REPUBLIC OF TURKEY
ÇUKUROVA UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LANGUAGE TEACHING

A LEARNER CORPUS-BASED STUDY ON THE USE OF ENGLISH PREPOSITIONAL VERBS OF TURKISH EFL LEARNERS

Sibel AYBEK

MASTER OF ARTS

ADANA / 2018

REPUBLIC OF TURKEY
ÇUKUROVA UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LANGUAGE TEACHING

A LEARNER CORPUS-BASED STUDY ON THE USE OF ENGLISH PREPOSITIONAL VERBS OF TURKISH EFL LEARNERS

Sibel AYBEK

Supervisor:	**Assoc. Prof. Dr. Cem CAN**
Jury Member:	**Prof. Dr. Hatice SOFU**
Jury Member:	**Asst. Prof. Dr. Meltem MUŞLU**

MASTER OF ARTS

ADANA / 2018

To Çukurova University Institute of Social Sciences;

We certify that this thesis is satisfactory for the award of the degree of Master of Arts in the Department of English Language Teaching.

Supervisor: Assoc. Prof. Dr. Cem CAN

Member: Prof. Dr. Hatice SOFU

Member: Dr. Meltem MUŞLU

ONAY

Yukarıdaki imzaların, adı geçen öğretim elemanlarına ait olduklarını onaylarım.

…/…/2018

Prof. Dr. Serap ÇABUK

Enstitü Müdürü

ETİK BEYANI

Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde ve ortaya çıkan sonuçlarda herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim. / / 2018

Sibel AYBEK

ÖZET

YABANCI DİL OLARAK İNGİLİZCE ÖĞRENEN TÜRK ÖĞRENCİLERİN İNGİLİZCE İLGEÇLİ EYLEM KULLANIMLARI ÜZERİNE DERLEME DAYALI BİR ÇALIŞMA

Sibel AYBEK

Yüksek Lisans Tezi, İngiliz Dili Anabilim Dalı
Danışman: Doç. Dr. Cem CAN
Ağustos 2018, 101 sayfa

Yabancı dil olarak İngilizce alanında öğrenici derlemlerinin yaygınlaşması ile birlikte, öğreniciler tarafından hedef dilde yazılmış metinleri kullanarak aradil hatalarını incelemek ve çözümlemek olası hale gelmiştir. Öğrenici derlemlerinin oluşturulması ile elde edilen veriler, bağlamları içerisinde düzenli bir örüntü içerisinde oluşan özgün hataların bireysel sıklıklarını tür ve türce bazında çözümleyerek, araştırmacılara aradil sürecinde oluşan hedef dilden sapmaları inceleme ve yorumlama olanağını sağlamaktadır. Bu özgün aradil hatalarının, dil öğretim öğrenceleri ve çözüme yönelik kaynak üretmede göz önüne alınmasının gerekliliği birçok araştırmada önemle vurgulanmıştır (örneğin, Juozulynas, 1994; Mitton, 1996; Cowan, Choi, & Kim, 2003; Ndiaye & Vandeventer Faltin, 2003; Allerton et al., 2004). Bu çalışma, Avrupa Dilleri Ortak Çerçeve Programı (Council of Europe, 2001) tarafından belirlenmiş olan altı dil yeterlilik düzeyinde (A1-A2; B1-B2; C1-C2) yabancı dil olarak İngilizce öğrenen Türk öğrencilerin yazılı dillerinde yaptıkları İngilizce ilgeçli eylem öbeklerinin kullanım hatalarına odaklanarak incelemeyi ve bu hataların kaynaklarına yönelik çözümleme yapmayı amaçlamaktadır.

