning was helpful for the learning of semantically non-transparent items in receptive knowledge, but not in productive knowledge. This could be due to the fact that receptive learning of vocabulary appears to be easier than productive learning. In fact, as previously mentioned in section 2.1, research suggests that receptive skills develop before productive skills (see Nation, 2013) and therefore the incidental learning, which took place in the control group may have been sufficient to improve receptive knowledge of the target items but insufficient to make a difference in the productive post-test scores. For example, when faced with a collocation such as *it boils down*

to or *the upshot*, the control group found it easier to translate from L2 to L1, but struggled with the translation from L1 to L2. However, even though the control group made gains in the receptive post-test, these were smaller than for the intervention groups. This could be attributed to the complexity to produce L2 target items (see Nation, 2001). Moreover, it has been previously suggested that intentional vocabulary learning tasks can result in higher vocabulary gains (see Horst et al., 1998; Hulstijn, 1992; Laufer & Girsai, 2008a), which would help explain some of the differences observed in the post-test results of the three different groups.

Another important factor is the complexity of the relationship between the development of receptive and productive knowledge. Multiple factors can affect learners' ability to recall vocabulary and successfully retrieve it. In addition to frequency of exposure, one factor that could help explain the difference in learners' performance in the receptive and productive tasks may be the participants' prior learning experiences. It has been acknowledged that teachers may find receptive tasks easier to "design, grade, and complete than productive tasks" (Webb, 2005, p. 34), which could mean that they use receptive tasks more often, which eventually leads L2 learners to be accustomed to and familiarised with receptive tasks much more than productive tasks (see Melka, 1997; Nation, 2001). Another reason could be that learning receptively occurs at an earlier stage of learning and prior to productive learning (Nation, 2001), and thus learners are generally more capable of acquiring receptive knowledge than productive knowledge (Nation, 2013; Schmitt, 2010).

In light of the above, it can be inferred that receptive vocabulary knowledge will always be much larger than productive vocabulary. Nation (2001) elaborated on the areas of receptive and productive vocabulary learning and attempted to describe the four differences between them i.e., *the amount of knowledge*, *practice*, *access*, and *motivation*. The first difference is the complexity

to produce L2 vocabulary as it requires extra knowledge of the new word whereas with receptive learning, learners require less knowledge to gain vocabulary knowledge. This is more obvious in languages that do not share similarities such as Arabic and English, where language elements (e.g., grammar, writing, etc.) are very different. The second difference implies that receptive tasks receive more practice than productive tasks (Melka, 1997; Nation, 2001). The third difference suggests *access,* as when using L2-L1 receptive knowledge, the learner is required to have one-on-one translation knowledge only, while for L1-L2 productive use, the learner needs knowledge such as grammatical forms, culture, and context along with L2 translation knowledge. Lastly, Nation underlines that if a learner lacks the motivation to use the vocabulary that has been previously learned, then learning new words may not be useful. The size gap between learners' receptive and productive vocabulary knowledge has implications for teachers, who may need to re-focus their priorities and question what it is they need to target so that they can develop learners' productive vocabulary knowledge. This also implies that productive tasks need attention in vocabulary learning.

Teachers can also apply explicit methods such as CAT to help their students engage more in productive activities bearing in mind that these productive tasks can affect receptive learning greatly. This was seen in Webb's study (2005) where he compared two experiments that had similar reading (receptive) tasks and writing (productive) tasks with the difference of allotted time. The first experiment had controlled time whereas the second experiment was without time constraints. He concluded that learners showed more vocabulary gains from writing than from reading. Also, his findings revealed that there was a correlation between productive tasks and productive learning, although this was not the case with receptive tasks and receptive learning. In

fact, he argued that receptive learning was more effective and that better learning came from productive tasks because students spent more time on writing tasks than on reading activities.

The current study revealed the usefulness of CAT instruction in incidental L2 learning for learning semantically non-transparent words and collocations. It is worth noting that within incidental learning, brief explicit teaching can be included such as CAT methods. Previous research (e.g. Alharbi, 2017; Laufer & Girsai, 2008a), as well as the findings of the current study, encourage using a CAT method in L2 vocabulary learning and making use of the L1 in L2 classrooms when it seems to be to learners' advantage (Schmitt, 2008).

5.2.2 The Role of L1 in L2 Learning in CAT Instruction

The interview data from the 11 think-aloud (TA) interviews, which included seven participants that received the CAT intervention and four from the control group, showed that participants made frequent references to the L1, suggesting that L1 plays an important role in the learning of semantic non-transparent words and collocations. For example, in the TA interviews, all participants used the L1 in their thinking process to answer most of the 19 fill-in-the-gaps questions. In fact, the strategy of L1 reliance was used in 202 occurrences, indicating that it was the most frequently used strategy. It is worth noting that the researcher did not specify which language learners should use in their TA interview. The students chose to use L1 while completing the task.

In addition, in the TA interviews, students reported that they relied on the L1 even when using other strategies (e.g., monitoring against sense and guessing). The students' comments also indicate that this type of learning, where the L1 words and collocations were compared to the meaning of the L2 vocabulary, was very useful. For example, one participant stated "…if there is

no Arabic translation, I feel I don't want to finish the class…but to see Arabic and English makes it easier and I can remember it better". These findings indicate that L1 was used in L2 learning and processing. In fact, L1 was used constantly with the target items and newly encountered words. This was done mainly by finding an equivalent L1 translation to L2 words.

This finding supports the vocabulary learning hypothesis that argues how L1 translations can help transfer the semantic concept to L2 and that learners can acquire vocabulary through word associations from L1 to L2 (Jiang, 2004). An earlier study conducted by Jiang (2002) found that learners associate semantic concepts in L1 and L2. The present study also found that performance scores improved when the first language and the target language had similar word translations compared to where there were no associations between the two languages. Moreover, learners had less difficulty in identifying target items, for which the translation was similar to Arabic. This was evident in the present study where learners were still heavily reliant on the L1 translation of the target items. For example, phrases such as *on the other hand* would be translated into Arabic as *from the other perspective* (من وجهة أخرى) or *from the other side* (من ناحية أخرى). Here there is a similarity between the word and concept *other*, which helped learners associate the meaning. In fact, this target item was identified correctly along with *in the long run* more often than any of the other items included in the test. For example, in the receptive pre-test, 12 out of 19 learners from the CAT group identified the target item *on the other hand* correctly while only two out of 19 learners produced the correct item in the productive pre-test. These scores improved considerably in the post-test, where all learners in the CAT group identified the target items correctly. In the productive post-test 17 out of 19 students were able to produce the correct item in the L2.

According to Jiang's (2000) three-stage model of adult vocabulary acquisition, learners in the first-stage understand the meaning of a word in the L2 by linking it to an existing semantic

structure in their L1. In the second stage, semantic content has been transferred from its L1 translation, which means that the L2 word is now directly linked to conceptual representations. In the third stage, lexical knowledge which is unique to the L2 word is integrated into the entry and the information in the L1 is eliminated. According to Jiang (2000), learners who have reached the third stage of lexical development are able to use L2 words more fluently, and with more automaticity, with little influence from translation from the L1.

It would appear that items with similar associations in L1 and L2 pose less challenge to learners and can thus result in an increase in the number of correct responses. Greater attention should therefore be placed on highlighting concepts which are conveyed differently or are absent in either the L1 or the L2 vocabulary. Take for example, when teaching learners the meaning of the phrase *on the other hand;* teachers can explain that *hand* means another point of view here and should not be confused with its literal meaning i.e. a part of the body. Another example is *the aftermath* where the teachers should clarify that *math* here is not to be understood as *mathematics*. A third example is the phrasal verb *it boils down to*, which is unlike the previous phrases, since there is no partial similarity in any of the words here. The Arabic translation corresponds to *it sums* (يتلخص) which would be similar to the phrase *to sum up* in English. Learners did not perform as well on this target item compared to others where there were similarities between the L1 and L2. The target item that posed greatest challenge was *thereupon* (see chapter four) which corresponds to *consequently* (بناء على ذلك) in Arabic. When teaching the target item, it was important to point out to learners that the term was not related to the meanings that the words *there* and *upon* have on their own. The adverb *thereupon* has a completely different meaning.

These examples of target items correspond to Lee et al.'s (2017) study, which suggests that particular vocabulary items might need specific treatments to be learned. For example, more time

could be allocated for more complex vocabulary items that do not have equivalents in the L1 (e.g., *thereupon*). Based on the findings of this study, it could be suggested that educators should consider explicit teaching methods such as CAT to explain the differences between L1 and L2 vocabulary. As shown, this can promote better and quicker retention of L2, especially for vocabulary items that are semantically non-transparent such as the words and collocations used in this study.

5.2.3 *CAT Instruction Understood Through the Lens of ILH*

When comparing the two treatment groups, as previously mentioned, findings show that the post-test results for the CAT group were slightly higher than those of the CAT+Corpus group in both receptive and productive knowledge. This finding indicates either that the CAT group were given tasks which resulted in an advantage over the CAT+Corpus group, or that the tasks used in the CAT+Corpus put this group at a disadvantage when comparing their outcomes to the CAT group. It appears that the type of activity was of key importance, since both groups received the same CAT instruction, allotted time for the intervention, while the participants were of the same proficiency level. It can only be suggested, therefore, that the difference lies in the involvement aspect.

By observing the data through the lens of ILH, the only difference found is the presence or absence of the corpus and hence the tasks the two groups completed. The CAT group received a paper-format consisting of sentences from the AEPC. These sentences were selected by the researcher to convey the exact meaning of the target items in context (see Appendix A2). The paper-format included two sheets, in which the first page had nine Arabic target items in a word-list and nine sentences in English, that included the English target items. The second page had a

word-list of nine target items in English and nine sentences in Arabic with the translated target item highlighted. After the CAT instruction, the CAT participants were asked to match each English target item with the correct L1 sentence and target item.

The CAT+Corpus group tasks required greater involvement from learners because in addition to the sentences in English and Arabic, which were the same ones used for the CAT group, they also had access to the AEPC. Learners had to look up the target items by using the search box. This would generate results consisting of sentences in English and their corresponding Arabic translation. The search would generate several sentences, sometimes reaching up to six examples of how the target item can be used in context. The target items entered into the search box were automatically highlighted, however their translation was not. This is because the corpus automatically highlights only the item that is entered but not its translation. The students would read both the English and the Arabic sentences. The researcher instructed the learners to ignore certain sentences from the AEPC because they did not convey the exact meaning of the target items. For example, the target item *the bottom line* would be literally translated as the last line in a text (الخط السفلي). Learners would then repeat the same procedure for all nine target items. Following that, the CAT+Corpus group also completed a modified worksheet from AEPC that included nine Arabic sentences with the target items highlighted. To complete this task, they were required to search for the English target items. Having this type of involvement might have affected their recall of the target items, which will be discussed in detail in the following section.

5.2.4 *ILH and the Three Components (Need, Search and Evaluation)*

The involvement load hypothesis (ILH) by Laufer and Hulstijn (2001) suggests that retention of unfamiliar words in the L2 is dependent upon the involvement load (IL) of a task. The

findings of the present study, however, seem to suggest otherwise since the CAT+Corpus group had slightly lower scores in the post-test compared to the CAT group, but the CAT+Corpus instruction had been deemed to induce greater IL. The ILH includes three components: *need*, *search* and *evaluation* (Hulstijn & Laufer, 2001). In order to calculate the independent load (IL) of a task, one needs to calculate the points for each of the three components (Yanagisawa & Webb, 2021).

The *need* component refers to the motivational factor. If unknown words are needed to complete an activity, then *need* is present, whereas it is absent (0 points) when an unfamiliar word is not required to fulfil the task. *Need* is moderate if the teacher asks learners to understand a word or instructs them to use this word (1 point). *Need* is strong when learners seek to understand the word themselves (2 points). An example of a moderate *need* would be when a learner is asked to write a sentence using an unfamiliar word, whereas *need* is strong if, for example, a learner needs to know what a word means because they want to use it in writing or speaking. In this study, *need* was moderate, since learners were asked by the researcher to use the target items.

Search is understood as the attempt to find the L2 equivalent of a word or its meaning. Two levels are suggested for *sear*ch: presence or absence. Where a learner needs to use external resources, such as a dictionary or ask a teacher, then *search* is present (1 point), whereas it is absent (0 points) if the L2 form and the meaning are provided within the same activity. In this study *search* was present, since learners had to identify the correct translation for the target items, although these were included in the task.

The third component, namely *evaluation*, comprises a comparison of an unfamiliar word's form and meaning with other words. The learner then needs to identify the most suitable word for the context. *Evaluation* is absent (0 points) where learners are not required to select which word

to use. On the other hand, it is moderate (1 point) when learners are given a context, such as a gap-fill activity, where they need to choose the most suitable word to use given the context. *Evaluation* is strong when learners need to use a word in a productive context. An example of a task with strong *evaluation* (2 points) would be composition writing using target words.

The current study took into account Yanagisawa and Webb's (2021) criteria in scoring the components i.e., *need* and *search* but with minor adjustments in scoring the component *evaluation* for the two treatment groups. For both groups, there was a moderate level of *need* where students were asked by their teacher to understand the target items; therefore, both groups had equal *need* which, according to the criteria, is equivalent to (1). Additionally, both groups searched for the item words: the CAT+Corpus participants searched by writing the English target item in the search box in the corpus, followed by a search for the English translation of the Arabic target item in the modified worksheet printed from AEPC. CAT participants searched for the correct target items in the word-list to align it with the translation. Both groups were provided with L2 form and meaning within the same activity. Hence, according to the criteria, *search* was present (1) for both groups.

Regarding the last component, *evaluation,* both groups read authentic texts and compared a word's form and meaning, which according to the criteria above, gives an evaluation score of 1 point. However, the CAT+Corpus group dealt with numerous sections of text, both in paper and digital formats, that involved accurate and inaccurate translations of each target item, while the CAT group was provided with one selected sentence with accurate translations for the target item. As such, when comparing the CAT group with the CAT+Corpus group, the task of the former was actually less demanding and involved less input. The CAT group were asked to compare and match the most suitable word to the given sentence, which could suggest that an *evaluation* point of 0.5 is actually more suitable than 1. This score is uncommon according to the ILH criteria but it can

be justified when looking at CAT's task, which was worth more than 0 points, yet less than the CAT+Corpus group, whose *evaluation* point was 1. The overall IL scores for the CAT group would then add up to 1.5 points and for the CAT+Corpus group to 2 points, because of the higher score for *evaluation.*

For the CAT+Corpus group, this *evaluation* score might have been a factor that influenced learners' results in the productive and receptive tests, which were slightly lower compared to students in the CAT group. This finding agrees with previous studies that have suggested that predictions of the ILH were not always accurate (see Folse, 2006; Keating, 2008). Reasons for this inconsistency include factors such as time spent on the task (e.g., Keating, 2008). While both the CAT and CAT+Corpus groups were allocated the same amount of time for the task, because the CAT+Corpus group had access to many more examples of sentences of how the target items were used in context, they may actually have had less time to focus on the purpose of the task.

In a comprehensive literature search, Yanagisawa and Webb (2021) found 42 studies that examined the predictions of ILH. Based on these studies, they analysed the scores and examined the extent to which ILH predicts incidental vocabulary learning and the intensity of impact of the three components of the ILH on learning. They also evaluated other factors (i.e., time spent on the task, frequency, and aspects of vocabulary knowledge). Among their findings, they found that *evaluation* was the component that contributed the most to learning, followed by *need*. However, they also stated that the component *search* did not have any effect in contributing to learning.

Additionally, the corpus might have been a distraction for the CAT+Corpus group which could have also affected their learning. It has previously been suggested that accessing external sources, such as a bilingual dictionary could actually act as a distraction to learners rather than assist L2 learning. Nawal (2018), for instance, found that the use of an Arabic-English dictionary

during a writing task seemed to have a negative influence on learners' L2 writing, possibly because it acted as a distraction. In a similar vein, results from the current study suggest that the CAT+Corpus group may have found that having access to an extensive amount of sentences within the corpus is a distraction rather than a helpful tool because of the higher cognitive skills (Flowerdew, 2015) needed to complete the task.

In a meta-analysis study, Lee et al. (2019) investigated the impact of corpus use on L2 vocabulary knowledge. They found that L2 proficiency, along with several characteristics of using a corpus (i.e., interaction types, corpus types, training, and duration), have a major influence on L2 vocabulary gains. In a previous study, Lee et al. (2017) found that the complexity of interaction in learning target items involves the following areas: the context of the target items, learners' clicking behaviours, the selected concordance lines, and the learner's proficiency background. For example, Lee et al. (2017) suggested that the most appropriate number of concordance lines for prompting inference of their meaning is three lines. This contrasts with the current study where, at times, more than three concordance lines were given, depending on the target items.

Based on the above, it could be argued that learners who are more proficient in the L2 may be able to draw greater benefit from CAT+Corpus approaches because the cognitive load would not be as high for them. The data in the current study indicated that students overall had not developed the reading skills necessary to draw the most benefit from engaging with authentic texts (such as the text in corpora). Furthermore, the findings suggest that not having an appropriate proficiency level limits learners from employing higher order cognitive skills (i.e., reading) especially when moving between languages. This is because corpora in general include authentic and unfamiliar texts unlike school-oriented materials which are tailored to the learner's proficiency level. Using corpora requires efficient reading strategies. In the present study, the information sheets showed

that out of the 56 participants, only 28.5% across all three groups read English books and 53.5% read English online. In relation to the CAT+Corpus group, which included 18 participants, only seven of them (38%) read English books and 10 out of 18 students stated that they read English online (55%). While this information is helpful, the study did not explore the relationship between participants' reading of books in English and test scores. However, further research could look into possible correlations between learners' time spent reading books in the L2 and ability to engage with corpora.

According to Haynes' (2010), review of Grabe's book of L2 English reading strategy research, learners who are good readers are more effective in strategy use in comparison to poor readers. This is because good readers are more capable of using a wider variety of reading strategies and are able to apply strategy use when facing complex texts. One important strategy in reading is making inferences effectively. Haastrup (1991) confirms that L2 proficiency is a key factor in lexical inferencing and that learners have to reach a specific level in L2 proficiency to efficiently use inferences strategy. Furthermore, to make effective inferences, one must be familiar with a certain percentage of words while reading. Earlier studies estimated that approximately 95% of words in a text should be known (e.g., Hirsh & Nation, 1992). More recent studies have increased this estimate to 98% (Hu & Nation, 2000; Nation, 2006). Based on his data within the British National Corpus, Nation (2013) calculated that approximately 8000-to-10,000-word families (with proper nouns) are needed to reach 98% coverage of the reading of texts to successively read texts. This suggests that the CAT+Corpus group might not have been able to make effective inferences while reading the concordance lines due to the limitation in their vocabulary background. As a result, the learning of the target items may have been affected. A

limitation in the current study was that there was no information regarding participants' vocabulary size.

5.3 *Strategies Used by the Three Groups*

To answer RQ2, the TA interview data were analysed. The study's findings showed that learners adopted particular vocabulary learning strategies, some of which have some overlap with those theorised in Nations's (2013) taxonomy of Vocabulary Learning Strategies (VLS). The strategies utilised by the participants in the study were the following: *the use of L1, guessing, monitoring against sense, elimination, elaboration*, *self-correction,* and *recall*. In particular, Nation's strategies of *sources* and *processes* (see section 2.1.3) have similarities with the strategies of guessing, monitoring against sense, elimination and recall. *Sources,* for example, refers to using context to infer the meaning of words, which is similar to the strategy of *guessing.* Similarly, *processes* includes noticing and retrieving, which has parallels with the strategies of *monitoring against sense* and *recall.*

Furthermore the findings show that the CAT+Corpus group used a wider range of strategies and more strategy occurrences compared to the other two groups (see Table 4.5 in chapter four). The CAT+Corpus group utilised: *the use L1, guessing, monitoring against sense, elimination, elaboration*, *self-correction,* and *recall*. This seems to suggest that the CAT+Corpus group may have developed additional strategies as a result of the intervention (although of course no pre-intervention areportssessment of learners' strategy use was made, making such a claim somewhat speculative). It is also possible that engaging with the corpus may have helped learners acquire transferable skills, such as more advanced reading skills and drawing on the context, especially when encountering difficulties with L2 vocabulary.

The fill-in the-gaps tasks along with TA interviews took place in week five and one week after the delayed post-test. Four learners from each of the three groups were selected to take part in this task. Participants were selected based on their results on the receptive delayed post-test (see Figure 3.1, chapter three, for original experimental procedure), and participants' proficiency was categorised as being either low, low-mid, mid-high or advanced based on these test scores. Results from the fill-in-the-gaps task indicated that the four participants from the CAT+Corpus group demonstrated the ability to draw on a variety of strategies when completing the task. Furthermore, their use of strategies could imply that they did not recall the learned target items and had to use strategies to understand them. A delayed-post-test, including all participants, would have been helpful to gain a better understanding of participants' retention of the target items.