İlgeçli eylemler (PV'ler), Baldwin (2005) 'e göre çoklu sözcük öğeleri (MWIs) veya çok sözcüklü ifadeler (MWE) adı verilen sözcükler ulamı altında sınıflandırılmaktadır. Bu sözlüksel birimler, deyimler, öbeksi eylemler, ilgeçli eylem öbekleri, basmakalıp sözler ve kalıplaşmış anlatım dizilerini içerir. Bunlar, zihinsel sözlükçede iki veya daha fazla sözcük ile oluşturulur, yabancı dilde bunların kullanımı ile öğrenciler Gardner & Davies'in (2007) de dediği üzere anadil kullanımına daha çok yaklaşmaktadırlar. Bu alandaki alanyazın ikinci dil öğreniminde ilgeçli eylemler öbeklerinin önemini vurgulamaktadır (Tetreault & Chodorow, 2008; Hong, Rahim, Hua

ve Salehuddin, 2011); Bununla birlikte, Özel ve Akademik Amaçlı yabancı dil yazılı anlatım alanlarında bu eylemlere yönelik çalışmalar yeterince yapılmamıştır. Bu çalışmanın yaklaşımı bilgisayar derlembilimi yöntemlerini kullanarak ve Corder'ın (1973) önerdiği Hata Çözümlemesini çerçevesinde araştırma yapmayı kapsamaktadır. Corder'ın bu yaklaşımda önerdiği adımlar hatayı belirleme, tanımlama, sınıflandırma ve açıklama olarak belirlenmiştir. Çalışmada veri tabanı olarak kullanılacak derlem, alandaki en geniş yabancı dil olarak İngilizce öğrenici derlemi Cambridge Learner Corpus olacaktır. Derlemin Türk öğrencilerin 1992 yılından beri yapılmış tüm Cambridge dil sınavlarındaki yazılı bölümleri çözümlemeye alınacak ve SkecthEngine derlem sayısal çözümleme araçları ile hatalar ve bağlamları incelenecektir.

Anahtar Kelimeler: Öğrenici derlemleri, hata çözümlemesi, yabancı dil olarak Ingilizce öğretimi, ilgeçli eylem öbekleri

ABSTRACT

A LEARNER CORPUS-BASED STUDY ON THE USE OF ENGLISH PREPOSITIONAL VERBS OF TURKISH EFL LEARNERS

Sibel AYBEK

Master Thesis, Department of English Language Teaching
Supervisor: Doç. Dr. Cem Can
August 2018, 101 pages

As learner corpora have presently become readily accessible, it is now practicable to examine interlanguage errors and carry out error analysis (EA) on learner-generated texts. The data available in a learner corpus enable researchers to investigate authentic learner errors and their respective frequencies in terms of types and tokens as well as contexts in which they regularly occur. The need to consider these authentic learner errors in the design of useful language learning programs and remedial teaching materials has been widely emphasized by many researchers (see e.g., Juozulynas, 1994; Mitton, 1996; Cowan, Choi, & Kim, 2003; Ndiaye & Vandeventer Faltin, 2003; Allerton et al., 2004). This study aims at analyzing the errors pertaining to the use prepositional verbs by Turkish EFL learners across six distinct proficiency levels, A1-A2; B1-B2; C1-C2, as defined by Common European Framework of Reference for Languages (henceforth CEFR) (Council of Europe, 2001).

Prepositional verbs (PVs) are classified under the category of lexical items called multiword items (MWIs), or multiword expressions (MWEs) according to Baldwin (2005). These lexical items include idioms, phrasal verbs, prepositional verbs, stock phrases, prefabrications, and formulaic sequences. They are built with two or more words in the mental lexicon enabling learners to sound more native-like claimed by Gardner & Davies (2007). The literature in this area highlights the importance of prepositional verbs in L2 learning (Tetreault & Chodorow, 2008; Hong, Rahim, Hua, and Salehuddin, 2011); however, there is a lack of empirical studies of these verbs in ESL and EAP writing. The corpus to be used in this particular study is the Cambridge Learner Corpus (CLC), the largest annotated test performance corpora which enables the investigation of the linguistic and rhetorical features of the learner performances in CEFR proficiency bands.

Approach of this study uses the techniques of computer corpus linguistics and has its roots in the Error Analysis framework as proposed by Corder (1973): identification, description, classification and explanation of errors.

Keywords: Learner corpus, error analysis, English as a foreign language, prepositionals verbs

ACKNOWLEDGEMENTS

First and foremost, I would like to express my greatest appreciation to my supervisor, Assoc. Prof. Cem Can, for his commitment, patient guidance, and limitless effort in any way. This work would not have been possible without him.

I also would like to offer my special thanks and gratitude to the examining committee, Prof. Dr. Hatice Sofu and Asst. Prof. Dr. Meltem Muşlu for their valuable feedback and comments and their commitments by coming my thesis defence on this very hot Adana day.

I owe special thanks to each of my teachers at the department and also Dr. Dilek Yavuz Erkan who always believed, supported and encouraged to make me believe what I have been doing.

Finally, I should also explain my gratitude to my lovely cat son, Çiko, for sleeping and snoring on my lap while I was writing this thesis.

This thesis has been supported by Cukurova University Scientific Research Projects Unit with 10017 ID and SYL-2018-0017 code

Sibel AYBEK

Adana / 2018

CONTENTS

CHAPTER V
DISCUSSION

CHAPTER VI
CONCLUSION AND SUGGESTIONS

LIST OF ABBREVIATIONS

EA: Error Analysis

TEA: Traditional Error Analysis

EFL: English as Foreign Language

ESL: English as Second Language

FLT: Foreign Language Teaching

CEFR: Common European Framework of Reference for Languages

PVs: Prepositional Verbs

MWIs: Multiword Items

MWEs: Multiword Expressions

CLC: Cambridge Learner Corpus

BNC: British National Corpus

DAT.: Dative

LOC.: Locative

ABL.: Ablative

L.SG.: First Person Singular

ABIL.: Abilitative

GEN.: Genitive

PR.PROG: Present Progressive

ACC.: Accusative

INF.: Infinitive

CEA: Computer-Aided Error Analysis

LSP: Language for Specific Purposes

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1. Background to the Study

1.1.2. Corpus Linguistics

The definition of corpus linguistics can be stated as a linguistic methodology which is based on the use of electronic collections of authentically occurring texts, to wit corpora, according to Granger (2002). In her study, she states that it is not a new branch or theory of language; however, its authenticity makes it a particularly efficacious methodology, which is likely to alter perspectives on language. Similarly, Leech (1991) also defines corpus linguistics not as a simple methodology, but more like a methodology rather than a stand-alone scientific domain. According to him, a developing collection of computer tools for searching, annotating, retrieving, and analysing electronic text data, such as taggers, parsers, and concordancer, have been enabled by using a corpus (Leech, 1992). He also asserts that "it is from these tools, as much as from the availability of abundant text data, that Corpus Linguistics has derived its special character and its unprecedented power to reveal the characteristics of real language use" (p. 158). For this reason, Corpus Linguistics could be considered as a methodologically-oriented branch of linguistics.

Within the same line of reasoning, pertaining to the ongoing debate in the field about whether Corpus Linguistics is a methodology, or it is more than a set of methods for dealing textual corpora, Leech (1992), McEnery and Wilson (1996), McEnery, Xiao and Tono (2006) refer to it as a methodology; while Hoey (1997) sees it as "the route into linguistics" (p. 6); Biber, Conrad and Reppen (1998) call it as "an approach", and Scott and Thompson (2001) describe it as "a means for accessing resources" (p. 36).