The fill-in-the-gaps tasks with TA interviews revealed that participants across the three groups most frequently made use of three out of the seven strategies mentioned above. These were: *use of L1, guessing and monitoring against sense.*

5.3.1 Use of L1

All 11 participants made use of L1 as a strategy, and out of the 202 incidents of strategy use, only seven of these did not include L1. This suggests that the norm was reliance on the L1 to solve the fill-in-the-gaps task. Six out of the seven incidents where a participant did not use translation to complete the task were attributed to one single learner (as mentioned in section 4.7.4, chapter four). Interestingly, this participant was a low-proficiency student from the control group. The participant was able to answer all six multiple choice gap-fill questions correctly and without hesitation. The six items comprised the last six questions on the test. While it is not possible to

draw any conclusions from a single case, the student's approach towards these six questions could suggest that she was able to recall the meaning of the target items from previous exposure.

What is interesting to note here is that the student did not *think aloud* in Arabic while completing these six questions, and there was no mention of the translation in Arabic. The student simply read out the questions in English and selected the correct target item to complete the sentence. It was quite intriguing that this student with a low proficiency level answered six consecutive questions without using L1, as no other students from her level or even from other levels, did so. This was especially surprising given that in both the receptive and productive tests, the control group performed worse than the treatment groups. Nevertheless, even though she correctly answered all six questions, her final grade on the fill-in-the-gaps test was 12. Another student with high proficiency who utilised more L1 during her TA interview, had a final score of 17.

It is possible that the low-proficiency student's six correct responses were due to those particular target items being easier to recall than other items. Furthermore, incidental learning of the target items through reading and multiple exposure (see Pellicer-Sánchez, 2017; Rott, 1999; Webb et al., 2013) could explain why the student was able to confidently answer these last six questions without making use of L1 as a strategy. It is also possible that, because of frequent exposure, the student was able to recall the meaning of the target items in the L2 without having to refer back to the L1 translation. This could suggest that, at least for some target items, the learner reached the third stage of lexical development (Jiang, 2000) where lexical knowledge unique to the L2 word is integrated into the entry and the L1 is eliminated. At this stage L2 words can be used with more automaticity and little reliance on the L1, which is what was observed when the

student completed the last six questions. Moreover, since the fill-in-the-gaps is considered a receptive test, this suggests that receptive learning did take place in the control group.

5.3.2 Monitoring Against Sense

Participants from all three groups referred to this strategy 118 times throughout the fill-in-the-gaps task. This strategy entailed selecting one of the multiple-choice options and checking if it made more sense compared to the others. Participants relied primarily on the sentence provided in the fill-in-the-gaps question to select their answer. One of the participants, a low-proficiency student from the control group, used different strategies to answer the questions despite answering incorrectly. Her thought processes indicated that she was trying to monitor against sense but also made use of translation and guessing as strategies. She read the sentences out loud but hesitated when she came across the term *henceforth*. She tried to pronounce the word correctly and then asked herself in Arabic "what does it mean?". The participant then tried placing *henceforth* into the gap to see if the word made sense in the sentence but later decided to select *in so far as*. When she was selecting this option, she commented "I feel it is the closest thing". Perhaps because she relied on the guessing strategy, her answer was incorrect.

This may suggest that lower-proficiency students may use strategies less effectively compared to higher proficiency level students. Furthermore, in comparison to the other groups, the control students drew on fewer strategies to answer the questions. While there are few studies which related vocabulary learning strategies to vocabulary knowledge depth, it has been suggested that students with greater vocabulary knowledge may be more successful in lexical inferencing. Nassaji (2006), for instance, found that students who had more in-depth vocabulary knowledge

were able to apply specific strategies (such as section repeating, self-inquiry and confirming, and monitoring procedures) more frequently and more successfully than their less competent peers.

5.3.3 Guessing

This strategy was the third most frequently used one by all participants. In the TA interviews, participants used terms such as *I predict, I think, maybe* when they were selecting their answers. This was consistent across students from low to advanced proficiency levels. One low-proficiency student from the control group stated that all her answers were "predictions and assumptions" at the start of the test. The CAT+Corpus group used the guessing strategy in many instances. Participants from that group used it in 38 occurrences and more frequently compared to the other two groups. At times, participants guessed the correct answer and at times they guessed incorrectly. One advanced student from the CAT+Corpus group was able to answer all the fill-in-the-gaps questions correctly. Interestingly, she referred to guessing 16 times, sometimes mentioning the incorrect answer, yet answered all questions correctly. The frequent reference to guessing seems to suggest that participants were unsure of their answers and did not necessarily recall the meaning of the target items. This may indicate that the CAT intervention was more effective in improving short-term recall than long-term recall of the meaning of the target items, since learners were unsure of their answers during the fill-in-the-gaps task, which took place in week five.

5.4 Students' Observations of the Effectiveness of CAT Instruction

To address RQ3, the CAT and the CAT+Corpus groups were asked at the end of the TA interview about their experience using CAT instruction (for both groups) and AEPC for the CAT+Corpus group. The findings showed that both the CAT and CAT+Corpus groups elaborated on the view that the presence of Arabic in the learning tasks was helpful. Indeed, the excerpts from the follow-up interview indicate the importance of using L1 in L2 vocabulary learning. For example, the low-proficiency level student in CAT commented on how she felt when all the learning only involved L2: "I feel I don't want to finish the class". According to her words, she was encouraged to learn in a familiar learning context and preferred the integration of L1 in L2 instructional methods. This suggests considering L1 in L2 learning and exploiting it to the L2 learner's advantage (Schmitt 2008). Similarly, the advanced student in the CAT group stated that it was difficult to translate word-by-word when learning L2. This shows that L2 learners naturally depend on their L1 and try to make sense of new words by referring them back to L1 concepts. This supports Jiang's view (2004) that L2 learners acquire vocabulary through word associations from L1 to L2. Therefore, it makes sense to utilise L1 in L2 learning.

Furthermore, the advanced student in the CAT+Corpus Group expressed the view that the corpus helped them by showing words in context. This is another aspect of learning a word (Nation, 2013). As discussed in the Literature Review in chapter two, Nation (2013) states there are three aspects of knowing a word: *form*, *meaning*, and *use*. From the students' observations, there was not only form and meaning but a focus on *use* as well, which includes the context in which words are used. Although there was no evidence from the quantitative results to support this point, it was only speculated from the students' comments in the TA interview. Indeed, students in CAT+Corpus made reference to a wider range of strategies used and there were more strategy

occurrences suggesting that using AEPC might have equipped them for incidental learning in future reading activities. With that being said, they did not outperform the CAT group despite the fact that learners in the CAT+Corpus group had greater opportunities for incidental vocabulary learning because they had access to the AEPC. Also, because the AEPC automatically generates excerpts from texts based on the search item, learners were provided with examples of the translation of the target item used in different contexts.

In order to determine which examples from AEPC contain the correct use of the target item, learners also needed to infer the meaning of unknown words. However, L2 learners do not always attempt to infer the meaning of unknown words from the context. Some research suggests that L2 learners often ignore unfamiliar words and only infer meaning when there is a specific need (e.g., Paribakht & Wesche, 1997; Fraser, 1999). Yet, it has been previously suggested that reading for comprehension can contribute to some incidental L2 vocabulary learning (see Hulstijn, 1992; Paribakht & Wesche, 1993, 1997; Webb & Chang, 2012). Incidental vocabulary learning mostly occurs through the process of inferring word meaning. In her study on lexical processing strategies used by L2 learners, Fraser (1999) found that reading for comprehension in an L2 can be a helpful tool to support incidental vocabulary learning. A further study found that reading a series of thematically linked texts can assist students in understanding unfamiliar words and help them become more aware of their appropriate use (Kang, 2015).

While reading in the L2 has been thought to boost general reading proficiency and increase motivation, Kang (2015) argues that when learners read thematically linked texts, this can enhance both productive and receptive vocabulary knowledge. It has been previously acknowledged that the impact of reading on vocabulary acquisition is particularly important for learners with limited exposure to formal English education (Nation, 2013). Furthermore, language-focused instruction

in classrooms provides only limited opportunities for direct vocabulary learning. For example, according to Tang and Nesi (2003), only around 12% of new words that appear in Chinese textbooks are explicitly covered in class. This suggests that learners need to acquire additional words through incidental learning (Reynolds & Ding, 2022) or through independent studies. As previously mentioned in section 5.2.4, participants in this study did not do a lot of independent reading in English. Reading is thus regarded as a major incidental vocabulary acquisition approach (Webb & Chang, 2015).

5.5 Conclusions and Limitations

This study sheds light on how both CAT and CAT+Corpus approaches impacted short-term recall when learning semantically non-transparent words. The findings showed that the two intervention groups who received CAT instruction outperformed the control group in both receptive and productive vocabulary learning. When comparing between both treatment groups, findings also revealed that the CAT group scored slightly higher than the CAT+Corpus group. This may have been because there is more involvement of the *evaluation* component, although contrary to the ILH (Laufer and Hulstijn, 2001), it's role as a learning aid is not entirely positive. he findings also revealed that the control group was able to acquire receptive vocabulary learning in incidental learning.

Findings further suggest that while learners had previously encountered the target items both in previous tests and in the intervention tasks (for the CAT and CAT+Corpus groups), many participants did not fully recall the meaning of these items in the fill-in-gaps task and used various strategies to complete the task. This seems to suggest that while CAT and CAT+Corpus approaches contributed to significant short-term recall of the target items, a delayed post-test could

have provided a better understanding of whether CAT instruction might actually help students in recalling the target items in the long-term.

The findings of the fill-in-the-gaps task suggest that learners who were more proficient in English were able to make greater use of vocabulary learning strategies. Furthermore, participants from both groups articulated in the follow-up interview the importance of being able to compare English and Arabic sentences and the use of L1 in L2 vocabulary.

Important points for practitioners to consider could include thinking carefully about the purpose of L2 tasks (e.g., learning vocabulary or reading and understanding context) and aligning the provided task (e.g., using a corpus) with the proficiency level of the learners. Furthermore, educators need to be aware that there are differences in proficiency among learners, even in courses where students are placed in a particular level as a result of a placement test score. Vocabulary knowledge among students within the same group may therefore differ. It is also difficult to make any conclusions about learners' knowledge of the L2 based on years of schooling. For example, a study on vocabulary size undertaken within the Omani context found that students entered university with a vocabulary size of about 2000 words (Laufer, 2000) which is around 1,350 hours of L2 instruction, while it would be expected for learners to have acquired at least 6000 words (Muqaibal, 2020), which would be considered typical for having learnt a language in school for 12 years. This is considerably lower than native speaker level, for which a vocabulary size of around 18,000 words is the norm (Schmitt, 2010). Vocabulary size is thus a key indicator of proficiency, and attention should be placed on developing this area within L2 instruction. Approaches, such as CAT, can be an effective tool to support educators to build vocabulary knowledge from an early stage.

Additionally, learners in the CAT+Corpus group also mentioned the benefits of having access to the AEPC, as this provided them with more examples of how the target items were used in context. However, learners commented that some of the sentences generated by the AEPC contained the incorrect usage of the target items. While this may have presented a challenge, it is suggested that students may also have benefited from this. Indeed, being aware that the same word can be used differently in different contexts is also a valuable and transferable skill in L2 learning. Therefore, while the CAT+Corpus group performed slightly lower compared to the CAT group in the post-test, it is suggested that both CAT and CAT+Corpus approaches were effective. Nevertheless, it is possible that CAT+Corpus approach may be better suited for students whose proficiency level is more advanced as they would be able to handle larger amounts of text in the L2 with greater ease compared to lower-ability students.

It is important to note that a corpus like AEPC is an ongoing project, and although it started with a vocabulary size of 10 million words, the aim is to increase it to 27.8 million words in Arabic, and 30.8 million English words (Alotaibi, 2017). Alotaibi (2017) argues that Arabic–English corpora in general are limited when it comes to size (1-3 million words only). Therefore, the AEPC was used because it had a wider vocabulary size, while it was also user-friendly because it had the feature of translation. However, there are some disadvantages to using AEPC. For example, some of the target items did not work very well with this corpus. Not all sentences provided the exact meaning of the target items, especially when learners searched for semantically non-transparent linking adverbials, which are words that are generally not clear – that is, vague in meaning.

In summary, this chapter discussed both the quantitative and the qualitative findings. The following chapter is the Conclusion, which will summarise the entire study.

Chapter Six

Conclusion

6.1 Introduction

The present study explored vocabulary learning through CAT instruction in the ESL classroom in a university in Saudi Arabia. The primary aim of the study was to explore the impact of two instructional methods: contrastive analysis and translation (CAT) and CAT+Corpus on short-term recall of semantically non-transparent linking adverbials (words and collocations) in receptive and productive knowledge. This chapter summarises the study's results and addresses the research questions, which guided the investigation. Along with key findings, section 6.2 outlines the areas in which this study adds to our understanding of the topic. This is followed by the study's limitations, which are covered in section 6.3. Section 6.4 discusses recommendations and implications for educators and institutions. In section 6.5, some recommendations for future investigations are made. Finally, section 6.6 highlights how this research contributes to the field of knowledge in relation to second language learning.

6.2 Key Findings

6.2.1 First Research Question

The first question in the study was: w*hat is the impact on short-term learning of semantically non-transparent words and collocations (form and meaning recall) among high-intermediate learners of English of (a) a contrastive analysis and translation (CAT) instruction by the teacher?; (b) a combined approach using both a parallel corpus and direct teacher CAT instruction?; and c) how do the previous two approaches compare with an approach using neither*

CAT nor a parallel corpus? The major statistical findings are outlined in the below subsections. These include: (1) the effects of CAT instruction on receptive and productive knowledge of the target words and collocations; (2) the effect of CAT instruction with parallel corpus on receptive and productive knowledge of the target words and collocations; (3) receptive and productive results for the control group; and (4) a comparison of the receptive and productive results of the three groups.

6.2.1.1 The Effect of CAT Instruction with a Parallel Corpus on Receptive and Productive Knowledge. Participants in this group took part in an intervention using CAT instruction with a parallel corpus that involved both receptive and productive tasks. Eighteen students took part in the intervention. The intervention session required students to read passages taken from the students' textbook, with some modifications made to the linking adverbials used in the original source. After having read the passage, students answered the questions and completed the worksheets. Afterwards, students began working on the receptive task, in which they were asked to translate the target items highlighted in bold using the bilingual corpus available on their computers. In this task, learners were required to translate the target items from English into Arabic. They were asked to read the texts in the corpus, search for the target items, and type in the Arabic translation beside them. After completing the receptive task, students completed the productive task, which involved reading the Arabic target items and sentences and searching for the correct English translation. When comparing pre-test and post-test scores, results indicate that students made significant vocabulary gains in both receptive and productive knowledge. This suggests that this type of combined approach can improve learning outcomes for semantically non-transparent words and collocations in both receptive and productive forms. The results also

illustrate that the features within a parallel corpus (i.e., the appearance of two languages) can complement CAT instruction in L2 learning.

6.2.1.2 The Effect of CAT Instruction by the Teacher on Receptive and Productive Knowledge. The intervention for the CAT group was similar in structure but did not include the use of a corpus. The participants here experienced CAT instruction delivered by the teacher alone using receptive and productive tasks. They were asked to match the target items in the L1 to their equivalent in the L2 and vice versa. The task was in paper format. Nineteen students took part in the intervention. Results show that they made significant vocabulary gains in their post-tests for both receptive and productive short-term recall. This finding indicates that CAT instruction was useful in learning semantically non-transparent words and collocations in both receptive and productive forms.

6.2.1.3 Receptive and Productive Results for the Control Group. The control group did not receive any treatment of the kind experienced by the other two groups (i.e., CAT instruction, or a combined approach of CAT instruction and use of a parallel corpus). Nineteen students were included in the control group. The participants in this group experienced incidental learning and completed receptive tasks (i.e., they read a passage and answered comprehension questions). Between the pre-test and post-test, the control group made significant improvement in receptive knowledge scores, but not for productive knowledge. It can thus be concluded that incidental learning activities without explicit (i.e., CAT) or supplemental learning (e.g., a parallel corpus) can improve receptive knowledge of semantically non-transparent words and collocations, but not productive knowledge.

6.2.1.4 A Comparison between Receptive and Productive Results Across the Three Groups.

The two treatment groups who received CAT and CAT+Corpus instruction made significantly greater vocabulary gains on their post-tests for both receptive and productive short-term recall than the control group who only experienced incidental learning. However, among the two groups who received CAT instruction, the findings of the immediate post-tests for both reception and production suggest that learners who received CAT instruction led by the teacher learned more vocabulary items than learners who received CAT instruction combined with a corpus, though the differences between the two groups were not significant. This could imply that even though both groups received explicit CAT vocabulary instruction, CAT teaching alone could be more useful than supplementing it with parallel corpus tasks. This could be verified with more research on both interventions.

Completing the parallel corpus task required more involvement and more evaluation on the part of learners; a component of the involvement load hypothesis (ILH) (Laufer & Hulstijn, 2001). Although research suggests that (e.g., Yanagisawa & Webb, 2021) the more evaluation there is, the more that language learning occurs, this was not the case in this study. It can be suggested that using CAT instruction with less task involvement in the evaluation component (through the use of a parallel corpus) can achieve better vocabulary gains on a short-term basis. This indicates that learning tasks and learners' cognitive load should be taken into account when using technology. These results also indicate that, from a pedagogical perspective, vocabulary learning requires a well-balanced strategy of task involvement and reasonable cognitive load, with teachers being clear about the objectives of their educational approach and taking into account the supplemental tools i.e., bilingual corpus, complexity of the target items, and the proficiency background of the learners.

6.2.2 Second Research Question

The second question was *What strategies do learners report using in each condition to understand semantically non-transparent words and collocations?* To address this question, an analysis was conducted on the qualitative data obtained from TA interviews. Results showed that participants from the CAT+Corpus group used a wider range of strategies and had more occurrences of strategy use than the other two groups. This could imply that being involved in certain learning tasks such as using a bilingual corpus could have potentially led to developing a wider range of strategies for use in sentence contexts. It may thus be beneficial for teachers to use a bilingual corpus in L2 classes to show learners how to employ different strategies when dealing with vocabulary learning within context.

Across the three groups there were three strategies that were used most frequently: *the use of L1, monitoring against sense, and guessing*. The students relied on the use of L1 in most of their answers suggesting it can be the norm to use L1 in L2 vocabulary acquisition. Therefore, employing teaching methods that involve L1 (such as CAT instruction) makes sense and can be an effective approach in L2 learning.

When learners were using the strategy of monitoring against sense, there were instances that showed lower-proficiency students using it less effectively compared to higher-proficiency level students. The strategy of guessing was frequently used, suggesting that participants were unsure of their answers and did not necessarily recall the meaning of the target items. The ways in which the latter two strategies were employed in the study shows that they often use them incorrectly. In this respect, teachers should take into account the vocabulary learning goals in classes and whether the aim is to recall the target for the short-term or long-term. Accordingly,

teachers can plan vocabulary learning methods that promote suitable vocabulary training and strategies that align with the learning outcomes.

6.2.3 Third Research Question.

The third question for this study was: *What are the views of the learners on the teaching they have received? Do these views illuminate the learners' strategy use and how well they did in the tests?* To address this question, a qualitative analysis was used. Following the TA interview, the two treatment groups were asked about their views concerning CAT instruction while the CAT+Corpus group were asked about their perceptions regarding the parallel corpus. The control group was not asked since they did not receive any intervention.

According to the results, both the CAT and CAT+Corpus groups thought that including Arabic in the learning activities was helpful. Additionally, some students discussed how crucial it was to include Arabic sentences in the task. Participants from both groups emphasised the importance of being able to compare English and Arabic sentences. By looking at the students' feedback, it could be concluded that L2 learners found L1 useful and significant when learning L2 vocabulary. The CAT+Corpus group noted that utilising the parallel corpus led to positive results. They reported that they made use of words surrounding the target items. This somewhat contradicts the quantitative results discussed in section 6.2.1.4, which indicated that using the bilingual corpus had more of a negative impact on the CAT+Corpus group's learning and scores Nevertheless, feedback from the follow-up interviews confirms that they found using a parallel corpus with CAT instruction useful in learning L2 vocabulary. For example, some of their comments referred to the usefulness of having access to the corpus in both L1 and L2, which, they argued, helped them notice words preceding and proceeding the target items (see section 4.7.4, chapter four).

6.3 Limitations of the Study

In relation to the present study, several limitations have emerged. One important issue is the sample, which included only Saudi female EFL students in their first year of study at a Saudi university. At the time the research was conducted, the preparatory year in the English Institution at the Saudi university had shorter terms than the other departments. The English institution had a six-week term, which was considerably shorter than the full term of other departments. This resulted in some methodological issues which will be discussed below.