1.1.3. Learner corpora

Over the past two decades, as Ludeling and Kytö (2009) describes, corpus has been defined as a large taxonomic compilation of written and/or spoken language stored on a computer and used in linguistic analysis, and corpus data bases not only provide opportunities for linguistics researches, but also provide sources for researches about teaching and learning of languages. Bernardini (2004) also indicates that recent

technological improvements, such as high capacity hard drives, fast processors in computing technology, constructing, downloading, buying, and/or accessing online texts have led the corpora to be accessed through easier and feasible ways. As he also highlights, (2004, p. 21) "More importantly perhaps, corpus has become less of a buzzword and more of a necessary, acknowledged reference source for students, linguists, language professionals (teachers, translators, technical writers, lexicographers etc.)" The way corpora adapted and/or made appealing to most students has led its implementation perceived to be as a discovery learning for many teachers, which brought data-driven teaching to the classroom setting along with.

Granger (2002) observes corpus-based studies of non-native varieties have been a recent divergence: it was not until the late 1980s and early 1990s that academics and publishers started compiling corpora of non-native Englishes, which is now referred as learner corpora.

Sinclair (1991) describes this relatively new type of corpora as "…electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance" (p. 2). The data in such corpora are typically the "data coming from the same genre," like argumentative student essays as the ones in International Corpus of Learner English (ICLE), or "data elicited from the same prompt," like a picture description task (Wulff, 2017).

1.1.4. Corpus and ELT

Analysing learner corpora helps us review how second language development is processed as O'Keeffe et al. (2007) posit and understood which might change "long-held" notions of education and pedagogy bridging the gap between the cognitive science of linguistics and many other areas, such as sociolinguistics and translation. Syllabus design has benefited from the studies on learner corpora, according to Aston (2000), as they provide insights on "the needs of specific learner populations" (Meunier 2002, p. 125).

When we consider the contributions of corpus studies in the material development field in teaching English as a foreign language (EFL), we see that the inadequacy of traditional prescriptive grammars has been addressed by corpus-based descriptive grammar approach like the *Longman Grammar of Spoken and Written English* (Biber et al., 1999) by providing corpus-derived perspective and information about frequencies.

These suggest that even corpus-based grammars do not propose the same potential as corpora to be able to develop language skills (some of which learning in general, and autonomous learning in non-institutional settings in particular would seem to rely) to "identify – classify – generalise" on the basis of language experience implementing corpus-based teaching approaches as displayed in recently emerging Data-driven Teaching techniques.

1.1.5. Data-driven Teaching

In order to make data-driven teaching possible for teaching and learning, language learners and teachers are supposed to start applying corpora and concordances themselves and discover about language patterning and the behaviour of words and phrases in an "autonomous" way rather than relying on a researcher as mediator or provider of corpus-based materials (Bernardini 2002, p. 165). Johns' (1991) study on *data-driven learning* established extremely influential and ground-breaking method to indicate how relevant corpus analysis techniques are for the wide and varied audience of language teachers and students around the world. Johns (1991) propounds that learners need to be guided to *discover* the foreign language, much in the same way as corpus linguists discover facts of their own language. Additionally, he also refers to the learner as a "language detective" and coined catchword "Every student a Sherlock Holmes!" (Johns, 1997, p. 101). According to Johns (1986; 1994), this approach, in which there is either an interaction between the learner and the corpus or, in a more controlled way, between the teacher and the corpus is now known as "data-driven learning" or DDL. Language learners can practice DDL activities based on generally (larger) general reference corpora or on (smaller) specialised corpora based on their specific purposes.

Particularly proposed for teaching writing, Leech (1997) suggests a similar viewpoint claiming that the critical and argumentative "type of essay assignment [. . .] should be balanced with the type of assignment [. . .] which invites the student to obtain, organize, and study real-language data according to individual choice" (p. 5). The latter type of task provides the learners with realistic expectations of breaking new ground as a 'researcher' and also, doing something which is a unique and individual contribution instead of reworking and evaluating of the research of others. This shift of emphasis from deductive to inductive learning methods has various effects on: (a) the teacher, who becomes a coordinator of research, or facilitator; (b) the learner, who learns how to learn

through exercises that involve the observation and interpretation of patterns of use; (c) the role of pedagogic grammars, whose level of abstraction often works against their effectiveness (Bernardini, 2004).

1.1.6. Importance of Error Analysis in ELT

Corder's "The significance of learner's errors" (1967) has gained quite popularity leaving Constrastive Analysis somewhat questionable. Corder (1967) puts forward several reasons why error analysis is useful and how it gives insight about the ways learners learn their second language. He (1967) explains that errors characterise our linguistic competence and can expose some of the learners' learning strategies. Determining whether the semantic content and the linguistic form of learners' communicative performances are erroneous or deviant from the norms are enabled by errors. Corder (1967, 1973 & 1974) classifies the errors into four different categories which are as follows:

a. Addition
b. Omission
c. Selection
d. Ordering

Error analysis (EA) has gained a significant impetus to Second Language Acquisition (SLA) research. Granger (2002) notes that practicing current EA is quite different from that of the 1970s. Decontextualization of errors, ignoring for learners' correct use of the language and non-standardised error classifications were the characteristics of the first EA, whereas today's EA focuses on contextualised errors along with the context of use and the linguistic context (co-text). Linguistic items with errors could be provided with one or more sentences, a paragraph, or the whole text, along with correct exemplifications (Granger, 2002).

Interlanguage Errors are essential parts of EFL learning processes just as error correction is an integral part of the teaching process and, unlike mistakes, systematic. According to Corder (1967), they map the EFL learner's linguistic competence and could lead us to obtain the clues about learners' learning strategies as well as processes. Detecting errors entails an arduous process related to deciding whether the semantic

content and the linguistic form of learners' communicative performances are deviant from the norms. This process can also be realised by studying related anecdotal evidence or by classroom observation as claimed by Swan and Smith (2001). However, as it requires a large number of samples of error types, a learner corpus is a more dependable option to examine the naturally occurring learner errors. Moreover, a learner corpus assists the researchers to pinpoint error typologies of learners from different native language backgrounds enabling a systematic coparison. Focusing on specific points, a specialized corpus could be utilised to advise language teaching. As Sinclair (2001) proposes, "descriptions of structure, reliable models of usage, how words and phrases are translated, determining the essentials in a syllabus, and identifying learners' errors can all be developed and/or supported by focusing on specific genres and subgenres" (p. 289).

Granger (2003) emphasizes on two methodological approaches which are generally involved in linguistic exploitation of learner corpora: Contrastive Interlanguage Analysis and Computer-aided Error Analysis. Contrastive Interlanguage Analysis (CIA) is a contrastive method carrying out comparisons of quantitative and qualitative data between native (NS) and non-native (NNS) varieties or between different varieties of non-native data.

1.1.7. Contrastive Interlanguage Analysis (CIA)

Interlanguage studies across various L1 backgrounds based on learner corpora focus on what Granger (2002) calls "Contrastive Interlanguage Analysis (CIA)", comparing learner data and the data produced by native speakers of the target language, or the learner's L1. The first type of comparison attempts to evaluate the level of under- or overuse of particular linguistic features in learner language. The second type tries to uncover L1 transfer or interference.

As mentioned above, Contrastive Interlanguage Analysis (CIA) involves two types of comparison. Lorenz (1999) indicates that NS/NNS comparisons aim to shed light on non-native characteristics of learner writing and speech with detailed comparisons of linguistic features in native and non-native corpora. CIA also involves NNS/NNS comparisons. According to Lorenz (1999), "By comparing different learner populations, researchers improve their knowledge of interlanguage" (p. 12). Also, comparisons of learner data of different L1 backgrounds reveal interlanguage universals shared by various learner populations for researchers. These features may be developmental and L1

dependent. For example, in Granger and Tyson's (1996) study (of connectors), overuse of sentence-initial connectors was determined to be developmental, and it was also indicated to be an interlanguage characteristic of the three learner populations which are French, Dutch, and Chinese. The individual connectors that has a wide variation of use among these learner groups provide evidence for interlingua characteristics.