Initially, the present study intended to adopt a quasi-experimental design in classes of six weeks, which included an introductory week in the first week and examinations in the last week. Therefore, the total number of weeks for the intervention was five weeks. During the third week, some of the participants started to attend other exams in other departments and therefore did not attend the second intervention and the delayed-post-test. Others did not attend because of the COVID-19 health-related warnings issued by the university and fears of becoming infected. This impacted the size of the sample making it smaller than the original plan and forced the researcher to adjust the data analysis from what was planned initially (pre-test, first post-test, second post-test, and delayed post-test) to analysing pre-tests and the first post-test only. As a result, the 19 words that were taught were not all used for the data analysis, and only nine target items were used for the current study.

Although the study sample was small, the quantitative and qualitative data together helped provide a comprehensive picture of how to use CAT instruction and how to apply a combined approach of CAT instruction and a parallel corpus. Therefore, the findings can still have an impact and help to inform vocabulary instruction in Saudi Arabian universities. However, there is more

research needed in the area of parallel corpus instruction. The CAT+Corpus quantitative results in this study showed a lack of significant effect of the use of the parallel corpus in comparison to the CAT quantitative results, but a larger sample size could provide a better insight on the benefits of CAT+Corpus and improve external validity. In the future, it is advised to select a larger sample, such as six classes rather than only three. This is because when increasing the sample size there is a potential effect on statistical power. The recommendation to recruit a number of trained teachers in CAT instruction could also assist in expanding the sample.

Additionally, due to strict guidelines from the English Language Institution, there was limited time for the interventions, tests, and TA interviews. There was insufficient time to train the students in CAT+Corpus to use the bilingual corpus i.e., AEPC. AEPC is generally considered user-friendly, which was confirmed in the pilot study, although this could have been due to the fact that it was used on a much smaller population in this preliminary phase. However, when it came to the main study, which was done on a larger scale, with longer and more detailed interventions, some unanticipated challenges emerged such as the need for participants to familiarise themselves with the extensive corpus language, and to feel comfortable navigating the structure and layout of the corpus. As such, the researcher felt that participants would have benefited from more time prior to the study for corpus training. A recommendation to overcome this issue could involve teachers planning ahead and training students in the use of a new online tool before the intervention i.e., during breaktime or as a homework assignment.

Moreover, this study only conducted research on high-mid level students in learning semantic non-transparent vocabulary. The curriculum design for the mid-high level required the students to use adverbials (e.g., as a result) in the form of words and collocations. Based on the

curriculum's requirement, the level was selected for this study. However, the study's relevance to teaching CAT instruction with or without a parallel corpus might also be applied to students in other levels, such as those in advanced courses, provided that the curricular objectives are taken into account.

In addition, researcher positionality as a teacher might be a limitation in this study. However, the fact that the researcher is a faculty member at the university and aimed to explore CAT instruction effectively, means the advantages of teaching the intervention herself outweighed the disadvantages. Among these advantages was that she was competent and very familiar in delivering the curriculum and lessons to her students. Furthermore, her familiarity with CAT instructional research enabled her to ensure the intervention was structured and delivered correctly. English teachers unfamiliar with CAT instruction could have possibly mistaken this approach with a regular translation exercise or may have intentionally used the L1 when teaching the control group. The researcher thus carried out all the interventions for all three groups in order to guarantee that each group received the necessary instruction. The researcher also assured the participants that their score in the study does not have an impact on their studies.

6.4. Implications and Recommendations for Teachers and Institutions

This study outlines the following implications for teachers and institutions. 1) The evidence from this study suggests that using explicit teaching such as CAT instruction for target vocabulary is useful for L2 learning and can be implemented in classrooms by teachers and institutions.

2) Teachers might be advised not to neglect the use of L1 in L2 practices and exercises as it can help in vocabulary learning.

3) In teaching vocabulary, it is recommended that teachers plan their goals and align them with learning outcomes taking into consideration a well-balanced strategy of task involvement and reasonable cognitive load on their students.

4) While taking into account learning goals and outcomes, it is recommended that teachers and institutions give students opportunities to become familiarised with online tools i.e., different parallel corpora, allowing them to explore the target vocabulary in authentic and real contexts. This could mean arrangements made in advance between administrative departments and teachers to have regular visits to computer labs.

5) Findings show that not only did the two treatment groups make significant vocabulary gains, but learners from the control group who experienced incidental learning with no explicit teaching activities, also improved their receptive vocabulary knowledge. Though the control group's gains were less than those of the two treatment groups, the results revealed vocabulary improvement in receptive knowledge. This implies that teachers should take into account incidental and intentional approaches in teaching L2 vocabulary and understand that both approaches complement each other.

6) It is important to provide students with enough time in the classroom to develop their learning. This includes learning new vocabulary or utilising new supplemental tools such as a parallel corpus. Teachers are encouraged to ensure that learners understand the vocabulary explanations

and how to use supplemental tools effectively. Perhaps less advanced students would require more assistance from the teacher or their peers.

7) It is important that teachers are sufficiently knowledgeable and are experts in CAT instruction. It is recommended that teachers have high proficiency in L1 and L2 to teach CAT.

8) Teachers are advised to help students make use of a parallel corpus sufficiently. This is because using a parallel corpus can have additional potential benefits, for example, using authentic language and understanding the contexts of sentences.

6.5 Future Research

1) This study explored students' short-term recall in vocabulary knowledge. It would be useful to conduct a study exploring other students' learning in other Saudi universities and in schools of all ages. By doing this, it may be possible to gain a clearer understanding of how to encourage the spread of CAT instruction and training.

2) It would be useful to conduct a study with different methodology tools. As this study used qualitative and quantitative tools, other techniques, such as stimulated recall interviews or focus groups interviews, might be considered to have more depth for researching L2 vocabulary learning and teaching.

3) Initially, this study intended to include a short-term and long-term analysis of the impact of the teaching interventions on vocabulary learning. However, due to external circumstances, the study focused on the short-term only. It would be useful to have a delayed post-test to understand

CAT impact on long-term learning e.g., a full term or a year. This would give a better insight into the impact of CAT instruction on long-term recall.

4) It would also be interesting to replicate the study with a larger number of participants both in other universities in Saudi Arabia and other countries where English is taught as a second language.

5) Finally, future research could explore the impact of learners' independent reading in the L2, previous familiarity with working with corpora, and vocabulary size on performance in CAT+Corpus based activities.

6.6 Contributions to Knowledge

1) The present study employed a mixed-methods approach to prevent potential methodological limitations. Quantitative analyses were employed to investigate the impact of CAT vocabulary instruction on vocabulary learning, followed by (TA) interviews with the purpose of exploring what learning strategies learners use when completing a fill-in-the-gaps task using the target items given by the teacher. The results of this study will help L2 teachers understand the benefits of CAT instruction in vocabulary knowledge and thus implement CAT in their activities and tasks.

2) The results can help L2 teachers structure their L1 use in the L2 classroom and not only depend on regular translations to teach target items.

3) The results could also play an important role in encouraging educators not to neglect the use of L1 and to understand that exploiting L1 in CAT instruction can improve vocabulary knowledge outcomes.

4) It is possible that this study, with the support of future empirical research, will help teachers to understand how to use a parallel corpus in L2 classes and show learners how to employ different strategies when dealing with vocabulary learning within context.

In conclusion, it is hoped that the findings of the current study will encourage scholars and educators to focus more on vocabulary development through CAT education.

References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199–226. https://doi.org/10.1080/09588220802090246

Abrams, Z. I. (2003). The effect of synchronous and asynchronous CMC on oral performance in German. *The Modern Language Journal, 87*(2), 157-167. https://doi.org/10.1111/1540-4781.00184

Ahmad, J. (2012). Intentional vs. incidental vocabulary learning. *ELT Research Journal,* 1(1)*, 71-79*. https://dergipark.org.tr/tr/download/article-file/63558

Al-Ajmi, H. (2003). Compiling an English-Arabic parallel text corpus. In M. Murata, S. Yamada, & Y. Tono, (Eds.), *Proceedings of Asian Association for Lexicography (ASIALEX), Meikai University* (pp. 51-54)

Albousaif, M. A. (2011). *Factors determining Saudi learners' difficulties in attaining EFL vocabulary* [Unpublished PhD thesis]. University of Newcastle.

Alghamdi, B., Alasmari, N., & Shukri, N. (2020). Academic challenges of female graduate TESOL students in the Saudi context. *International Journal of English Language Education, 8*(1), 20-33. https://doi.org/10.5296/ijele.v8i1.15902

Al-Guayyed, E. H. (1997). The future of English teaching in Saudi Arabia. *Knowledge*, 53(1), 62-69.

Alharbi, R. M. S. (2017). *Acquisition of lexical collocations: A corpus-assisted contrastive analysis and translation approach.* (Publication No.0000 0004 6425 6793) [Doctoral dissertation, Newcastle University]. British Library EThOS

Al-Khairy, M. A. (2013). Saudi English-major undergraduates' academic writing problems: A Taif university perspective. *English Language Teaching*, 6(6) 1-12. https://doi.org/10.5539/elt.v6n6p1

Alotaibi, H. M. (2017). Arabic-English parallel corpus: A new resource for translation training and language teaching. *Arab World English Journal, 8*(3), 319-337. https://doi.org/10.24093/awej/vol8no3.21

Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International Journal of Teaching and Education*, *3*(3), 21-34. https://doi.org/10.20472/TE.2015.3.3.002

Alruwaili, A. K. (2018). *Integrating corpus linguistics in second language vocabulary acquisition* [Unpublished PhD thesis]. University of Nottingham.

Alshammari, S. R. (2020). EFL vocabulary learning strategies used by Saudi Arabia university students. *Advanced Education*, *7*(16), 28–38. https://doi.org/10.20535/2410-8286.202436

Alshehri, A. A. (2021). The mapping process from L2 lexical forms to L1 meaning. *Journal of Language and Linguistic Studies*, *17*(1), 481–492. https://doi.org/10.52462/jlls.31

Aronoff, M. (1976). *Word formation in generative grammar.* Cambridge, MA: T*he MIT Press.*

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101-114. https://doi.org/10.1016/0346-251X(93)90010-E

Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System*, *53*, 84–95. https://doi.org/10.1016/j.system.2015.07.006

Batstone, R. (1996). Noticing. *ELT Journal*, *5*0(3), 273. https://doi.org/10.1093/elt/50.3.273

Becker, J. (1975). The phrasal lexicon. In B. Nash-Webber, & R. Schank (Eds.), *Theoretical issues in natural language processing* (pp. 70-73)*.* Beranek & Newman.

Beinborn, L., & Choenni, R. (2020). Semantic drift in multilingual representations. *Computational Linguistics*, 46(3), 571-603. https://doi.org/10.1162/coli_a_00382

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.

Bjork, R. A., & Kroll, J. F. (2015). Desirable difficulties in vocabulary learning. *American Journal of Psychology*, *128*(2), 241–252. https://doi.org/10.5406/amerjpsyc.128.2.0241

Bjork, R. A. (2018). Being suspicious of the sense of ease and undeterred by the sense of difficulty: Looking back at Schmidt and Bjork (1992). *Perspectives on Psychological Science*, *13*(2), 146–148. https://doi.org/10.1177/1745691617690642

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. https://doi.org/10.1111/lang.12224

Boulton, A., & Pérez-Paredes, P. (2014). Special Issue: Researching uses of corpora for language teaching and learning editorial researching uses of corpora for language teaching and learning. *ReCALL, 26*(2), 121-127. https://doi.org/10.1017/S0958344014000068

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. Routledge.

Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *Recall, 17*(1), 47-64. https://doi.org/10.1017/S0958344005000510

Brewster, J., Ellis, G., & Girard, D. (1992). *The primary English teacher's guide*. Penguin

Brown, D. F. (1974). Advanced vocabulary teaching: The problem of collocation. *RELC Journal, 5*(2), 1–11.

Bryman, A. (2012). *Social research methods*. Oxford University Press.

Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching*. Routledge

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9(2), 111-125.

Chan, T., & Liou, H. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb – noun collocations. *Computer Assisted Language Learning*, *18*(3), 231–251. https://doi.org/10.1080/09588220500185769

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. John Benjamins Publishing.

Chi, A. M., Wong, K. P., & Wong, M. C. (1994). Collocational problems amongst ESL learners: A corpus-based study. *Entering Text. Proceedings of the Joint Seminar on Corpus Linguistics and Lexicology, Guangzhou and Hong Kong*, 157-165. https://repository.vtc.edu.hk/ive-la-sp/17

Chun, D. M. (2016). The role of technology in SLA research. *Language Learning & Technology,* 20(2), 98–115. http://llt.msu.edu/issues/june2016/chun.pdf

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, *25*(3), 301–315. https://doi.org/10.1016/s0346-251x(97)00024-9

Cobb, T. (1999). Applying constructivism: A test for the learner-as-scientist. *Educational Technology Research and Development*, *47*(3), 15–31. https://doi.org/10.1007/bf02299631

Cobb, T., Spada, N. & Zahar, R. (2001). Acquiring vocabulary through reading: effects of frequency and contextual richness. *Canadian Modern Language Review-Revue Canadienne Des Langues Vivantes*, *57*(4), 541–572. https://doi.org/10.3138/cmlr.57.4.541

Cohen, A. D. (2011). *Strategies in Learning and Using a Second Language*. Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates

Cooper, G. (1998). *Research into cognitive load theory and instructional design at UNSW.* University of New South Wales.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Creswell, J. W. (2012). *Educational research: Planning, conducting, evaluating*. Pearson Education.

Cruse, D. (1986). *Lexical semantics*. Cambridge University Press.

Dallmeier, F., Szaro, R., Alfanso, A., & Henderson, A. (2013). Framework for assessment and monitoring of biodiversity. *Sciencedirect*, 545-559. https://doi.org/10.1016/B978-0-12-384719-5.00316-6

Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19(4), 450-466

Dörnyei, Z., Csizér, K., & Németh, N. (2006). *Motivation, language attitudes and globalisation: A Hungarian perspective*. Multilingual Matters.

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.

Doughty, C. (1991). Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13*(4), 431-469. https://doi.org/10.1017/S0272263100010287

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, *16*(2), 227–252. https://doi.org/10.1177/1362168811431377

Ellis, N. C. (2002). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(2), 297–339. https://doi.org/10.1017/S0272263102002140

Ellis, R. (2001). Introduction: Investigating Form-Focused instruction. *Language Learning,* 51(1), 1-46. https://doi.org/10.1111/j.1467-1770.2001.tb00013.x

Fan, M. Y. (2003). Frequency of Use, Perceived Usefulness, and Actual Usefulness of Second Language Vocabulary Strategies: A Study of Hong Kong Learners. *The Modern Language Journal (Boulder, Colo.)*, *87*(2), 222–241. https://doi.org/10.1111/1540-4781.00187

Fender, M. (2003). Word recognition and word integration skills of native Arabic and Japanese-speaking learners of English as a second language. *Applied Psycholinguistics, 24*(2)*,* 289-315. https://doi.org/10.1017/S014271640300016X

Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language, 20*(1), 19-42.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th Kindle ed.). SAGE Publications.

Figueredo, L. (2006). Using the known to chart the unknown: A review of first-language influence on the development of English-as-a-second-language spelling skill. *Reading and Writing, 19*(8), 873-905. https://doi.org/10.1007/s11145-006-9014-1

Flowerdew, J. (2015). Data-driven learning and language learning theories. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15–36). John Benjamins.

Folse, K. S. (2006). The effect of type of written exercise on L2 vocabulary retention. *TESOL Quarterly*, *40*(2), 273-293. https://doi.org/10.2307/40264523

Franke, U., & Weber, R. (2012). At the Papini hotel: On pragmatism in the study of international relations. *European Journal of International Relations, 18*(4), 669-691. https://doi.org/10.1177/1354066111404444

Fraser, C. E. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, *21*(2), 225–241. https://doi.org/10.1017/s0272263199002041

Friedrichs, J., & Kratochwil, F. (2009). On acting and knowing: How pragmatism can advance international relations research and methodology. *International Organization, 63*(4), 701-731. https://doi.org/10.1017/S0020818309990142

Gass, S., & Mackey, A. (2017). *Stimulated recall methodology in applied linguistics and L2 research*. Routledge. https://doi.org/10.4324/9781315813349

Glaser, B. G. (1992). *Basics of grounded theory analysis: emergence versus forcing*. Sociology Press.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Clarendon Press.

Gu, P.Y. and Johnson, R.K. (1996). Vocabulary Learning Strategies and Language Learning Outcomes. *Language Learning*, 46, 643-679. http://dx.doi.org/10.1111/j.1467-1770.1996.tb01355.x

Gu, P. Y. (2018). Validation of an online questionnaire of vocabulary learning strategies for ESL learners. *Studies in Second Language Learning and Teaching*, *8*(2), 325–350. https://doi.org/10.14746/ssllt.2018.8.2.7

Guba, E. G. (1990). The alternative paradigm dialog. In G.E. Guba (Ed.), *The paradigm dialog* (pp. 17-27). Sage Publications.

Haastrup, K. (1991). *Lexical inferencing procedures or talking about words*. Gunter Narr.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *International Review of Applied Linguistics in Language Teaching (IRAL), 40*(2), 69-87. https://doi.org/10.1515/iral.2002.008

Hammersley, M. (1989). Pragmatism. In M. Hammersley (Ed.), *Dilemma qualitative method* (pp. 49-70). Routledge. https://doi.org/10.4324/9780203392904-10

Haynes, M. (2010). [Review of the book *Reading in a second language: Moving from theory to practice,* by W. Grabe]. *Studies in Second Language Acquisition, 32*(4), 648-650. https://doi.org/10.1017/s0272263110000355

Heath, H. F., & Cowley, S. (2004). Developing a grounded theory approach: A comparison of Glaser and Strauss. *International Journal of Nursing Studies*, *41*(2), 141–150. https://doi.org/10.1016/s0020-7489(03)00113-5

Hemchua, S. and Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect, 21*(3), 3-25.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *International Review of Applied Linguistics in Language Teaching (IRAL), 41*(2), 87-106. https://doi.org/10.1515/iral.2003.007

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified Texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.

Hornby, A. S. (1974). *Oxford advanced learners' dictionary*. Oxford University Press.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*(2), 207-223.

Hu, H. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load

hypothesis versus technique feature analysis. *System*, 56, 28–39. https://doi.org/10.1016/j.system.2015.11.001

Hu, M. & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language 13*(1), 403-430.

Huang, S., Willson, V.L., & Eslami, Z. R. (2012). The effects of task involvement load on L2 incidental vocabulary learning: A meta-analytic study. *The Modern Language Journal*, *96*(4), 544–557. https://doi.org/10.1111/j.1540-4781.2012.01394.x

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). MacMillan.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge University Press.

Hulstijn, J. H. (2003). Incidental and Intentional Learning. In *The Handbook of Second Language Acquisition* (pp. 349–381). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470756492.ch12

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal, 80*(3), 327-339. https://doi.org/10.1111/j.1540-4781.1996.tb01614.x

Hulstijn, J. & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558. https://doi.org/10.1111/0023-8333.00164

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

James, W. (1995). *Pragmatism.* Dover Philosophical Classics. (Original work published 1907).

Jenkins, J. R., Stein, M., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, *21*(4), 767–787. https://doi.org/10.3102/00028312021004767

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Jiang, N. (2002). Form–meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *24*(4), 617–637. https://doi.org/10.1017/s0272263102004047

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal*, *88*(3), 416–432. https://doi.org/10.1111/j.0026-7902.2004.00238.x

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge University Press.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research Paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

Kang, E. S. (2015). Promoting L2 vocabulary learning through narrow reading. *RELC Journal*, *46*(2), 165–179. https://doi.org/10.1177/0033688215586236

Keating, G. C. (2008). Task effectiveness and word learning in a second language: The involvement load hypothesis on trial. *Language Teaching Research*, *12*(3), 365–386. https://doi.org/10.1177/1362168808089922

Khoshsima, H., & Eskandari, Z. (2017). Task effectiveness predictors: Technique feature analysis versus involvement load hypothesis. *Iranian Journal of English for Academic Purposes,* 6(2), 50-69. https://dorl.net/dor/20.1001.1.24763187.2017.6.2.5.3

Kim, Y. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, *58*(2), 285–325. https://doi.org/10.1111/j.1467-9922.2008.00442.x

Kroll, J. F., & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–199). Lawrence Erlbaum.

Kurosaki, S. (2012). *An analysis of the knowledge and use of English collocations by French and Japanese learners.* [Doctoral dissertation, University of London]. Royal Holloway Repository. https://repository.royalholloway.ac.uk/items/d65f4c4b-8bac-d738-3735-83b5e5e031a7/8/

Lærd Statistics (2015). Two-way mixed ANOVA using SPSS Statistics. *Statistical tutorials and software guides.* Retrieved from https://statistics.laerd.com/

Lado, R. (1957). *Linguistics across cultures; applied linguistics for language teachers.* University of Michigan Press.

LaFlair, G. T., & Settles, B. (2019). *Duolingo English test: Technical manual*. Duolingo Research Report.

Larsen-Walker, M. (2017). Can data driven learning address L2 writers' habitual errors with English linking adverbials? *System*, *69*, 26–37. https://doi.org/10.1016/j.system.2017.08.005

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS.* Routledge.