According to CA hypothesis, Dagneaux et al., (1998) describe that communication strategy-based errors may occur when the learner seeks to convey a spontaneous speech with an inadequate grasp of the target language system, i.e. with his/her interlanguage. The holistic strategy can also be termed as approximation, that is, when learners lack the demanded form, they tend to use a near-equivalent L2 item they have learnt. For example, in their study with the Chinese students, some uses of similar synonyms; (*decide for determine), super-ordinate term (*flower for roses), *face missing work (face losing a job), *learn to everybody (learn from everybody), etc. According to Dagneaux et al., (1998) traditional EA has still been a worthwhile enterprise. Its basic principles hold but its methodological weaknesses need to be covered.

Generating comprehensive lists of specific error types, counting, and sorting them in various ways and tagging them in their context along with their instances of non-errors could be carried out by Computer-Aided Error Analysis (CEA). Completing the gaps in traditional EA, it is a powerful technique assisting ELT teachers and syllabus designers in their production of innovative and remedial pedagogical tools by being more "learner aware" as well as "noticeable".

1.2. Statement of the Problem

To be able to provide the most useful input to the learners, teachers and material designers need large amount of information about what learners may be expected to have acquired by what stage according to Granger et al. (2002). For this reason, analysing interlanguage errors is a valuable source of information. In fact, it is not meant that classroom activities need focusing on errors, however, more 'learner-aware teaching' could be advantageous.

Can (2017) indicates that analysing interlanguage and performing error analysis (EA) on learner-generated texts are practicable thanks to readily accessible learner corpora. Authentic learner errors and their respective frequencies in terms of types and tokens as well as contexts to be investigated are facilitated through the data available in

the learner corpora. The necessity of the consideration of authentic learner errors needs to be considered while forming effective language learning programs as well as remedial teaching materials which has also been widely emphasized by many researchers (Juozulynas, 1994; Mitton, 1996; Cowan, Choi, & Kim, 2003; Ndiaye & Faltin, 2003; Allerton et al., 2004).

Prepositional verbs (PVs) have been among one of the most common interlanguage errors of Turkish EFL learners. Analysing PVs on such an authentic data in terms of gaining insights about learners' stages and strategies is needed. That's why, this study aims at analysing the errors pertaining to the use of prepositional verbs by Turkish EFL learners across six distinct proficiency levels, A1-A2; B1-B2; C1-C2, as defined by Common European Framework of Reference for Languages (henceforth CEFR) (Council of Europe, 2001). This study's approach uses the techniques of computer corpus linguistics and has its roots in the Error Analysis framework as proposed by Corder (1973): identification, description, classification, and explanation of errors.

1.3. Purpose of the Study and The Research Questions

The aim of this explorative study is to obtain empirical evidence of the use of PVs in Turkish EFL learners' written productions across A1-C2 proficiency levels according to CEFR as displayed in CLC. Using the Sketch Engine Tools, Turkish EFL learners' use of PVs have been extracted and then analysed for their frequency and accuracy.

Research questions of the study in this regard include:

1. How frequent are PVs used in the written productions by Turkish EFL learners, as represented in their sub-corpus of the CLC?
2. What are the most frequent PVs used by Turkish EFL learners across CEFR proficiency levels?
3. To what extent are PVs found in Turkish EFL learner's corpus well-formed across CEFR proficiency levels, i.e., verbs paired with appropriate preposition?
4. What PVs, if any, seem to be most problematic for Turkish EFL learners across CEFR proficiency levels?

1.4. Significance of the Study

In their study of L2 collocations, Laufer and Waldman (2011) mention three most common ways to examine interlanguage errors which are traditional analysis of errors in

samples, eliciting collocations by various data collection methods, and utilizing learner corpora.