Laufer, B. (2000). Task effect in instructed vocabulary learning: The hypothesis of involvement. *Selected papers from AILA '99 Tokyo,* 47-62. Waseda University Press.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *The Canadian Modern Language Review*, 59, 567-587. http://dx.doi.org/10.3138/cmlr.59.4.567

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook, 5*(1), 223 – 250. https://doi.org/10.1075/eurosla.5.11lau

Laufer, B. (2010). Form focused instruction in second language vocabulary learning. In R. Chacón Beltrán, C. Abello-Contesse, M. M. Torreblanca-López, & M. D. López-Jiménez (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 15-27). Multilingual Matters.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202-226. https://doi.org/10.1191/0265532204lt277oa

Laufer, B., & Girsai, N. (2008a). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*(4), 694-716. https://doi.org/10.1093/applin/amn018

Laufer, B., & Girsai, N. (2008b). The use of native language for improving second language vocabulary. In A. Stavans & I. Kupferberg (Eds.), *Studies in language and language Education. Essays in honor of Elite Olshtain* (pp.261-275). The Hebrew University Magnes Press.

Laufer, B. & Hulstijn. J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement, *Applied Linguistics,* 22(1), 1-26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33-51. https://doi.org/10.1177/026553229901600103

Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC Journal 28*(1), 89-108. https://doi.org/10.1177/003368829702800106

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.

Lee, H., Warschauer, M., & Lee, J. G. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, *40*(5), 721–753. https://doi.org/10.1093/applin/amy012

Levelt, W. (1989). *Speaking: From intention to articulation*. MIT Press.

Lewis, M. (1993). *The lexical approach: The state of ELT and way forward*. Language Teaching Publications.

Lewis, M. (1993). *The Lexical Approach: The State of ELT and way forward*. Language Teaching Publications.

Lewis, M. (1997a). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255-270). Cambridge University Press.

Lewis, M. (1997b). *Implementing the lexical approach: Putting theory into practice*. Language Teaching Publications.

Lewis, M. (2000). *Teaching collocation: Further developments in the lexical approach*. Language Teaching Publications.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language 61*(1), 30-44. https://doi.org/10.1006/brln.1997.1876

Linck, J. A., & Weiss, D.J. (2015). Can working memory and inhibitory control predict second language learning in the classroom? *SAGE Open*, 5(4). https://doi.org/10.1177/2158244015607352

Linck, J. A., Osthus, P., Koeth, J. T., & Buntin, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review, 21*, 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lindstromberg, S., & Boers, F. (2008). The mnemonic effect of noticing alliteration in lexical chunks. *Applied Linguistics*, *29*(2), 200–222. https://doi.org/10.1093/applin/amn007

Lindstromberg, S., & Eyckmans, J. (2019). The effect of frequency on learners' ability to recall the forms of deliberately learned L2 multiword expressions. *ITL - International Journal of Applied Linguistics, 170*(1), 161-193. https://doi.org/10.1075/itl.18005.lin

Logan, G., Taylor, S., & Etherton, J. (1996). Attention in the acquisition and expression of automaticity. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(3), 620-638. https://doi.org/10.1037/0278-7393.22.3.620

Long, M. (1988). Instructed interlanguage development. In L. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115–141). Newbury House.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39-52). John Benjamins.

Macis, M., & Schmitt, N. (2017). Not just 'small potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research, 21*(3), 321-340. https://doi.org/10.1177/1362168816645957

Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics, 27*(3), 405-430. https://doi.org/10.1093/applin/ami051

Mackey, A. & Gass, S. (2005). *Second language research: Methodology and design.* Lawrence Erlbaum Associates.

Mahboob, A., & Elyas, T. (2014). English in the kingdom of Saudi Arabia. *World Englishes, 33(*1), 128-142. https://doi.org/10.1111/weng.12073

Malone, J. D. (2018). Incidental vocabulary learning in SLA. *Studies in Second Language Acquisition*, *40*(3), 651–675. https://doi.org/10.1017/s0272263117000341

Maskor, Z. M., & Baharudin, H. (2016). Receptive vocabulary knowledge or productive vocabulary knowledge in writing skill, which one important? *International Journal of Academic Research in Business & Social Sciences*, *6*(11), 261-271. https://doi.org/10.6007/ijarbss/v6-i11/2395

Mattiello, E., & Dressler, W. U. (2018). The morphosemantic transparency/opacity of novel English analogical compounds and compound families. *Studia Anglica Posnaniensia, 53*(1), 67-114. https://doi.org/10.2478/stap-2018-0004

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching & Linguistics: Abstracts, 13*(4), 221-246.

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, *28*(1), 19–30. https://doi.org/10.1016/S0346-251X(99)00058-5

Melka F. (1997). Receptive vs. productive aspects of vocabulary. In Schmitt N., McCarthy M. (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 84–102). Cambridge University Press.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey Bass.

Miles, M. B., & Hubermann, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage Publications.

Ministry of Education. (2021, September). *The Ministry of Education provides students with electronic textbooks for the new curricula through Madrasati.* Saudi Ministry of Education. https://moe.gov.sa/en/mediacenter/MOEnews/Pages/ele-t-2021-872.aspx

Mohajan, D., & Mohajan, H. (2023). Straussian grounded theory: An evolved variant in qualitative research. *Studies in Social Science & Humanities*, *2*(2), 33–40. https://doi.org/10.56397/sssh.2023.02.06

Mondria, J. A. (2003). The effects of inferring, verifying and memorizing on the retention of L2 word meanings: An experimental comparison of the 'meaning inferred method' and 'meaning given method'. *Studies in Second Language Acquisition, 25*(4), 473-499. https://doi.org/10.1017/S0272263103000202

Morton, J. (1979). Word recognition. In J. Morton & J. Marshall (Eds.), *Structures and processes 2* (pp. 106-156). Cornell University Press.

Muqaibal, M. H. (2020). *Evaluating the effectiveness of spaced practice using computer-assisted language learning (CALL) in teaching and learning English vocabulary in the classroom: The case of Oman.* [Doctoral dissertation, University of Reading]. Central Archive at the University of Reading. https://centaur.reading.ac.uk/96634/

Nagy, W. E., Anderson, R. A., & Herman, P. M. (1987). Learning Word Meanings From Context During Normal Reading. *American Educational Research Journal*, *24*(2), 237–270. https://doi.org/10.3102/00028312024002237

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17–38. https://doi.org/10.1080/09588221.2010.520675

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal, 90(3)*, 297-455. https://doi.org/10.1111/j.1540-4781.2006.00431.x

Nassaji, H., & Hu, M. (2012). The relationship between task-induced involvement load and learning words from context. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *50*(1), 69-86. https://doi.org/10.1515/iral-2012-0003

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian*

Modern Language Review, 63(1), 59-82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge university press.

Nation, I. S. P., & Webb, S.A. (2011). *Researching and analyzing vocabulary*. Heinle & Heinle Publishers.

Nawal, A. F. (2018). Cognitive load theory in the context of second language academic writing. *Higher Education Pedagogies*, *3*(1), 385–402. https://doi.org/10.1080/23752696.2018.1513812

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242. https://doi.org/10.1093/applin/24.2.223

Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.

Österman, A. (1997). There compound in the history of English. In M. Rissanen, M. Kytö, & K. Heikkonen, (Eds.), *Grammaticalization at work: Studies of long-term developments in English* (pp.191-276). Mouton De Gruyter.

Paribakht, T. S., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11,* 9–29.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–200). Cambridge University Press.

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381-402. https://doi.org/10.1177/1362168815618428

Pérez-Paredes, P., Sánchez-Tornel, M., Calero, J. M. A., & Jiménez, P. (2011). Tracking learners' actual uses of corpora: Guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, *24*(3), 233–253. https://doi.org/10.1080/09588221.2010.539978

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*(1), 25-36. https://doi.org/10.1037/0022-0663.90.1.25

Plonsky, L. & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 Research. *Language Learning*, 64 (4): 878-912. https://doi.org/10.1111/lang.12079

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish

compound words. *Language and Cognitive Processes, 20*(1-2), 261-290. https://doi.org/10.1080/01690960444000098

Porte, G. K. (2002). *Appraising research in second language learning: A practical approach to critical analysis of quantitative research*. John Benjamins.

Ramachandran, S. D., & Rahim, H. A. (2004). Meaning recall and retention: The impact of the translation method on elementary level learners' vocabulary learning. *RELC Journal 35*(2), 161-178.

Ramadhan, J. M. (2017). *The influence of features of collocations on the collocational knowledge and development of Kurdish high school students*: *A longitudinal study.* [Doctoral dissertation, University of Exeter]. Open Research Exeter. http://hdl.handle.net/10871/30115

Reinhart, A., Evans, C., Luby, A., Orellana, J., Meyer, M., Wieczoreke, J., Elliott, P., Burckhardt, P., & Nugent, R. (2022). Think-aloud interviews: A tool for exploring student statistical reasoning. *Journal of Statistics and Data Science Education, 30*(2), 100-113. https://doi.org/10.1080/26939169.2022.2063209

Reppen, R., & Simpson-Vlach, R. (2020). Corpus linguistics. In N. Schmitt & M. P. H. Rogers (Eds.), *An introduction to applied linguistics* (pp. 90-108). Routledge.

Reynolds, B. L., & Ding, C. (2022). Effects of word-related factors on first and second language English readers' incidental acquisition of vocabulary through reading an authentic novel. *English Teaching: Practice & Critique*, 21(2), 171-191. https://doi.org/10.1108/ETPC-05-2021-0049

Reynolds, B. L. (2015). Action research: Applying a bilingual parallel corpus collocational concordancer to Taiwanese medical school EFL academic writing. *RELC Journal*, *47*(2), 213–227. https://doi.org/10.1177/0033688215619518

Richards, J. C. (2008). *Moving beyond the plateau*: *From intermediate to advanced levels in language learning* Cambridge University Press.

Richards, J., & R. Schmidt. (2002). *Longman dictionary of language teaching and ppplied linguistics*. Pearson Education.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*(2), 283-331. https://doi.org/10.1111/j.1467-1770.1995.tb00441.x

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: A replication and extension of Reber, Walker and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences in instructed language learning* (pp. 211-266). John Benjamins.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*(4), 589-619. https://doi.org/10.1017/S0272263199004039

Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research, 6*(3), 183–222.

Rubin, J. (1975). What the "Good language learner" can teach us. *TESOL Quarterly*, 9, 41-51.

Rubin, J. (1987). Learner strategies: Theoretical assumptions, research history and typology. In A. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 15–30). Prentice-Hall International

Ruiz, S., Rebuschat, P., & Meurers, D. (2021). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research*, *25*(4), 510–539. https://doi.org/10.1177/1362168819872859

Saigh, K., & Schmitt, N. (2012). Difficulties with vocabulary word form: The case of Arabic ESL learners. *System, 40*(1), 24-36. https://doi.org/10.1016/j.system.2012.01.005

Saldaña, J. (2009). T*he coding manual for qualitative researchers.* Sage Publications.

Salhi, H. (2013). Investigating the complementary polysemy and the Arabic translations of the noun destruction in EAPCOUNT. *Meta: Translators' Journal*, *58*(1), 227–246.

Sangawa, K. H. (2014). The learner as lexicographer: Using monolingual and bilingual corpora to deepen vocabulary knowledge. *Acta Linguistica Asiatica*, *4*(2), 53–65. https://doi.org/10.4312/ala.4.2.53-65

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. W. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165-209). Academic Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp.1-63). Foreign Language Resource Center.

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In

W. M. Chan, S. Chi, K. N. Chin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker (Eds.), *Proceedings of CLaSIC 2010* (pp. 721-737). National University of Singapore, Centre for Language Studies.

Schmidt, R. W. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. Bhatt, & I. Walker (Eds.), *Perspectives on individual characteristics and foreign language education* (pp. 27–50). Mouton de Gruyter.

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt and M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy*. Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, *52*(2), 261–274. https://doi.org/10.1017/S0261444819000053

Schmitt, N., & McCarthy, M. (Eds.). (1997). *Vocabulary: Description, acquisition and pedagogy*. Cambridge university press.

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*(2), 419–440. https://doi.org/10.1006/jmla.2001.2813

Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in Education and the Social Sciences* (4th ed.). Teachers College Press.

Silver-Pacuilla, H., Brown, S., Overton, C., & Stewart, A. (2011). *Assistive technology research matters.* American Institutes for Research.

Sinclair, J. M. (1987). Collocation: A progress report. In R. Steele, & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday Vol. 2* (pp. 319-331). John Benjamins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Singleton, D. (1999). *Exploring the second language lexicon*. Cambridge University Press.

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *English Language Teaching Journal, 64*(3), 253-260. https://doi.org/10.1093/elt/ccp059

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*(2), 139–152. https://doi.org/10.1080/09571730802389975

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition, 28*(3), 387-422. https://doi.org/10.1017/S0272263106060177

Swain, M. (1995). Three functions of output in second language. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson* (pp. 125-144). Oxford University Press.

Swain, M. (2005). The output hypothesis: theory and research. In Hinkel, E. (Ed.) *Handbook of research in second language teaching and learning*. (Vol I. pp. 471- 483). Routledge.

Swan, M. (1997). The influence of the mother tongue on second language vocabulary acquisition and use. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp.156-180). Cambridge University Press.

Sweller, J. (2017). Cognitive load theory and teaching English as a second language to adult learners. *Contact Magazine*, 43(1), 10-14.

Tang, E., & Nesi, H. (2003). Teaching vocabulary in two Chinese classrooms: Schoolchildren's exposure to English words in Hong Kong and Guangzhou. *Language Teaching Research*, 7(1), 65–97. https://doi.org/10.1191/1362168803lr113oa

Tang, C., & Treffers-Daller, J. (2016). Assessing incidental vocabulary learning by Chinese EFL learners: A test of the involvement load hypothesis. In G. Yu, & Y. Jin (Eds.), *Assessing Chinese Learners of English* (pp. 121–149). Palgrave Macmillan. https://doi.org/10.1057/9781137449788_7

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Sage Publications.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage Publications.

Thorndike, E. & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Columbia University.

Thurstun, J. & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, *17*(3), 267-280. https://doi.org/10.1016/S0889-4906(97)00013-6

Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. *Proceedings of the 8th*

International Conference on Language Resources and Evaluation (LREC '12), European Language Research Association, 2214–2218. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf

Uchihara, T., Eguchi, M., & Clenton, J. (2022). The contribution of guessing from context and dictionary use to receptive and productive vocabulary knowledge: A structural equation modeling approach. *Language Teaching Research*, 1-24. https://doi.org/10.1177/13621688221122138

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343

University of Pennsylvania (2023). Linguistic Data Consortiums http://catalog.ldc.upenn.edu

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52. https://doi.org/10.1017/S0272263105050023

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46-65. https://doi.org/10.1093/applin/aml048

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30*(1), 79-95. https://doi.org/10.1017/s0272263108080042

Webb, S., & Chang, A. M. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review- La Revue canadienne des langues vivantes*, *68*(3), 267–290. https://doi.org/10.3138/cmlr.1204.1

Webb, S. & Chang, A. C. S. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*(6), 667-686. https://doi.org/10.1177/1362168814559800

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, *63*(1), 91-120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Weinreich, U. (1953). *Languages in Contact.* Mouton Publishers.

West, M. (1953). *A general service list of English words*. Longman.

Westbrook, C., Baker, L., & Sowton, C. (2019). *Unlock level 3: Reading, writing and critical thinking*. Cambridge University Press

Wong, T., & Lee, J. S. Y. (2016). Corpus-based learning of Cantonese for Mandarin speakers. *ReCALL, 28*(2), 187-206. https://doi.org/10.1017/S0958344015000257

Wray, A., Bell, H., & Jones, K. (2016). How native and non-native speakers of English interpret unfamiliar formulaic sequence. *European Journal of English Studies*, *20*(1), 47-63. https://doi.org/10.1080/13825577.2015.1136163

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1-28. https://doi.org/10.1016/S0271-5309(99)00015-4

Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, *71*(2), 487–536. https://doi.org/10.1111/lang.12444

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System*, *70*, 38–49. https://doi.org/10.1016/j.system.2017.09.012

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology, 10*(3), 85-101. http://dx.doi.org/10125/44076

Youssef, A. M., & Omar, M. (2020). The effectiveness of corpus-based approach on vocabulary learning gains and retention in Saudi tertiary EFL context. *Journal of English Language Teaching and Applied Linguistics*, *2*(3), 1–15. https://doi.org/10.32996/jeltal.2020.2.3.1

Zhang, P. (2018). *Comparing different types of EFL vocabulary instruction for Chinese senior secondary school learners of English* (Publication No. 0000 0004 7233 5934) [Doctoral dissertation, University of Reading]. British Library EThOS.

Appendices

Appendix A – Tasks, Tests and Intervention Material

Appendix A.1.

MAIN STUDY

The Receptive Tests (translating the target vocabulary L2 to L1) for pre-tests & delayed tests on 19 target items.

Translate the following sentences from English to Arabic. Make sure you translate the underlined words.

ترجمي الجمل الآتية مع التأكد من ترجمة الكلمة/الكلمات التي تحتها خط.

1) It's difficult to choose which car to buy, but in the end it boils down to the cost.

2) If your flight is late, the bottom line is that you will miss your connected flight.

3) The man was very tired; nevertheless, he went for a run.

4) We should go to the concert, seeing that we've already paid for the tickets.

5) We haven't had any difficulties thus far.

6) Khalid walked around with the pot on his head, thereby causing him to bump into a wall.

7) The upshot of a daily exercise routine is that you increase and strengthen muscles.

8) I didn't attend the meeting owing to a headache.

9) He is not very hopeful about the outcome of the interview.

10) I expect you to follow the university's rules henceforth.

11) I think the price is fair, but on the other hand, I really can't afford to spend that much money.

12) I am required, in so far as I am able, to write about the world as I find it to be.

13) The committee reviewed the papers and thereupon decided to accept the idea.

14) The aftermath of the explosion was horrific and horrible.

15) You couldn't trust her to look after your dog, let alone your child.

16) The book is too long but, nonetheless informative and entertaining.

17) Build good habits over months and years and you will find many benefits in the long run.

18) He's got a house in London, not to mention the villa in Spain.

19) Michael can hardly boil an egg, much less cook dinner.

The Productive Tests (translating the target vocabulary L1 to L2) for pre-tests & delayed tests on 19 target items.

Please provide the most appropriate translation(s) for the following underlined words and phrases. I understand some of these questions can be difficult just please try your best. Thank you.

الرجاء كتابة الترجمة المناسبة للكلمات الى تحتها خط، بعض الاجابات تتكوّن من كلمة او كلمتان او ٣ كلمات. قد تثير بعض الأسئلة القليل من الارتباك، ارجو منكم حلها بقدر الامكان. وشكرا.

١) إذا رحلتك تأخرت، الخلاصة ان الرحلة المتصلة ستفوتك.
The b_____ _____

٢) يجب ان نذهب الى الحفل الموسيقي بما اننا دفعنا ثمن التذاكر.
Se_____ _____

٣) خالد تجول وعلى رأسه وعاء ، وبالتالي تسبب في اصطدامه بجدار.
T_____

٤) لم أحضر الاجتماع بسبب الصداع.
Ow_____ _____

٥) أتوقع منك أن تتبع قواعد الجامعة من الآن فصاعدا.
He_____

٦) أنا مضطر ، بقدر ما أستطيع ، أن أكتب عن العالم كما أجده.
In so _____ _____

٧) كانت آثار الانفجار مروعة ورهيبة.
The af_____

٨) الكتاب كان طويلا جدا، رغم ذلك غني بالمعلومات وممتع.

Non_____

٩) لديه بيت في لندن، ناهيك عن فيلا في اسبانيا

Not ____ _____

١٠) بالكاد يستطيع مايكل غليان البيض ،وبالأخص طهي الطعام.
Mu____ _____

١١) ابني عادات جيدة على مدى أشهر وسنوات ، وسوف تجد العديد من الفوائد على المدى البعيد.
In the _____ _____

١٢)لا يمكن أن تثق بها لرعاية كلبك، ناهيك عن طفلك.
L_____ _____

١٣) استعرضت اللجنة الأوراق وقررت عندئذ قبول الفكرة.
T_____

١٤) أعتقد أن السعر عادل، لكن من ناحية أخرى لا أستطيع حقا دفع هذا القدر من المال .
On the o_____ _____

١٥) إنه ليس متفائلا جدا بنتيجة المقابلة.
The out_____

١٦) إن نتيجة التمرين اليومي الروتيني هي زيادة العضلات وتقويتها.
The up_____

١٧) لم نواجه اي صعوبات حتى الآن.
Th_____ _____

١٨) لقد كان الرجل متعبا جدا، ومع ذلك ذهب للجري.
Nev_____

١٩) من الصعب اختيار السيارة لشرائها، ولكن في النهاية تتلخص في التكلفة.
It bo_____ _____

Appendix 2.A

Intervention 1 for CAT Group:

Q: Match the Arabic vocabulary with the correct English word or collocation in the sentences.
Intervention 1

ناهيك عن، نتيجة، من وجهة نظر أخرى، النتيجة، يتلخص، على المدى البعيد، حتى الآن، على الرغم، عندئذ

1) Also, it is important to note the area of the face or body at which one directs one's gaze, as this affects **the outcome** of a deal.

2) A daily dose of challenge **on the other hand** keeps you alive, fresh and moving.

3) **The upshot** was that he didn't really listen to the rest of our presentation and we had to work hard to rebuild credibility.
4) **It all boils down** to the reader's ability to decode a person's reactions to statements made and to questions asked, and by information gathered from simple observation about a person's appearance.

5) It is better to let students see that you know what they are doing and that you have higher expectations for them. You will get huge results and rewards **in the long run.**

6) The natural 11-year cycle is currently approaching its peak, but **thus far** it has been the weakest solar maximum in a century. This could help to explain both the hiatus and the discrepancy in the model simulations, which include a higher solar output than Earth has experienced since 2000.

7) Filled with her own anxieties, she **nevertheless** watched him with affectionate pride, for Gerald was an excellent horseman.

8) However, after discussing the matter further with a CostPlus representative, it becomes apparent that the actual costs of installation would exceed £4,000. Philip **thereupon** withdraws his order.

9) According to Naomi Quinn and Dorothy Holland, "Largely tacit and unexamined, [cultural] models embed a view of 'what is' and 'what it means' that seems wholly natural__a matter of course. Alternative views are not even recognized**, let alone** considered".

let alone, thereupon, nevertheless, thus far, in the long run, the upshot, the outcome, it all boils down, on the other hand.

١.و من المهم أيضاً ملاحظة منطقة الوجه التي يوجه الشخص نظره إليها حيث أنها تؤثر على **نتيجة** التعامل.

٢. من وجهة نظر أخرى، جرعة يومية من التحدي كفيلة بإبقائك حياً و منتعشاً و مستمراً في طريق.

٣. والنتيجة هي أنه لم يستمع حقاً إلى بقية عرضنا وعملنا جاهدين على إعادة بناء المصداقية.

٤. ويتلخص كل ذلك في قدرة القارئ على حل شفرة ردات فعل شخص ما حول التصريحات المدلاة، والأسئلة المطروحة، والمعلومات المجموعة من ملاحظة بسيطة لمظهر الشخص.

٥. ومن الأفضل أن يرى الطلاب أنك تعلم ما يفعلونه ولديك توقعات عالية منهم، وبذلك سوف تحصل على مكافأة ونتائج ضخمة **على المدى البعيد.**

٦. وتصل دورة السنوات الإحدى عشرة إلى قمتها حاليًا، لكنها **حتى الآن** مثلت أضعف حد أعلى من الإشعاع الشمسي خلال قرن كامل. وهذا يمكن أن يسهم في تفسير كلٍّ من تباطؤ الحرارة في الغلاف الجوي، والاختلافات في نماذج المحاكاة التي تفترض إشعاعًا شمسيًا أعلى من الذي تعرضت له الأرض منذ عام 2000.

٧. وعلى الرغم مما اعتراها من هواجس وأفكار مقلقة, فقد راحت تراقبه بإعتزاز مفعم بالحب. فقد كان جيرالد فارساً ماهراً.

٨. ومع ذلك وبعد مناقشة المسألة مع ممثل كوست بلس، اتضح أنَّ التكاليف الفعلية للتركيب تتجاوز (4.000) جنيه إسترليني؛ **عندئذٍ** انسحب فيليب

٩.هووفقا لنعومي كوين ودوروثي هولند فإن "النماذج [الثقافية] الضمنية وغير المدققة, بطبيعة الحال, ترسخ فكرة عامة عن "ماهية" الشيء و "ماهية معناه" التي تبدو طبيعية بالكامل. و لا يعترف بالأفكار البديلة **ناهيك عن** عدم أخذها بعين الإعتبار" (1987:11)

Intervention 2 for CAT Group:

(نظرا ، نتيجة)، بينما، (رأى بأن ، حين وجد)، مع ذلك، وبالتالي، الخلاصة، منذ ذلك الحين، آ ثار الكارثة، الى حد ما ، ناهيك عن ,فضلا عن

1) Adequate sleep is necessary for the proper balance of the hormones that affect appetite and fat storage, **not to mention** the fact that being tired may worsen depression or other mood disorders and sap your motivation.

2) One report suggests additional barriers: greater financial needs and being more location-bound **owing to** finances or family obligations.

   This was previously difficult **owing to** the small size and fragility of diamond-cell samples.

3) Then **seeing that** her feelings had not changed, he went on, I thought you would do this for me, Lizzie. But I can soon put things right.

   But **seeing that** his young relative looked worried, he added a few soothing words, and thus wiped out any impression that his wife might have made.

4) These conditions do not exist in the same combination elsewhere in the world, but the development of shale gas and oil is **nonetheless** beginning to happen, albeit at a slower pace.

5) Water buoyancy reduces weight-bearing stress on joints and **thereby l**essens the risk for injuries.

6) **The bottom line** is to pay attention to your physical state when you're making a decision.

7) He is then immediately taken from his proud yet sorrowing parents and adopted by some childless Equilateral, who is bound by oath never to permit the child **henceforth** to enter his former home or so much as to look upon his relations again, for fear lest the freshly developed organism may, by force of unconscious imitation, fall back again into his hereditary level.

8) Even after the wind has abated, **the aftermath** in the countryside may take some time and a great deal of care to recover.

9) Yes, interviews are a bit like exams **in so far as** that you're asked a number of questions to which you need to respond intelligently, but there the similarities end.

10) He loved his wife and never raised his voice to her, **much less** his hand. He valued her counsel and found genuine pleasure in her companionship.

Owing to, seeing that, nonetheless, thereby, the bottom line, henceforth, the aftermath, in so far as, not to mention, much less

١.فالنوم الكافي ضروري للتوازن المناسب للهرمونات التي تؤثر على الشهية وتخزين الدهون، **ناهيك عن** حقيقة عندما تكون متعباً قد يتفاقم الاكتئاب أو اضطرابات المزاج الأخرى وتستنزف تحفيزك.

٢.ويشير أحد التقارير إلى وجود عوائق إضافية، تتمثل في الاحتياجات المالية الكبيرة، وفي كونهم مرتبطين بمناطقهم بشكل أكبر، **نتيجة** الْتِزامات مالية أو أُسَرِيّة

كان هذا صعبًا في السابق، **نظرًا** لصغر حجم وهشاشة عينات خلية ألماس.

٣..عندئذ, وبعد أن **رأى بأن** مشاعرها لم تتغير, تابع: فكرت بأنك ستفعلين هذا من أجلي ياليزي لكن سرعان ما سأتمكن من وضع الأمور في نصابها.

لكنه **حين وجد** قريبه الشاب مضطرباً، أضاف بضع كلمات مهدئة، وبذلك أزال أي ضيق تسببت به زوجته .

٤.وهذه الظروف لا توجد بنفس الشكل في أي مكان آخر في العالم لكن تطوير غاز الإردواز والنفط **مع ذلك** بدأ بالحدوث وإن كان بوتيرة أبطأ

٥.الطفو على الماء يقلل من ضغط الأوزان على المفاصل **وبالتالي** يقلل من خطورة الإصابة بالجروح.

٦.والخلاصة هنا أنه يجب عليك أن تنتبه لحالتك الجسدية عندما تتخذ قراراً ما .

٧. ويأخذونه على الفور من أبويه اللذين تتنازعهما مشاعر الفخر به والحزن لفراقه، ويتبناه أحد المضلعات المنتظمة التي لم ترزق بذرية، ويتعهد بألا يسمح للطفل **منذ ذلك الحين** بزيارة بيته السابق أو حتى إلقاء نظرة على أقربائه مرة ثانية، خشية أن يرتد — بفعل المحاكاة غير الواعية — إلى مستواه الموروث وهو لم يزل بَعدُ حديث عهد بالتطور.

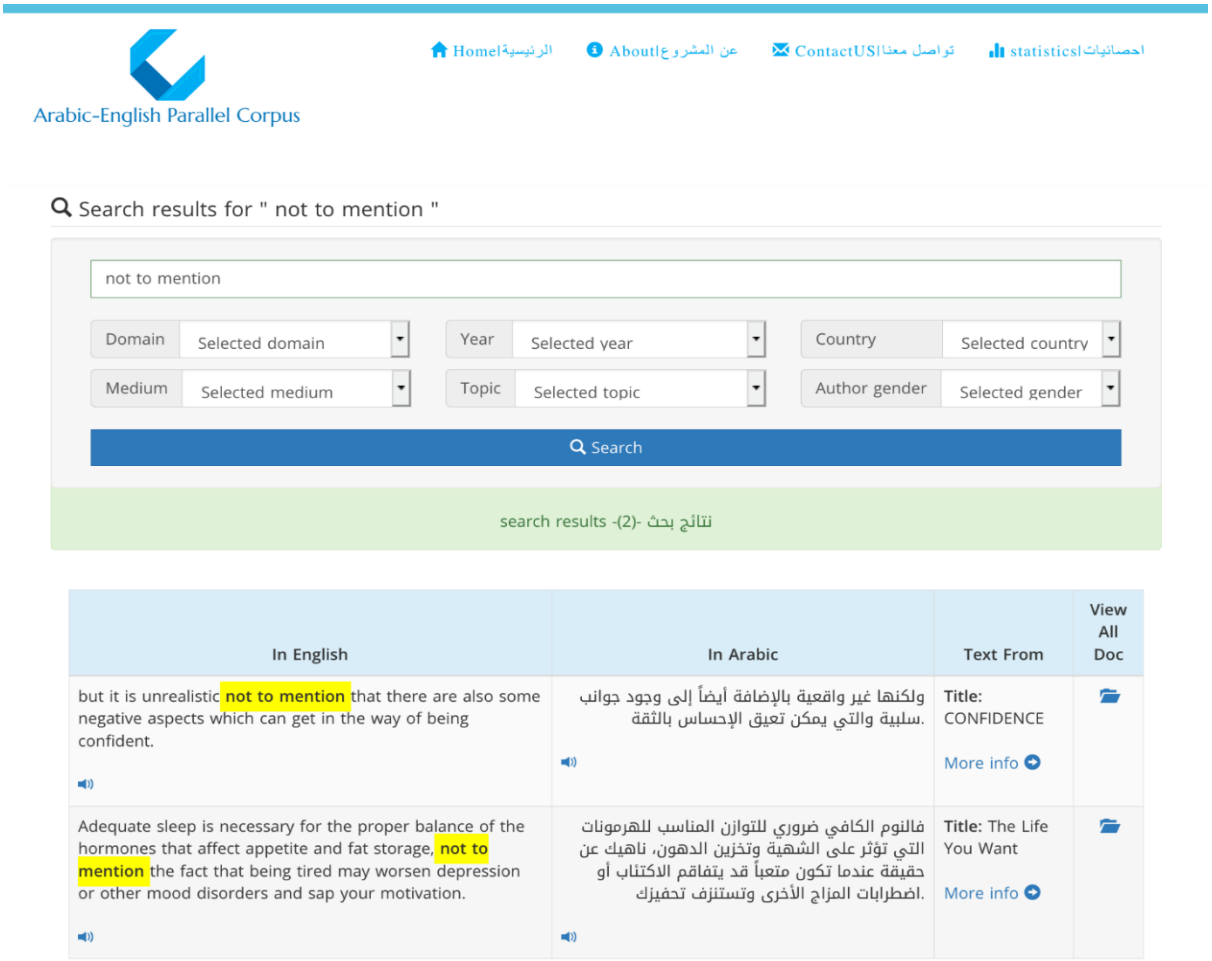
٨. حتى بعدما تخف الرياح، فربما تتطلب **آثار الكارثة** في الريف بعض الوقت وقسطًا كبيرًا من العناية للتعافي.

٩.هذا صحيح، فالمقابلات الشخصية تشبه الاختبارات **إلى حدٍ ما** حيث يتم سؤالك عدداً من الأسئلة التي تحتاج إلى إجابتك بذكاء. ولكن عند هذا ينتهي وجه التشابه.

١٠) . كان يحب زوجته ولم يرفع صوته عليها قط، **فضلاً** عن يده، وكان يقدر مشورتها ويجد مسرة عظيمة في صحبتها.

Intervention for CAT+ Corpus:

This is similar to the pilot study. It is worth noting that AEPC is no longer running.



In English	In Arabic
A daily does of challenge on the other hand keeps you alive, fresh and moving	من وجهة نظر أخرى، جرعة يومية من التحدي كفيلة بإبقائك حياً و منتعشاً و مستمراً في طريقك

Appendix A.3.

MAIN STUDY
Think-Aloud Interview and Fill-In-The-Gaps Task (a test consisting of 19 exercises with a gap in each sentence; each exercise has three multiple-choice options).

THINK ALOUD Guidelines:

First thing **before** the actual interview is to:
1) reiterate the reason the participants are being asked to THINK aloud
For example:
In this experiment I am interested in what you think about when you complete these tasks. In order to find out, I am going to ask you to THINK ALOUD as you work through these questions.

This experiment is designed to how second language learners think about while they solve vocabulary items.

We ask you to Talk ALOUD as you go through the questions. What we mean by "Talk Aloud" is that I want you to say out loud everything that you would say to yourself silently while you think. Just act as if you were alone the room speaking to your elf. Don't try to explain your thoughts.

Fill in the Gap Task
Choose the best answer and fill in the gap:

1) The costs, _____ the risks, of moving the satellite to and from space would have been difficult.
   b) not to mention
   c) owing to
   d) thereby

2) I know about all the problems, but what is _____? What will happen?
   a) on the other hand
   b) the bottom line
   c) in so far as

3) There is not much we can do, _____they've already made their decision.
   a) in the long run
   b) the outcome
   c) seeing that

4) She forgot to tie her shoes, _____ tripping and falling down the stairs.
   e) thereby
   f) in the long run
   g) seeing that

5) The baby can't even walk, _____ run.
   a) on the other hand
   b) much less
   c) thereby

6) _____ the heavy traffic, he was late.
   a) The upshot
   b) Seeing that
   c) Owing to

7) I promise not to lie to you _____
   a) in so far as
   b) henceforth
   c) much less

8) This is the truth_____ I know it.
   a) in so far as
   b) henceforth
   c) seeing that

9) In _____of the hurricane and the strong wind, many people's homes were destroyed.
   a) the bottom line
   b) thus far
   c) the aftermath

10) Though she is a fool, I like spending time with her _____.
   a) Owing to
   b) nonetheless
   c) not to mention

11) Being alone can seem scary and uncomfortable at first but it will strengthen you_____.
   a) in the long run
   b) let alone
   c) seeing that

12) There isn't room for us, _____ any guests.
   a) let alone
   b) henceforth
   c) thereby

13) The police arrived. _____ the thieves ran away.
   a) Thereupon
   b) In the long run
   c) The bottom line

14) I want to go to the party, but _____I ought to be studying.
a) owing to
b) let alone
c) on the other hand

15) _____ of their decision is still unknown.
a) Let alone
b) The outcome
c) In so far as

16) _____ of reading many books is that you will have a lot of information.
a) The upshot
b) Nevertheless
c) Nonetheless

17) We haven't had any problem_____
a) henceforth
b) thus far
c) thereby

18) It all _____ to money in the end.
a) the bottom line
b) let alone
c) boils down

19) The mother was very tired;_____, she cooked dinner for her kids.
a) nevertheless
b) in the long run
c) owing to

Appendix A.4:

MAIN STUDY
A.4.1 Comprehension Questions for Reading Task (CAT+Corpus, CAT+ Control Groups)
For Reading Passage ONE: OUR CHANGING PLANET:

1. **Read the article and number the main ideas in the order which they appear.**

   solution to the problem_____

   changing ecosystems_____

   melting glaciers_____

   causes of climate change_____

2. **Read the article again then complete the sentences using the words and phrases in the box.**

CO2 Levels	coral reefs	extinction	farming
global temperatures	mangrove forests	sea levels	

   1) Over the last century,_____ have gun up by 0.75 Celsius.
   2) Global increases in temperature could cause the_____ of 30% of land species.
   3) _____ could rise by about 30.5 cm by the end of the century.
   4) Recently, over a third of the world's_____ have been destroyed.
   5) Twenty percent of the Earth's _____ have been lost in the last few decades.
   6) Twenty-five percent of the land on earth is used for_____.
   7) _____ are at their highest for 800,000 years.

3. **Read the article again and complete the table with supporting details.**

1. Country where the Upsala glacier is located	
2. Name of sea route through the Arctic ice	
3.Why forests are being cut down all over the world	
4.Medical problem caused by pollen	
5.Main chemicals responsible for climate change	
6.Human activities that reduce the amount of oxygen in the atmosphere	
7.What we should do to reduce the amount of CO2 in the atmosphere	

For Reading Passage TWO: THE CAUSES OF FORESTATION:

1. **Read the essay and complete the summary using the words in the box.**

animals	crops	decade	deforestation	effects	environment	erosion	habitats	protected	warming

The essay discusses the human causes of_____and the_____on the environment. Trees are removed for grazing of_____ and growing _____ like soy and palm oil. Farmers traditionally leave the land for a_____ before using it, but if the land is constantly reused, it results in _____ of the soil. Deforestation allows CO2 to escape into the atmosphere and contribute to global _____. It also affects bio diversity because it leads to the loss of _____. Government should make sure forests are_____ from logging. Otherwise, deforestation will have terrible consequences for the _____.

2. **Match the words in each box to make collocations about the environment. Sometimes more than one collocation is possible**.

Carbon climate environmental greenhouse natural power tropical	change dioxide gas group plant rain forest resources

3. Complete the sentences with the correct form of the collocations from Exercise 2.

1 In my town, about 70% of the electricity comes from a _____which uses coal for energy.

2 Carbon dioxide and methane are examples of _____.

3 Almost all scientists these days agree that _____ is happening and is a serious threat to our planet.

4 Trees absorb _____ and give off oxygen.

5 All over the world, _____ are working to educate people about the dangers of deforestation and habitat destruction.

6 Fresh water is the most precious _____ on Earth.

7 Thousands of unique plants, animals, birds and insects live in the _____South America and Southeast Asia

A.4.2 ADDIDIONTAL Comprehension Questions for Reading Task (Control Group ONLY)

For Reading Passage ONE: THE CAUSES OF FORESTATION:

4. Fill in the blank:

Absorb, construction, destruction, effect, farming, logging, rainforest

1) Clothes made from plants, like cotton or bamboo, _____water more easily than man-made materials like polyester.

2) _____has been my family's occupation since my grandfather bought his first cow 75 years ago.

3) _____ hurts native people because it destroys the forest that provides them with food, shelter, and medicine.

4) The Amazon_____ in south America received 200 to 600 cm of rain every year.

5) Because of heavy snow, the_____of the new road stopped for more than two months.

6) Sunburn is just one of the harmful_____ of too much sun on sensitive skin.

7) In 2017, Hurricane Harvey cause serious_____in the US state of Texas and killed more than two months.

5. Work with a partner. Answer the following questions:

1) What does the writer mean by the phrases ***industrial logging*** and ***commercial farming***?
2) What will happen if the Amazon rainforest disappears?
3) Why does deforestation reduce future sources of food and medicine?

6. Read the essay again and correct the factual mistakes in the sentences.

1) In Indonesia, trees are cut down to make way for olive plantations.

2) Farmers can graze animals on their land for 10 years.

3) The rain forest of the Amazon cover an area 2.5 times the size of the UK.

4) Government should plant more trees to absorb oxygen.

5) Deforestation protects future sources of food and medicine.

6) Small-scale deforestation will have disastrous effects on the environment.

For Reading Passage TWO:

4) Work with a partner. Discuss the questions.

1 What was the author's main purpose in writing this article?
- a) to inform the reader about the causes and effects of global warming
- b) to describe changes in mangrove forests and coral reefs
- c) to persuade people to help reduce CO2 levels by using renewable energy

2) Who is the intended audience for this article?
- a) advanced science students
- b) general adult leaders
- c) university professors

5) Work with a partner. Discuss the questions and write the answers in a paragraph for each question.

1) Are there any advantages to the melting of the glaciers in the Northwest Passage? Give reasons for your answer.

2) What are some possible disadvantages of using renewable energy like solar energy or wind power?

Appendix A.5.6:

Reading Passage for Pilot Study 1 + Intervention 1 (CAT+Corpus)

A.5.6.1 Reading Passage: From *Unlock Level 3* (Westbrook et al., 2019)

Our Changing Planet

> The Upsala Glacier in Argentina Used to be one of the biggest glaciers in South America. In 1928, it was covered in ice and snow, but now the glacier is melting at an annual rate of about 200 m, so the area is covered in water this is evidence of global warming.

In the last 100 years, the global temperature has gone up by around 0.75 degrees °C.
This may not sound much, but such a small increase is causing sea levels to rise threatening the habitat of many species of plants and animals. An increase of 2 °C in global temperatures could result in extinction for 30% of the world's land species.