Although they have essential differences in terms of their methodologies, theoretical assumptions, data and terminology, Contrastive Analysis (CA) and Error Analysis (EA) studies in interlanguage research try to attain the same goal of sheding light on the language learning process to understand it better. The essential difference between them lies in the approach to the learner's performance and the errors. While CA studies do not take any position on this issue, traditional EA considers errors to be detrimental and tries to eliminate them. Combining the core principles of interlanguage (IL) in which the deviations from the TL norms are considered as integral components and indicators of the learner's system, computer-aided error analysis, based on learner corpora, on the other hand, “represents a major improvement over traditional EA, not least because instead of considering errors in isolation, but also it sees them as part of a system which includes both correct and incorrect uses” (Gilquin and Granger, 2015, p. 427), combining the underlying principles of IL and EA. Referring to the use of learner corpus based error analysis and the weaknesses of traditional error analysis, Ellis (1994) stresses "the importance of collecting well-defined samples of learner language so that clear statements can be made regarding what kinds of errors the learners produce and under what conditions" and claims that "many EA studies have not paid enough attention to these factors, with the result that they are difficult to interpret and almost impossible to replicate" (p. 49). Dagneaux et al., (1998) also criticizes the nature of data and related error typologies used by traditional EA and lists five limitations in this regard:

- Limitation 1: EA is based on heterogeneous learner data;
- Limitation 2: EA categories are fuzzy;
- Limitation 3: EA cannot cater for phenomena such as avoidance;
- Limitation 4: EA is restricted to what the learner cannot do;
- Limitation 5: EA gives a static picture of L2 learning (p. 164).

According to Schachter (1974), EA can only be used to identify and analyse errors that learners actually produce and ignores potential errors that learners do not make because they avoid difficult structures. Supporting these claims, Gilquin (2007) affirms that traditional error analysis results could reveal only a partial picture of the learners’ proficiency and target-like language use is much more than the absence of errors. She

emhasizes that "[A]n error analysis . . . only lifts a corner of the veil. Equally important are indications as to what learners get right, what they underuse and what they overuse" (2007, p. 288). Because of these limitations, learner-corpus-based approach to error analysis has been considered as a new direction in the field with its pedagogical implications being perceptible as a useful guidance for Foreign Language (FL) teaching. For this reason, there is an academic and pedagogic need for more research into the field of learner corpora in order to raise discussion concerning the linguistic patterns of the interlanguage produced by EFL learners.

Being one of these patterns, according to Biber et al. (1999), PVs appear to be more common in written registers than their phrasal verb counterparts. Due to their complex behaviour, PVs display a challenge for EFL learners. Tetreault & Chodorow (2008) claim that one of the causes of such difficulty stems from the use of prepositions. Also, it is a set of forms that most L2 learners find difficult according to Hong et al. (2011). Because of their patterning challenge, even the native speakers of English have difficulty in choosing the correct preposition in a cloze-test task, yielding a 75% success rate.

Along with speaking, reading, and listening, writing is another crucial skill for all L2 learners. Therefore, it is necessary to obtain understanding of how these verbal patterns are learnt and used in authentic written learner interlanguage. Considering the previous researches, this study aims to contribute to research and pedagogy in relation to PVs in EFL writing. Using a bottom-up methodological approach, this research also seeks to contribute to the current literature by shortening the gap between the theory of second language learning and the interlanguage that students actually produce.

For this reason, focusing on the prepositional verbs, this corpus-based study aims at analyzing the errors by Turkish EFL learners across six distinct proficiency levels, A1-A2; B1-B2; C1-C2, as defined by CEFR.

CHAPTER II

REVIEW OF LITERATURE

2.1. Corpus Linguistics: Methodology or not?

In order to decide whether Corpus Linguistics is a methodology or not, Sardinha (2011, p. 33) suggests determining "who uses it and for what purpose it is used." According to him, it can be considered as a method, which is a set of procedures for collecting and analysing data, as it could