The Northwest Passage is a sea route, which runs along the northern coast of Canada between the Atlantic and the Pacific Oceans. In the past, it was often difficult to use because the waters were frozen; however**,** increasing temperatures and the subsequent deglaciation have made it easier for ships to travel through this route. The trouble is that the melting of the ice is leading to loss of habitat for the polar bears and other species that live in this area.

Seeing that sea levels in the UK have increased by around 10 cm in the last 100 years, experts predict the global sea levels could rise by the 30.5-122 cm by the end of the century. Consequently, some areas that were land a few hundred years ago are now under water, and many low-lying islands maybe under water in the future.

In the aftermath of the climate change, the world's ecosystems are also changing faster than ever before. More than one-third of the world's mangrove forests and around 20% of the world's coral reefs have been destroyed in the last few decades. Forests are being cut down to provide land for food because the human population is growing at such a rapid rate. Approximately a quarter of the land on Earth is now used for growing food. Also the higher temperatures and higher levels of carbon dioxide in the atmosphere causes plants to produce more pollen **thereby** leading to more cases of asthma, a medical condition which makes it hard to breathe.

What is causing climate change? The main cause of climate change is the huge amount of greenhouse gases, such as methane and carbon dioxide (CO2), in the atmosphere, but the reason for this is the world's population- you and me. As the population increases, more land is needed to provide food and energy. Burning fossil fuels for heating, lighting, transport, electricity or manufacturing produces CO2**. Not to mention** that humans breathe out CO2 and trees 'breathe in' CO2 and produce oxygen, so by cutting down trees, we are increasing the amount of CO2 in the atmosphere and reducing the amount of oxygen. **Owing to** these human activities, CO2 levels are now at their highest in 800,000 years.

The biggest challenge we all face is to prevent further environmental disasters. **Nonetheless,** we must do something before it's too late. We are required, **in so far as** we can, to reduce the amount of CO2 in the atmosphere. We are not reducing our use of fossil fuels enough, **much less** finding alternatives for them. **Henceforth**, we need to start using renewable energy. We can get enough energy from renewable fuels like solar energy, hydroelectric energy or wind power, to be able to stop using fossil fuels completely. **The bottom line** is we need to sign a petition to get governments to take action before it's too late!

A.5.6.2 Reading Comprehension Questions from *Unlock Level 3* (Westbrook et al., 2019)

4. Read the article and number the main ideas in the order which they appear.

solution to the problem_____

changing ecosystems_____

melting glaciers_____

causes of climate change_____

5. Read the article again then complete the sentences using the words and phrases in the box

CO2 Levels	coral reefs	extinction	farming
global temperatures	mangrove forests	sea levels	

8) Over the last century,_____ have gun up by 0.75 Celsius.

9) Global increases in temperature could cause the_____ of 30% of land species.

10) _____ could rise by about 30.5 cm by the end of the century.

11) Recently, over a third of the world's_____ have been destroyed.

12) Twenty percent of the Earth's _____ have been lost in the last few decades.

13) Twenty-five percent of the land on earth is used for_____.

14) _____ are at their highest for 800,000 years.

6. Read the article again and complete the table to supporting details.

1. Country where the Upsala glacier is located	
2. Name of C route through the Arctic ice	
3.Why forests are being cut down all over the world	
4.Medical problem caused by pollen	
5.Main chemicals responsible for climate change	
6.Human activities that we do is the amount of oxygen in the atmosphere	
7.What we should do to reduce the amount of CO2 in the atmosphere	

A.5.6.3 CAT tasks using Bilingual Corpus

1) **The Productive tasks were copied examples, adjusted and printed in papersheets from the Arabic English Parallel Corpus**

In English	In Arabic
One report suggests additional barriers: greater financial needs and being more location-bound owing to finances or family obligations	ويشير أحد التقارير إلى وجود عوائق إضافية، تتمثل في الاحتياجات المالية الكبيرة، وفي كونهم مرتبطين بمناطقهم بشكل أكبر، نتيجة الْتِزامات مالية أو أُسَرِيّة
This was previously difficult owing to the small size and fragility of diamond-cell samples.	كان هذا صعبًا في السابق، نظرًا لصغر حجم وهشاشة عينات خلية ألماس.

In English	In Arabic
Then seeing that her feelings had not changed, he went on, I thought you would do this for me, Lizzie. But I can soon put things right.	عندئذ, وبعد ان راى بان مشاعرها لم تتغير, تابع: فكرت بانك ستفعلين هذا من أجلي يالِيزي, لكن سرعان ما سأتمكن من وضع الأمور في نصابها.
But seeing that his young relative looked worried, he added a few soothing words, and thus wiped out any impression that his wife might have made.	لكنه حين وجد قريبه الشاب مضطربا، اضاف بضع كلمات مهدئة، وبذلك أزال أي ضيق تسببت به زوجته

In English	In Arabic
The bottom line is to pay attention to your physical state when you're making a decision.	والخلاصة هنا انه يجب عليك ان تنتبه لحالتك الجسدية عندما تتخذ قراراً ما.

In English	In Arabic
These conditions do not exist in the same combination elsewhere in the world, but the development of shale gas and oil is nonetheless beginning to happen, albeit at a slower pace.	وهذه الظروف لا توجد بنفس الشكل في اي مكان اخر في العالم لكن تطوير غاز الإردواز والنفط مع ذلك بدأ بالحدوث وإن كان بوتيرة أبطأ

In English	In Arabic
Water buoyancy reduces weight-bearing stress on joints and thereby lessens the risk for injuries.	الطفو على الماء يقلل من ضغط الاوزان على المفاصل وبالتالي يقلل من خطورة الإصابة بالجروح.

In English	In Arabic
He is then immediately taken from his proud yet sorrowing parents and adopted by some childless Equilateral, who is bound by oath never to permit the child henceforth to enter his former home or so much as to look upon his relations again, for fear lest the freshly developed organism may, by force of unconscious imitation, fall back again into his hereditary level.	ويأخذونه على الفور من ابويه اللذين تتنازعهما مشاعر الفخر به والحزن لفراقه، ويتبناه أحد المضلعات المنتظمة التي لم ترزق بذرية، ويتعهد بألا يسمح للطفل منذ ذلك الحين بزيارة بيته السابق أو حتى إلقاء نظرة على أقربائه مرة ثانية، خشية أن يرتد — بفعل المحاكاة غير الواعية — إلى مستواه الموروث وهو لم يزل بَعدُ حديث عهد بالتطور

In English	In Arabic
Even after the wind has abated, the aftermath in the countryside may take some time and a great deal of care to recover.	حتى بعدما تخف الرياح، فربما تتطلب اثار الكارثة في الريف بعض الوقت وقسطًا كبيرًا من العناية للتعافي

2) **<u>The Receptive tasks: The Arabic English Parallel Corpus was used for this task.</u>**

A.5.6.4 Pilot Pre and Delayed Post-tests

The Productive Test (pre-test and delayed post-test):

Please provide the most appropriate translation(s) for the following underlined words and phrases.

الرجاء كتابة الترجمة المناسبة للكلمات الى تحتها خط:

١) إذا رحلتك تأخرت، الخلاصة ان الرحلة المتصلة ستفوتك.
The b_____ _____

٢) يجب ان نذهب الى الحفل الموسيقي بما اننا دفعنا ثمن التذاكر.
Se_____ _____

٣) خالد تجول وعلى رأسه وعاء ، وبالتالي تسبب في اصطدامه بجدار.
T_____

٤) لم أحضر الاجتماع بسبب الصداع.
Ow_____ _____

٥) أتوقع منك أن تتبع قواعد الجامعة من الآن فصاعدا.
He_____

٦) أنا مضطر ، بقدر ما أستطيع ، أن أكتب عن العالم كما أجده.
In so _____ _____

٧) كانت آثار الانفجار مروعة ورهيبة.
The af_____

٨) الكتاب كان طويلا جدا، رغم ذلك غني بالمعلومات وممتع.
Non_____

٩) لديه بيت في لندن، ناهيك عن فيلا في اسبانيا
Not ____ _____

١٠) بالكاد يستطيع مايكل غليان البيض ،وبالأخص طهي الطعام.
Mu____ _____

١١) ابني عادات جيدة على مدى أشهر وسنوات ، وسوف تجد العديد من الفوائد على المدى البعيد.
In the _____ _____

١٢) كان الرجل غير مهذب، أضف الى ذلك كان وقحا جدا.
Not ___ _____

١٣)لا يمكن أن تثق بها لرعاية كلبك، ناهيك عن طفلك.
_____ L_____

١٤) استعرضت اللجنة الأوراق وقررت عندئذ قبول الفكرة.
T_____

١٥) أعتقد أن السعر عادل، لكن من ناحية أخرى لا أستطيع حقا دفع هذا القدر من المال .
On the o_____ _____

١٦) إنه ليس متفائلا جدا بنتيجة المقابلة.
The out_____

١٧) إن نتيجة التمرين اليومي الروتيني هي زيادة العضلات وتقويتها.
The up_____

١٨) لم نواجه اي صعوبات حتى الآن.
Th_____ _____

١٩) لقد كان الرجل متعبا جدا، ومع ذلك ذهب للجري.
Nev_____

٢٠) إنه يعمل ببطء وبدقة بينما أميل الى التسرع في الأمور وارتكاب الأخطاء.
Wh_____

٢١) من الصعب اختيار السيارة لشرائها، ولكن في النهاية تتلخص في التكلفة.
It bo_____ _____

2.The Receptive Test (pre-test and the delayed post-test)

Translate the following sentences from English to Arabic. Make sure you translate the underlined words.

1) If your flight is late, the bottom line is that you will miss your connected flight.

2) We should go to the concert, seeing that we've already paid for the tickets.

3) Khalid walked around with the pot on his head, thereby causing him to bump into a wall.

4) I didn't attend the meeting owing to a headache.

5) I expect you to follow the university's rules henceforth.

6) I am required, in so far as I am able, to write about the world as I find it to be.

7) The aftermath of the explosion was horrific and horrible.

8) The book is too long but, nonetheless informative and entertaining.

9) He's got a house in London, not to mention the villa in Spain.

10) Michael can hardly boil an egg, much less cook dinner.

11) Build good habits over months and years and you will find many benefits in the long run.

12) He was impolite, not to say very rude!

13) You couldn't trust her to look after your dog, let alone your child.

14) The committee reviewed the papers and thereupon decided to accept the idea.

15) I think the price is fair, but on the other hand, I really can't afford to spend that much money.

16) He is not very hopeful about the outcome of the interview.

17) The upshot of a daily exercise routine is that you increase and strengthen muscles.

18) We haven't had any difficulties thus far.

19) The man was very tired; nevertheless, he went for a run.

20) He works slowly and precisely whereas I tend to rush things and make mistakes.

21) It's difficult to choose which car to buy, but in the end it boils down to the cost.

A.5.6.5 STIMULATED RECALL INTERVIEW (SRI) AND FILL-IN-GAPS TASK:

Transcription:
R: I want know what were you thinking about while doing the fill in the gap test, which is this (referring to the fill-in-th-gap test) this is called fill-in-the-gap-test.
I want you remember what you were thinking of at that moment you were answering not now. It is very important while you are answering to remember what were you thinking about while answering. Okay? This is the point. Anything you remember is what you remember while doing the test in particular ,not what you feel now. Okay?

R: How did you feel about it in general?

S:I felt it was a little bit above my level. But I tried to understand the sentence, the content of the sentence. And illuminate the wrong answers. I was only able to that action.

R: Even now after you did this (referring and pointing at the fill-in-the-gap-test that was just taken by the student)?

S: No

R: That was in the beginning?

S: Because the exercises were different. And the words were repeated, so I was able to understand them over and over.

R: So to be clear, that was in the very first fill in the gap test. You used the strategy of illumination and trying to infer the word meaning from the sentence.

S: Yes

R: And now after you took the lesson, were you able to know the answer or did you use the same strategy of illuminating the answers?

S: No, it was easier. I did not need to use any strategy, Everything was clear.

R: So you understood while answering the fill-in-the-gap-test (pointing to the stimuli: the fill-in-the-gap- test)

S: Yes

R: In general you pointed out that the tests was a little

S: difficult (student proceeds with answer)

R: difficult**.** And now after you took the lesson it was….(researcher waits for student to finish the sentence.

S: easier

R: easier. While you were solving the questions, did you feel you might use these words in the future?

S: In the fist time no, because I did not understand the literal meaning of the words. But after I understood them and we took the lesson, I felt these words can be beneficial to me a lot.

R: So you felt you can use them?

S: Yes

R: In general, while you were doing this fill-in the gap test, how did you feel the about CAT teaching…did you feel it was helpful while you were answering the test?

S: Yes, for sure. It was much easier to understand in Arabic and English and know how to use it originally.

R: So you felt it was helpful for you?

S: Yes

R: Can you give me an example, in the aftermath word for example, did you feel the explanation in class was helpful when you answered here?

S: Yes.

R: Can you tell me what you remembered here?

S: When I read the sentence, [reading in English] in …of the hurricane and the strong wind. This is a *kartitha* (catastrophe) that happened. [Reading in English] many people's homes were destroyed. That means after the *karitha***,** the aftermath means *awakib*. So I quickly understood that they are talking about something that happened after the *karitha*. So for sure it's *awakib*.

R: Okay, so you remembered the Arabic word while answering? You remembered *awakib?*

S: Yes.

R: Can you give me another example as well?

S: No to mention. The cost and then there was another thing the risk, *naheek an*, I quickly understood I'm talking about something, *naheek an* (not to mention) the other details of the other thing.

R: So again for you the Arabic was helpful?

S: Yes it was helpful.

R: Can you remember anything unique or special while you were doing the test?

S: Here [pointing and reading the example] the heavy traffic he was late. In the first test, the answer was for sure *because*, I was thinking it was for sure *because*. But I did not ever expect *owing to,* I think I put *seeing that.* But now I feel *owing to* is very logic.

R: So in the first test you do not recall what you answered?

S: Yes

R: You do not recall but you have doubts that it is *seeing that,* correct?

S: Yes, meaning that I illuminated *owing to*.

R: But after you understood the meaning of the words, did you feel the word *owing to* had a meaning while answering.

S: Yes

Fill in the gap-task

Choose the best answer and fill in the gap:

1) The costs, _____ the risks, of moving the satellite to and from space would have been difficult.

a) not to mention
b) whereas
c) thereby

2) I know about all the problems, but what is _____? What will happen?

a) on the other hand
b) the bottom line
c) in so far

3) There is not much we can do, _____they've already made their decision.

a) in the long run
b) the outcome
c) seeing that

4) She forgot to tie her shoes, _____ tripping and falling down the stairs.

a) thereby
b) in the long run
c) whereas

5) The baby can't even walk, _____ run.

a) on the other hand
b) much less
c) thereby

6)_____ the heavy traffic, he was late.

a) The upshot
b) Seeing that
c) Owing to

7) I promise not to lie to you _____

a) in so far
b) henceforth
c) much less

8) This is the truth_____ I know it.

a) in so far as
b) henceforth
c) seeing that

9) In _____of the hurricane and the strong wind, many people's homes were destroyed.

a) the bottom line
b) thus far
c) the aftermath

10) Though she is a fool, I like spending time with her _____.

a) Owing to
b) nonetheless
c) not to mention

11) Being alone can seem scary and uncomfortable at first but it will strengthen you_____.

a) in the long run
b) let alone
c) whereas

12) It would be a mistake, _____ stupid, to leave your first job after only four months.

a) whereas
b) not to say
c) it boils down

13) There isn't room for us, _____ any guests.

a) let alone
b) henceforth
c) thereby

14) The police arrived. _____ the thieves ran away.

a) Thereupon
b) In the long run
c) The bottom line

15) I want to go to the party, but _____I ought to be studying.

a) whereas
b) let alone
c) on the other hand

16) _____ of their decision is still unknown.

a) Let alone
b) The outcome
c) In so far as

17)_____ of reading many books is that you will have a lot of information.

a) The upshot
b) Nevertheless

c) Nonetheless

18) We haven't had any problem_____
a) henceforth
b) thus far
c) thereby

19) The old system was difficult _____ the new system is really very simple.
a) owing to
b) not to mention
c) whereas

20) It all _____ to money in the end.
a) the bottom line
b) let alone
c) boils down

21) The mother was very tired;_____, she cooked dinner for her kids.
a) nevertheless
b) in the long run
c) owing to

Appendix B – Productive Test

Table 1. Dependent Variables.

Within-Subjects Factors

Measure: response

Week	Dependent Variable
1	Pre-prodc
2	Post-prodc

Table 2. Sample Size by Group.

Between-Subjects Factors

		Value Label	N
Groups	1	CAT	19
	2	CAT+C	18
	3	Control	19

Table 2.a. Test of Normality for the CAT, CAT+Corpus and Control Groups (Shapiro-Wilk).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
Groups		Statistic	Df	Sig.	Statistic	df	Sig.
1 CAT	Pre-prodc	.495	19	.000	.460	19	.000
	Post-.prodc	.236	19	.007	.872	19	.016
2 CAT+C	Pre-prodc	.387	18	.000	.651	18	.000
	Post-prodc	.171	18	.176	.923	18	.145
3 Control	Pre-prodc	.495	19	.000	.460	19	.000
	Post-prodc	.398	19	.000	.639	19	.000

a. Lilliefors Significance Correction

Table 2.b. Normality Check: Skewness & Kurtosis.

Case Summaries

Groups		Pre.rodc	Post.prodc
1 CAT	Kurtosis	6.883	.859
	Std. Error of Kurtosis	1.014	1.014
	Skewness	2.658	-.881
	Std. Error of Skewness	.524	.524
2 CAT+C	Kurtosis	3.384	1.136
	Std. Error of Kurtosis	1.038	1.038
	Skewness	1.889	-.459
	Std. Error of Skewness	.536	.536
3 Control	Kurtosis	6.883	3.498
	Std. Error of Kurtosis	1.014	1.014
	Skewness	2.658	1.909
	Std. Error of Skewness	.524	.524

Table 3. Parametric Test Used to Compare Variation in Multivariate Samples.

Box's Test of Equality of Covariance Matrices[a]

Box's M	16.708
F	2.633
df1	6
df2	68703.124
Sig.	.015

Table 4. Overall Significant Difference Between the Means at the Different Time Points.

Tests of Within-Subjects Effects

Measure: response

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Week	Sphericity Assumed	448.146	1	448.146	318.502	.000	.857
	Greenhouse-Geisser	448.146	1.000	448.146	318.502	.000	.857
	Huynh-Feldt	448.146	1.000	448.146	318.502	.000	.857
	Lower-bound	448.146	1.000	448.146	318.502	.000	.857
Week * Groups	Sphericity Assumed	190.918	2	95.459	67.844	.000	.719
	Greenhouse-Geisser	190.918	2.000	95.459	67.844	.000	.719
	Huynh-Feldt	190.918	2.000	95.459	67.844	.000	.719
	Lower-bound	190.918	2.000	95.459	67.844	.000	.719
Error(week)	Sphericity Assumed	74.573	53	1.407			
	Greenhouse-Geisser	74.573	53.000	1.407			
	Huynh-Feldt	74.573	53.000	1.407			
	Lower-bound	74.573	53.000	1.407			

Table 5. The Variances of the Differences Between All Combinations of Related Groups (Levels) are Equal.

Mauchly's Test of Sphericity[a]

Measure: response

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Week	1.000	.000	0	.	1.000	1.000	1.000

Table 6. Levene's Test of Equality of Error Variances for the Simple Effect Group at Both a) Pre-Test and b) Post-Test for the Productive Task.

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
Pre-.Prodc	Based on Mean	3.141	2	53	.051
	Based on Median	1.189	2	53	.313
	Based on Median and with adjusted df	1.189	2	42.854	.314
	Based on trimmed mean	2.818	2	53	.069
Post-.Prodc	Based on Mean	2.593	2	53	.084
	Based on Median	2.689	2	53	.077
	Based on Median and with adjusted df	2.689	2	45.377	.079
	Based on trimmed mean	2.968	2	53	.060

Table 7. Main Effect of Group.

Tests of Between-Subjects Effects

Measure: response

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	596.217	1	596.217	295.194	.000	.848
Groups	204.730	2	102.365	50.682	.000	.657
Error	107.047	53	2.020			

Table 8. Simple Effects Tests of Group at Time 1 and Time 2.

Test	Levene's Test of Equality of Error Variances		ANOVA			Effect Size (ηp^2)
	Levene Statistic	p	F	Df	P	
Pre	2.950	.061	1.119	2 52	.336	.041
Post	2.345	.081	61.78	2 52	<.001	.704

Table 9. Significance Level for Pairwise Differences Between the Individual Groups.

Pairwise Comparisons

Measure: response

Week	(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
						Lower Bound	Upper Bound
2	CAT	CAT+C	1.073	.569	.065	-.069	2.215
		Control	6.053*	.562	.000	4.926	7.179
	CAT+C	CAT	-1.073	.569	.065	-2.215	.069
		Control	4.980*	.569	.000	3.838	6.122
	Control	CAT	-6.053*	.562	.000	-7.179	-4.926
		CAT+C	-4.980*	.569	.000	-6.122	-3.838

Table 10. Simple Effect Tests Investigating the Simple Effect of Time for Each Group.

Paired Samples Test

Groups			Paired Differences					t	df	Sig. (2-tailed)	Estimate Point Cohen's d
			Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
						Lower	Upper				
1 CAT	Pair 1	Pre-Prodc – Post-Prodc	-6.47368	1.71167	.39268	-7.29868	-5.64868	-16.486	18	.000	-3.78
2 CAT+C	Pair 1	Pre-Prodc – Post-.Prodc	-5.11111	2.16629	.51060	-6.18838	-4.03384	-10.010	17	.000	-2.307
3 Control	Pair 1	Pre-.Prodc – Post-Prodc	-.42105	.96124	.22052	-.88435	.04225	-1.909	18	.072	-.452

Table 11.a. Results of Kruskall Wallis Test of the Difference Between Groups at Productive Pre-Test.

Total N	56
Test Statistic	2.237
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	0.327

Table 11.b. Results of Kruskall Wallis Test of the Difference Between Groups at Productive Post-Test.

Total N	56
Test Statistic	34.806
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

12.a Man-Whitney U Post Hoc Test for the CAT and CAT+Corpus Groups in the Productive Area.

Test Statistics[a]

	Pre-Prodc	Post-Prodc
Mann-Whitney U	140.500	113.500
Wilcoxon W	330.500	284.500
Z	-1.236	-1.772
Asymp. Sig. (2-tailed)	.217	.076
Exact Sig. [2*(1-tailed Sig.)]	.358[b]	.081[b]

Table 12.b. Man-Whitney U Post Hoc Test for the CAT and Control Groups in the Productive Area.

Test Statistics[a]

	Pre-Prodc	Post-Prodc
Mann-Whitney U	180.500	2.000
Wilcoxon W	370.500	173.000
Z	.000	-5.253
Asymp. Sig. (2-tailed)	1.000	.000
Exact Sig. [2*(1-tailed Sig.)]	1.000[b]	.000[b]

Table 12.c. Man-Whitney U Post Hoc Test for the CAT+Corpus and Control Groups in the Productive Area.

Test Statistics[a]

	Pre-Prodc	Post-Prodc
Mann-Whitney U	140.500	14.000
Wilcoxon W	330.500	185.000
Z	-1.236	-4.813
Asymp. Sig. (2-tailed)	.217	.000
Exact Sig. [2*(1-tailed Sig.)]	.358[b]	.000[b]

Table 13.a. Wilcoxon Test Showing Significance Between Time 1 and Time 2 for the CAT Group in Productive Area.

Total N	19
Test Statistic	190.000
Standard Error	24.696
Standarized Test Statistics	3.847
Asymptotic Sig. (2-sided test)	.000

Table 13.b. Wilcoxon Test Showing Significance Between Time 1 and Time 2 for the CAT+Corpus Group in Productive Area.

Total N	18
Test Statistic	153.000
Standard Error	21.045
Standarized Test Statistics	3.635
Asymptotic Sig. (2-sided test)	.000

Table 13.c. Wilcoxon Test Showing Insignificance Between Time 1 and Time 2 for the Control Group in Productive Area.

Total N	19
Test Statistic	31.000
Standard Error	6.964
Standarized Test Statistics	1.867
Asymptotic Sig. (2-sided test)	.062

Table 14. Median for Pre-Test and Post-Test of the Three Groups in Productive Area.

Report

groups		Pre-Prodc	Post-Prodc
CAT	N	19	19
	Median	.0000	7.0000
CAT+C	N	18	18
	Median	.0000	5.0000
Control	N	19	19
	Median	.0000	.0000
Total	N	56	55
	Median	.0000	5.0000

Appendix C – Receptive Test

Table 1. Dependent Variables.

Within-Subjects Factors

Measure: response

Week	Dependent Variable
1	Pre-Receptive
2	Post-.Receptive

Table 2. Sample Size by Group.

Between-Subjects Factors

		Value Label	N
Groups	1	CAT	19
	2	CAT+C	18
	3	Control	19

Table 2a. Test of Normality for the CAT, CAT+Corpus, control groups (Shapiro-Wilk).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
groups		Statistic	df	Sig.	Statistic	df	Sig.
1 CAT	Pre-Receptive	.162	19	.200*	.890	19	.032
	Post-Receptive	.306	19	.000	.756	19	.000
2 CAT+C	Pre-Receptive	.165	18	.200*	.921	18	.135
	Post-Receptive	.268	18	.001	.846	18	.007
3 Control	Pre-Receptive	.167	19	.171	.924	19	.134
	Post-.Receptive	.295	19	.000	.829	19	.003

Table 2b. Normality Check Skewness & Kurtosis.

Case Summaries

Groups		Pre-Receptive	Post-Receptive
1 CAT	Kurtosis	-1.378	-.764
	Std. Error of Kurtosis	1.014	1.014
	Skewness	-.008	-.914
	Std. Error of Skewness	.524	.524
2 CAT+C	Kurtosis	-1.344	-.324
	Std. Error of Kurtosis	1.038	1.038
	Skewness	.163	-.909
	Std. Error of Skewness	.536	.536
3 Control	Kurtosis	-1.123	-.686
	Std. Error of Kurtosis	1.014	1.014
	Skewness	.310	-.886
	Std. Error of Skewness	.524	.524

Table 3. Parametric Test Used to Compare Variation in Multivariate Samples.

Box's Test of Equality of Covariance Matrices[a]

Box's M	11.116
F	1.751
df1	6
df2	68703.124
Sig.	.105

Table 4. Overall Significant Difference Between the Means at the Different Time Points.

Tests of Within-Subjects Effects

Measure: Response

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Week	Sphericity Assumed	344.306	1	344.306	161.041	.000	.752
	Greenhouse-Geisser	344.306	1.000	344.306	161.041	.000	.752
	Huynh-Feldt	344.306	1.000	344.306	161.041	.000	.752
	Lower-Bound	344.306	1.000	344.306	161.041	.000	.752
Week * Groups	Sphericity Assumed	31.686	2	15.843	7.410	.001	.219
	Greenhouse-Geisser	31.686	2.000	15.843	7.410	.001	.219
	Huynh-Feldt	31.686	2.000	15.843	7.410	.001	.219
	Lower-Bound	31.686	2.000	15.843	7.410	.001	.219
Error(week)	Sphericity Assumed	113.314	53	2.138			
	Greenhouse-Geisser	113.314	53.000	2.138			
	Huynh-Feldt	113.314	53.000	2.138			
	Lower-Bound	113.314	53.000	2.138			

Table 5. The Variances of the Differences Between all Combinations of Related Groups (Levels) are Equal.

Mauchly's Test of Sphericity[a]

Measure: Response

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-Bound
Week	1.000	.000	0	.	1.000	1.000	1.000

Table 6. Levene's Test of Equality of Error Variances for the Simple Effect Group at Both a) Pre-Test and b) Post-Test for the Receptive Task.

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
Pre-Receptive	Based on Mean	.200	2	53	.819
	Based on Median	.130	2	53	.878
	Based on Median and with Adjusted df	.130	2	47.837	.878
	Based on Trimmed Mean	.205	2	53	.816
Post-Receptive	Based on Mean	4.816	2	53	.012
	Based on Median	1.400	2	53	.256
	Based on Median and with Adjusted df	1.400	2	48.334	.256
	Based on Trimmed Mean	4.103	2	53	.022

Table 7. Main Effect of Group.

Tests of Between-Subjects Effects

Measure: Response

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	2803.093	1	2803.093	334.812	.000	.863
groups	40.276	2	20.138	2.405	.100	.083
Error	443.724	53	8.372			

Table 8. Simple Effect of Group at Time 1 and Time 2.

Test	Levene's Test of Equality of Error Variances		ANOVA			Effect Size (ηp^2)
	Levene Statistic	*P*	*F*	*Df*	*p*	
Pre	.200	.819	.022	2 53	.987	.001
Post	.205	.012	6.39[a]	2 33.43	.004	.227

a= robust tests of equality of means for heterogeneity of Variances

Table 9. Significance Level for Differences Between the Individual Group.

Pairwise Comparisons

Measure: Response

(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
CAT	CAT+C	.620	.707	.384	-.798	2.038
	Control	2.632*	.697	.000	1.233	4.030
CAT+C	CAT	-.620	.707	.384	-2.038	.798
	Control	2.012*	.707	.006	.594	3.430
Control	CAT	-2.632*	.697	.000	-4.030	-1.233
	CAT+C	-2.012*	.707	.006	-3.430	-.594

Table 10. Simple Effect Tests Investigating the Simple Effect of Time for Each Group.

Paired Samples Test

Groups			Paired Differences					t	df	Sig. (2-tailed)	Estimate Point Cohen's d
			Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference					
						Lower	Upper				
1 CAT	Pair 1	Pre-Receptive – Post-Receptive	-4.52632	1.95415	.44831	-5.46819	-3.58444	-10.096	18	.000	-2.316
2 CAT+C	Pair 1	Pre-Receptive – Post-Receptive	-3.94444	1.92422	.45354	-4.90134	-2.98755	-8.697	17	.000	-2.050
3 Control	Pair 1	Pre-Receptive – Post-Receptive	-2.05263	2.29670	.52690	-3.15961	-.94566	-3.896	18	.001	-.894

Table 11. Welch Test for Receptive Post-Test.

	Statistic[a]	df1	df2	Sig.
Welch	6.392	2	33.433	.004

a. Asymptotically F distributed.

Table 12.a. Results of Kruskall Wallis Test of the Difference Between Groups at Receptive Pre-Test.

Total N	56
Test Statistic	.022
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.989

Table 12.b. Results of Kruskall Wallis Test of the Difference Between Groups at Receptive Post-Test.

Total N	56
Test Statistic	12.406
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.002

Table 13.a. Man-Whitney U Post Hoc Test for the CAT and CAT+Corpus Groups in the Receptive Area.

Test Statistics[a]

	Pre-Receptive	Post-Receptive
Mann-Whitney U	169.000	136.000
Wilcoxon W	359.000	307.000
Z	-.061	-1.117
Asymp. Sig. (2-tailed)	.951	.264
Exact Sig. [2*(1-tailed Sig.)]	.964[b]	.298[b]

Table 13.b. Man-Whitney U Post hoc Test for the CAT and Control Groups in the Receptive Area.

Test Statistics[a]

	Pre-Receptive	Post-Receptive
Mann-Whitney U	177.000	69.500
Wilcoxon W	367.000	259.500
Z	-.103	-3.300
Asymp. Sig. (2-tailed)	.918	.001
Exact Sig. [2*(1-tailed Sig.)]	.931[b]	.001[b]

Table 13.c. Man-Whitney U Post Hoc Test for the CAT+Corpus and Control Groups in the Receptive Area.

Test Statistics[a]

	Pre-Receptive	Post-Receptive
Mann-Whitney U	166.500	90.500
Wilcoxon W	356.500	280.500
Z	-.138	-2.481
Asymp. Sig. (2-tailed)	.890	.013
Exact Sig. [2*(1-tailed Sig.)]	.893[b]	.013[b]

Table 14.a. Wilcoxon Test Showing Significance Between Time 1 and Time 2 for the CAT Group in the Receptive Area.

Total N	19
Test Statistic	190.000
Standard Error	24.777
Standarized Test Statistics	3.834
Asymptotic Sig. (2-sided test)	.000

Table 14.b. Wilcoxon Test Showing Significance Between Time 1 and Time 2 for the CAT+Corpus Group in the Receptive Area.

Total N	18
Test Statistic	171.000
Standard Error	22.891
Standarized Test Statistics	3.735
Asymptotic Sig. (2-sided test)	.000

Table 14.c. Wilcoxon Test Showing Significance Between Time 1 and Time 2 for the Control Group in the Receptive Area.

Total N	19
Test Statistic	113.000
Standard Error	17.496
Standarized Test Statistics	3.029
Asymptotic Sig. (2-sided test)	.002

Table 15. Median for Pre-Test and Post-Test of the Three Groups in the Receptive Area.

Report

Groups		Pre-Receptive	Post-Receptive
CAT	N	19	19
	Median	4.0000	9.0000
CAT+C	N	18	18
	Median	3.5000	8.0000
Control	N	19	19
	Median	3.0000	6.0000
Total	N	56	56
	Median	3.5000	7.5000

Appendix D - Ethical Forms Approved in English and Arabic

University of Reading
Institute of Education
Ethical Approval Form A (version May 2019)

University of Reading

Tick one: Staff project: _____ PhD ✓ EdD ____

Name of applicant (s): Sarah Alamoudi

Title of project: A Sequential Explanatory Design on Non- Semantic Transparent Adverbials within Contrastive Analysis and Translation Setting

Name of supervisor (for student projects): Suzanne Graham

Please complete the form below including relevant sections overleaf.

	YES	NO
Have you prepared an Information Sheet for participants and/or their parents/carers that:		
a) explains the purpose(s) of the project	✓	
b) explains how they have been selected as potential participants	✓	
c) gives a full, fair and clear account of what will be asked of them and how the information that they provide will be used	✓	
d) makes clear that participation in the project is voluntary	✓	
e) explains the arrangements to allow participants to withdraw at any stage if they wish	✓	
f) explains the arrangements to ensure the confidentiality of any material collected during the project, including secure arrangements for its storage, retention and disposal	✓	
g) explains the arrangements for publishing the research results and, if confidentiality might be affected, for obtaining written consent for this	✓	
h) explains the arrangements for providing participants with the research results if they wish to have them	✓	
i) gives the name and designation of the member of staff with responsibility for the project together with contact details, including email . If any of the project investigators are students at the IoE, then this information must be included and their name provided	✓	
k) explains, where applicable, the arrangements for expenses and other payments to be made to the participants	✓	
j) includes a standard statement indicating the process of ethical review at the University undergone by the project, as follows: 'This project has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct'.	✓	
k)includes a standard statement regarding insurance: "The University has the appropriate insurances in place. Full details are available on request".	✓	
Please answer the following questions		
1) Will you provide participants involved in your research with all the information necessary to ensure that they are fully informed and not in any way deceived or misled as to the purpose(s) and nature of the research? (Please use the subheadings used in the example information sheets on blackboard to ensure this).	✓	
2) Will you seek written or other formal consent from all participants, if they are able to provide it, in addition to (1)?	✓	
3) Is there any risk that participants may experience physical or psychological distress in taking part in your research?		✓
4) Staff Only - have you taken the online training modules in data protection and information security (which can be found here: http://www.reading.ac.uk/internal/humanresources/PeopleDevelopment/newstaff/humres-		

MandatoryOnlineCourses.aspx Please note: students complete a Data Protection Declaration form and submit it with this application to the ethics committee.			
5) Have you read the Health and Safety booklet (available on Blackboard) and completed a Risk Assessment Form to be included with this ethics application?	✓		
6) Does your research comply with the University's Code of Good Practice in Research?	✓		
	YES	NO	N.A.
7) If your research is taking place in a school, have you prepared an information sheet and consent form to gain the permission in writing of the head teacher or other relevant supervisory professional?	✓		
8) Has the data collector obtained satisfactory DBS clearance?			✓
9) If your research involves working with children under the age of 16 (or those whose special educational needs mean they are unable to give informed consent), have you prepared an information sheet and consent form for parents/carers to seek permission in writing, or to give parents/carers the opportunity to decline consent?			✓
10) If your research involves processing sensitive personal data[1], or if it involves audio/video recordings, have you obtained the explicit consent of participants/parents?			✓
11) If you are using a data processor to subcontract any part of your research, have you got a written contract with that contractor which (a) specifies that the contractor is required to act only on your instructions, and (b) provides for appropriate technical and organisational security measures to protect the data?			✓
12a) Does your research involve data collection outside the UK?	✓		
12b) If the answer to question 12a is "yes", does your research comply with the legal and ethical requirements for doing research in that country?	✓		
13a) Does your research involve collecting data in a language other than English?	✓		
13b) If the answer to question 13a is "yes", please confirm that information sheets, consent forms, and research instruments, where appropriate, have been directly translated from the English versions submitted with this application.	✓		
14a. Does the proposed research involve children under the age of 5?		✓	
14b. If the answer to question 14a is "yes": My Head of School (or authorised Head of Department) has given details of the proposed research to the University's insurance officer, and the research will not proceed until I have confirmation that insurance cover is in place.			✓
If you have answered YES to Question 3, please complete Section B below			

- Complete **either** Section A **or** Section B below with details of your research project.
- Complete a risk assessment.
- Sign the form in Section C.
- Append at the end of this form all relevant documents: information sheets, consent forms, tests, questionnaires, interview schedules, evidence that you have completed information security training (e.g. screen shot/copy of certificate).
- Email the completed form to the Institute's Ethics Committee for consideration.

Any missing information will result in the form being returned to you.

A: My research goes beyond the 'accepted custom and practice of teaching' but I consider that this project has **no** significant ethical implications. (Please tick the box.)	✓
Please state the total number of participants that will be involved in the project and give a breakdown of how many there are in each category e.g. teachers, parents, pupils etc. Students: Pilot Study 30 students Main Study: 180 students	

[1] Sensitive personal data consists of information relating to the racial or ethnic origin of a data subject, their political opinions, religious beliefs, trade union membership, sexual life, physical or mental health or condition, or criminal offences or record.

Give a brief description of the aims and the methods (participants, instruments and procedures) of the project in up to 200 words noting:
title of project:

1. purpose of project and its academic rationale
2. brief description of methods and measurements
3. participants: recruitment methods, number, age, gender, exclusion/inclusion criteria
4. consent and participant information arrangements, debriefing (attach forms where necessary)
5. a clear and concise statement of the ethical considerations raised by the project and how you intend to deal with then.
6. estimated start date and duration of project

A Sequential Explanatory Design on Non- Semantically Transparent Adverbials within Contrastive Analysis and Translation Setting

Linking adverbials are important in writing and linking thoughts and ideas. Some linking adverbials are semantically transparent and can be guessed and inferred by students (such as 'first of all') and thus, easily translated. However, there are non-semantically transparent linking adverbials that could not easily be inferred or translated (such as 'whereas' and 'hence') are potentially harder to learn. This study aims to investigate through a mixed-methods approach the effectiveness of teaching English language learners to learn 21 non-semantically transparent linking adverbials by using a Contrastive Analysis and Translation Method (CAT) instructional approach. CAT is "the kind of instruction which leads to learners' understanding of the similarities and differences between their L1 and L2 in terms of individual words and the overall lexical system" (Laufer and Girsai, 2008, p. 696). The participants are first year undergraduate students in Saudi Arabia. They have all registered to take English language courses as a requirement for their first year. Their ages range between 17-21. The participants have been selected because of their level of English proficiency, which is high-intermediate through convenience sampling, from those at a university known to the applicant. The researcher will start with providing the Head Department with information about the study. Once receiving the approval and the consent form, the researcher will explain and translate the information sheet and consent form to the participants. After receiving the participants' consent forms, the researcher will start by collecting their standardised test scores upon university entry and language background questionnaires (translated into Arabic, the students' first language). The students that do not wish to participate will be in class during reading sessions and tasks but they will NOT take any tests related to the study. Instead, they will receive comprehension questions related to their books, which will be in paper format. They will not be interviewed and their personal information and data will not be used.

The study will take place in two modules. The procedure in both modules are exactly the same, however the participants will be different. The procedure will start by following a pre to post and delayed-test design lasting approximately 4 weeks for each module. Three classes will be randomly assigned to three groups, all of which will be exposed to reading passages containing the target vocabulary items: the first two will undertake contrastive analysis and translation type activities, with one group using a computer corpus plus input from the teacher for those tasks, the other group just teacher input. The third group will be the control group who will just read the passages and answer comprehension questions on them. The third group will not receive computer corpora tasks nor CAT activities. The researcher will teach all three groups and the teaching for the intervention will occur during two periods of 90 minutes each. Prior to the intervention, all students will complete pre-tests to assess their learning of the target vocabulary items, consisting of

1) Receptive tests (translating the 21 target vocabulary L2 to L1)
2) Productive tests (translating 21 target vocabulary from L1 to L2)
3) Fill-in-the-gap test
4) A sample of two students in each class will then complete a Stimulated Recall Interview (SRI) on their responses to the fill-in the gap test. The SRI will take about 20 minutes and will be audio-recorded, translated, and transcribed as stated in the consent form. (All recordings and data will be securely saved in the researcher's password-protected computer and locked filing cabinet. Only the researcher and the supervisors can access the data and recordings).

Immediately after the teaching intervention, all students will complete a post-test consisting of the receptive and productive tests which will include eleven of the target vocabulary items. One week after the intervention, they will complete a delayed post-test consisting of the receptive and productive tests including the same eleven target vocabulary items. The same procedures will then be followed for the remaining 10 items. The design for the study is show below:

Module 1

WEEK 1

Pre-test, fill in the gap, SRI (all 21 items)

WEEK 2

Intervention for first 11 items, immediate post-test

WEEK 3

Delayed post-test (first 11 items)

FOLLOWING DAY

Intervention for next 10 items, immediate post-test

WEEK 4

Delayed post-test, fill in the gap, SRI (next 10 items)

Module 2

Same phase, same procedure, and same vocabulary items as Module 1, but with different participants.

The pilot study will be in October 2019 and the main the study will start in January and ends in May 2020.

B: I consider that this project **may** have ethical implications that should be brought before the Institute's Ethics Committee.	

Please state the total number of participants that will be involved in the project and give a breakdown of how many there are in each category e.g. teachers, parents, pupils etc.

Give a brief description of the aims and the methods (participants, instruments and procedures) of the project in up to 200 words.

1. title of project
2. purpose of project and its academic rationale
3. brief description of methods and measurements
4. participants: recruitment methods, number, age, gender, exclusion/inclusion criteria
5. consent and participant information arrangements, debriefing (attach forms where necessary)
6. a clear and concise statement of the ethical considerations raised by the project and how you intend to deal with then.
7. estimated start date and duration of project

RISK ASSESSMENT: Please complete the form below

Brief outline of Work/activity:	The Researcher will investigate Contrastive Analysis and Translation (CAT) instructional approach on English learners. CAT tasks help learners to understand similarities and differences in their L1 and L2 (Laufer and Girsai, 2008). The researcher will provide reading sessions and comprehension questions for three groups. The first two groups will receive CAT activities and tasks, one group will incorporate a computer corpus along with input from the teacher for those tasks and the other group will only receive input from the teacher. The third group, which is the control group, will take on the reading sessions followed by the comprehension questions without any computer corpora activities or CAT tasks. There will be two phases, whereas in the first phase the

	participants will receive CAT tasks on 11 vocabulary items. In the second phase, the participants will receive CAT instruction and tasks on 10 vocabulary items. The researcher will assess the participants' learning of the 21 (11+10) target items through (pre-tests, post-tests, and delayed post-tests), fill-in-the-gap tests, and Stimulated Recall Interviews (2 students for each group). The study will last approximately 4 weeks for Module 1 and approximately 4 weeks for Module 2. The two modules will include the same phases, the same procedure, and the same vocabulary items. The participants will not be the same.

Where will data be collected?	English Language Institute, King Abdul Aziz University

Significant hazards:	None

Who might be exposed to hazards?	None

Existing control measures:	The researcher will highly consider each participant's well-being and will follow the ethic protocols, which includes information sheets and consent forms to the head department and participants. The study will take place on usual and regular teaching schedule, usual site, questions and assessments are related to the curriculum, interviews are related to the student's learning and assessments, identities will be anonymous, and confidentiality of the participants and the institute will be protected. If students decide not to participate, they will be in class during reading sessions and tasks but they will NOT take any tests related to the study. Instead, they will receive comprehension questions related to their books, which will be in paper format. They will not be interviewed and their personal information and data will not be used. The researcher will treat each participant respectfully, fairly, sensitively, and with dignity and freedom from prejudice. The researcher will not not attempt to harm, embarrass, or shame the participants. The researcher will not put herself or participants in any physical or moral harm.

Are risks adequately controlled:	No risks

If NO, list additional controls and actions required:	Additional controls	Action by:
	Freedom to withdraw from the study at any stage.	Participants and Head department

C: SIGNATURE OF APPLICANT:

Note: a signature is required. Typed names are not acceptable.

I have declared all relevant information regarding my proposed project and confirm that ethical good practice will be followed within the project.

Signed: Print Name Sarah Alamoudi Date 8/27/2019.

STATEMENT OF ETHICAL APPROVAL FOR PROPOSALS SUBMITTED TO THE INSTITUTE ETHICS COMMITTEE

This project has been considered using agreed Institute procedures and is now approved.

Signed: Print Name: Karen Jones
(IoE Research Ethics Committee representative)*

Date: 2 October 2019

(IoE Research Ethics Committee representative)*

* A decision to allow a project to proceed is not an expert assessment of its content or of the possible risks involved in the investigation, nor does it detract in any way from the ultimate responsibility which students/investigators must themselves have for these matters. Approval is granted on the basis of the information declared by the applicant.

University of Reading

Participant Information Sheet

Researcher: Sarah Alamoudi

Contact s.alamoudi@pgr.reading.ac.uk

Supervisors: Prof. Suzanne Graham and Dr Daisy Powell

Contact s.j.graham@reading.ac.uk; d.a.powell@reading.ac.uk

Project: Vocabulary Learning

Dear Student,
I am a PhD student at the university of Reading, UK. As part of the data collection stage of my thesis, I would like to invite you to take part in a research study about vocabulary learning.

Why have I been invited to take part?
You have been invited to take part because you are in the high- intermediate English program at the English Language Institute and I am interested in finding out how students at that level learn vocabulary.

What will be involved if I agree to take part?
With your consent, I would firstly like to have access to your placement test results. During the course of this module, you will be asked to complete some short vocabulary tasks as part of your normal classes. You will also be asked to complete two to four tasks every week. The tasks involve translating (English to Arabic and vice versa) and fill-in-the-gap activities in English. You will also be asked to complete a questionnaire about your experience of learning English. The tasks and questionnaire should take no more than 15minutes each to complete during class time, and you will complete them with others in your class. The information gathered from the tasks will not affect your GPA or your academic scores but will be used within my thesis.

You might also be selected to participate in an interview with the researcher on two occasions. The interview will also be related to the lessons we take and should not take more than 20 minutes, at a time and place convenient to you. The interview will be in Arabic and will be recorded, translated, and transcribed with your permission. The transcription will be shown to you in order for you to check its accuracy and to confirm that you are still happy for its contents to be used. The information gathered will be used by the researcher for data analysis.

Will anyone know about my answers?
No. Your name and answers will remain confidential and will only be seen by the researcher and the supervisors listed at the top of the letter. You will be assigned an identification number (ID) only to distinguish your responses from those of other participants. This ID is in no way associated with your name. The records of this study will be kept private. No identifiers linking you to the study will be included in any sort of report that might be published.

What will happen to the data?
Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you will be included in any sort of report that might be published. Participants will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet

and on a password-protected computer and only the researcher Sarah Alamoudi and the Supervisors will have access to the records. In line with the University's policy on the management of research data, anonymised data gathered in this research may be preserved and made publicly available for others to consult and re-use. The results of the study will be presented at national and international conferences, and in written reports and articles. We can send you electronic copies of these publications if you wish.

Will I benefit by taking part?

We anticipate that the findings of the study will be useful for teachers and students in planning how they to learn vocabulary.

Do I have to take part?

No, not at all. Your choice take part is voluntary. If you decide not to participate, you will be in class during reading sessions and tasks but you will not take any tests related to the study. Instead, you will receive comprehension questions related to your books and your data will not be used.
Also, you are free to leave the project at any time, without giving a reason. You can do this by letting me know that you don't want to take part or continue.

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form and return it to me.

This project has been reviewed following the procedures of the University Research Ethics Committee and has been given a favourable ethical opinion for conduct. The University has the appropriate insurances in place. Full details are available on request.

Signed: Sarah Alamoudi

Date: 08/05/2019

University of Reading

Consent Form

Researcher: Sarah Alamoudi s.alamoudi@pgr.reading.ac.uk

Supervisors: Suzanne Graham and Dr Daisy Powell s.j.graham@reading.ac.uk;d.a.powell@reading.ac.uk

Project title: Learning Vocabulary

I have read and had explained to me by Sarah Alamoudi the Information Sheet relating to this project.

I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

I understand that I will be asked to complete vocabulary tasks and a questionnaire in class

I understand that I may be interviewed and that the interview will be recorded, translated, and transcribed.

I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time, without giving a reason and without repercussions.

I have received a copy of this Consent Form and of the accompanying Information Sheet.

Please tick as appropriate:

I consent to be involved in the project	_____ yes	_____ no
I consent to have my standardized scores accessed	_____ yes	_____ no
I consent to be interviewed	_____ yes	_____ no
I consent to this interview being recorded	_____ yes	_____ no

Name:

Signed:

University of Reading

Head of department information sheet

Research Project: Vocabulary Learning.
Project Researcher: Sarah Alamoudi
Supervisors: Professor Suzanne Graham & Dr Daisy Powell

Dear xxxx,

I am writing to invite your department to take part in a research study about learning vocabulary in English.

What is the study?
A you know I am a PhD student at the University of Reading. As a fulfilment of my PhD requirements and as part of the data collection stage of my thesis, I invite the English learners in the high-intermediate- level at your department to take part in my research. I am interested in finding out how students at this level learn vocabulary.

The study aims to apply different ways in teaching vocabulary and understand and compare the learning outcomes and the learning process for English learners at university level. It hopes to make recommendations regarding how teachers can best help learners to make progress in English vocabulary learning.

Why have my students been chosen to take part?
Your students have been invited to take part because they are undergraduates enrolled in the English language courses in the English Language Institute. Their participation will assist my research in the collection of appropriate data.

Do my students have to take part?
It is entirely up to you whether you give permission for your students to participate. You may also withdraw your consent to participation at any time during the project, without any repercussions to you, by contacting the researcher; Tel: , e-mail: s.alamoudi@pgr.reading.ac.uk
This study is voluntary and students have the decision to participate or not. If they do not participate, they will be in class during reading sessions and tasks but they will not undertake any tests related to the study. Instead, they will receive comprehension questions related to their books, which will be in paper format. They will not be interviewed and their personal information and data will not be used.

What will happen if the school takes part?
With your agreement and the students' consent, three classes will take part in the project. The students would experience two English reading sessions followed by comprehension questions and tasks related to their curriculum. Each session will take 90 minutes. Students will also be asked to complete two to four tasks every week. The tasks involve translating (English to Arabic and vice versa) and fill-in-the-gap activities in English. They will also be asked to complete a questionnaire about their experience of learning English. The tasks and questionnaire should take no more than 15minutes each to complete during class time

Two interviews (20 minutes each interview) will also be conducted in Arabic with two students for each class, outside of class time. These interviews will be audio recorded, translated, and transcribed with your consent and the students' consent. Finally, it would be helpful to have access to the scores from your

students' English standardized test that they take at enrollment to the university. Students will be asked to give their permission in the consent forms to access these scores.

All the tasks would take place during class time. All efforts would be made to seek an appropriate time for your students to take part and to ensure that their normal study is not adversely affected. There will be a pilot study that will take place on October 2019. The main study will run from January 2020 until May 2020.

What are the risks and benefits of taking part?
The information given by participants in the study will remain confidential and will only be seen by the research team listed at the start of this letter. Neither you or the students will be identifiable in any published report resulting from the study. Information about individuals will not be shared with the school.

Participants in similar studies have found it interesting to take part. We anticipate that the findings of the study will be useful for teachers in planning how they teach English.

What will happen to the data?
Any data collected will be held in strict confidence and no real names will be used in this study or in any subsequent publications. The records of this study will be kept private. No identifiers linking you or the department to the study will be included in any sort of report that might be published. Participants will be assigned a number and will be referred to by that number in all records. Research records will be stored securely in a locked filing cabinet and on a password-protected computer and only the research team will have access to the records. In line with the University's policy on the management of research data, anonymised data gathered in this research may be preserved and made publicly available for others to consult and re-use. The results of the study will be presented at national and international conferences, and in written reports and articles. We can send you electronic copies of these publications if you wish.

What happens if I change my mind?
You can change your mind at any time without any repercussions. If you change your mind after data collection has ended, we will discard your student's data.

What happens if something goes wrong?
In the unlikely case of concern or complaint, you can contact Professor Suzanne Graham or Dr Daisy Powell, University of Reading; email: s.j.graham@reading.ac.uk ; d.a.powell@reading.ac.uk

We do hope that you will agree to your participation in the study. If you do, please complete the attached consent form.
Thank you for your time.

Yours sincerely,

Sarah Alamoudi

Head of Department of the English Language Institute

C **Information Management and Policy Services**

I have read the information sheet about the project and received a copy of it.

I understand what the purpose of the project is and what is required of me. All my questions have been answered.

Name of Head of Department: _____

Name of Department: _____

Please tick as appropriate:

I consent to the involvement of my department in the project as outlined in the Information Sheet ☐

Signed:_____

Date: _____

DATA PROTECTION DECLARATION FOR ETHICAL APPROVAL

University of Reading

This document can be used to provide assurances to your ethics committee where confirmation of data protection training and awareness is required for ethical approval.

By signing this declaration I confirm that:

- I have read and understood the requirements for data protection within the *Data Protection for Researchers* document located here:

  http://www.reading.ac.uk/web/files/imps/Data_Protection_for_Researchers__Aug_18.v1.pdf

- I have asked for advice on any elements that I am *unclear on* prior to submitting my ethics approval request, either from my supervisor, or the data protection team at: imps@reading.ac.uk

- I understand that I am responsible for the secure handling, and protection of, my research data

- I know who to contact in the event of an information security incident, a data protection complaint or a request made under data subject access rights

Researcher to complete

Project/Study Title_____ **A Sequential Explanatory Design on Non- Semantic Transparent Linking Adverbials within Contrastive Analysis and Translation Setting**

NAME	STUDENT ID NUMBER	DATE
Sarah Alamoudi		08/01/2019

Supervisor signature

Note for supervisors: Please verify that your student has completed the above actions

NAME	STAFF ID NUMBER	DATE
Suzanne Graham		11/08/19

Submit your completed signed copy to your ethical approval committee.
Copies to be retained by ethics committee.

VERSION	KEEPER	REVIEWED	APPROVED BY	APPROVAL DATE
1.0	IMPS	Annually	IMPS	

University of Reading

ورقة معلومات المشارك

الباحثة : سارا العامودي

ايميل:

s.alamoudi@pgr.reading.ac.uk

المشرفين: البروفوسورة سوزان جراهام و الدكتورة ديزي باول

ايميل:

s.j.graham@reading.ac.uk; d.a.powell@reading.ac.uk

البحث: تعلم المفردات

عزيزتي الطالبة

أنا طالبة دكتوراه في جامعة ريدينج بالمملكة المتحدة. كجزء من مرحلة جمع البيانات في رسالتي ، أود أن أدعوك للمشاركة في دراسة بحثية حول تعلم المفردات.

لماذا دعيت للمشاركة؟

تمت دعوتك للمشاركة لأنك في برنامج اللغة الإنجليزية عالي المستوى في معهد اللغة الإنجليزية وأنا مهتم بمعرفة كيف يتعلم الطلاب في هذا المستوى المفردات.

ما الذي سيتضمن إذا وافقت على المشاركة؟

بموافقتك ، أود أولاً الوصول إلى نتائج اختبار تحديد المستوى لديك. خلال هذا الكورس ، سيُطلب منك إكمال بعض مهام المفردات القصيرة كجزء من مهام الفصل العادية. سيُطلب منك أيضًا إكمال مهمتين إلى أربع مهام كل أسبوع. تتضمن المهام ترجمة (الإنجليزية إلى العربية ومن العربية للإنجليزية) وأنشطة املائي الفراغ باللغة الإنجليزية. سيُطلب منك أيضًا إكمال استبيان حول تجربتك في تعلم اللغة الإنجليزية. يجب ألا تستغرق المهام والاستبيان أكثر من ١٥ دقيقة لإكمالها خلال وقت الفصل الدراسي ، وسوف تكملها مع الآخرين في الفصل. المعلومات التي تم جمعها من المهام لن تؤثر على المعدل التراكمي الخاص بك أو درجاتك الأكاديمية ولكن سيتم استخدامها في أطروحتي.

قد يتم اختيارك أيضًا للمشاركة في مقابلة مع الباحث في مناسبتين. ستكون المقابلة أيضًا مرتبطة بالدروس التي نتعلمها ويجب ألا تستغرق أكثر من ٢٠ دقيقة في وقت ومكان مناسب لك. ستكون المقابلة باللغة العربية وسيتم تسجيلها وترجمتها ونسخها

بإذن منك. سيتم عرض النسخة لك حتى تتمكن من التحقق من دقتها والتأكيد على أنك ما زلت سعيدًا لاستخدام محتوياته. سيتم استخدام المعلومات التي تم جمعها من قبل الباحث لتحليل البيانات.

هل سيعلم أحد عن إجاباتي؟

لا. سيظل اسمك وإجاباتك سرية ولن يراها إلا الباحث والمشرفون المدرجون في الجزء العلوي من الرسالة. سيتم تعيين رقم تعريف لك فقط لتمييز إجاباتك عن ردود المشاركين الآخرين. هذا المعرّف لا يرتبط بأي حال باسمك. وستبقى سجلات هذه الدراسة خاصة. لن يتم تضمين أي معرفات تربطك بالدراسة في أي نوع من التقارير التي قد تنشر.

ماذا سيحدث للبيانات؟

سيتم الاحتفاظ بأي بيانات تم جمعها بسرية تامة ولن يتم استخدام أسماء حقيقية في هذه الدراسة أو في أي منشورات لاحقة. وستبقى سجلات هذه الدراسة خاصة. لن يتم تضمين أي معرفات تربطك في أي نوع من التقارير التي قد تنشر. سيتم تخصيص رقم للمشاركين وسيتم الإشارة إليهم بهذا الرقم في جميع السجلات. سيتم تخزين سجلات البحث بشكل آمن في خزانة ملفات مغلقة وعلى كمبيوتر محمي بكلمة مرور والباحثة فقط سارا العمودي والمشرفون يمكنهم الوصول إلى السجلات. تماشيا مع سياسة الجامعة بشأن إدارة البيانات البحثية ، يمكن الحفاظ على البيانات مجهولة المصدر التي تم جمعها في هذا البحث وإتاحتها للجمهور للآخرين للتشاور وإعادة استخدامها. سيتم عرض نتائج الدراسة في المؤتمرات الوطنية والدولية وفي التقارير والمقالات المكتوبة. يمكننا أن نرسل لك نسخًا إلكترونية من هذه المنشورات إذا كنت ترغب في ذلك.

هل سأستفيد بالمشاركة؟

نتوقع أن تكون نتائج الدراسة مفيدة للمعلمين والطلاب في تخطيط كيفية تعلم المفردات.

هل يجب علي المشاركة؟

لا ، ليس على الإطلاق. اختيارك المشاركة هو تطوعي. إذا قررت عدم المشاركة ، فسوف تكون في الفصل أثناء جلسات القراءة والمهام ولكنك لن تأخذ أي اختبارات متعلقة بالدراسة. بدلاً من ذلك ، ستتلقى أسئلة فهم تتعلق بكتبك ولن يتم استخدام بياناتك. أيضًا ، أنت حرة في مغادرة المشروع البحثي في أي وقت ، دون إبداء سبب. يمكنك القيام بذلك عن طريق إخباري أنك لا تريد المشاركة أو المتابعة.

نأمل أن توافق على مشاركتك في الدراسة. إذا قمت بذلك ، يرجى ملء نموذج الموافقة المرفق وإعادته لي.

تمت مراجعة هذا المشروع البحثي باتباع إجراءات لجنة أخلاقيات البحوث الجامعية وتم منح رأي أخلاقي مناسب بشأن السلوك. الجامعة لديها التأمينات المناسبة في المكان. متوفرة عند الطلب التفاصيل الكاملة.

التوقيع: سارا العامودي
التاريخ: ٢٠١٩/٨/٥

University of Reading

نموذج الموافقة

الباحثة : سارا العامودي

ايميل:

s.alamoudi@pgr.reading.ac.uk

المشرفين: البروفوسورة سوزان جراهام و الدكتورة ديزي باول

ايميل:

s.j.graham@reading.ac.uk; d.a.powell@reading.ac.uk

البحث: تعلم المفردات

لقد قرأ وشُرح لي من قبل سارا العامودي عن ورقة المعلومات المتعلقة بهذا البحث.

لقد شُرح لي أغراض البحث وما هو مطلوب مني ، وتم الرد على الاسئلة كافة. أوافق على الترتيبات الموضحة في ورقة المعلومات من حيث صلتها بمشاركتي.

أفهم أنه سيُطلب مني إكمال مهام المفردات واستبيان في الفصل.

أتفهم أنه قد يتم إجراء مقابلات معي وأن المقابلة سيتم تسجيلها وترجمتها ونسخها.

أدرك أن مشاركتي تطوعية تمامًا ولديّ الحق في الانسحاب من المشروع البحثي في أي وقت ، دون إبداء أسباب ودون أي تداعيات.

لقد استلمت نسخة من نموذج الموافقة هذا وورقة المعلومات المرفقة.

يرجى وضع علامة صح على الآتي:

أوافق على المشاركة في البحث	نعم _____	لا _____
أوافق على الاطلاع لنتيجة اختبار تحديد المستوى	نعم _____	لا _____
وأوافق على إجراء المقابلات	نعم _____	لا _____
أوافق على تسجيل هذه المقابلة الصوتية	نعم _____	لا _____

الاسم: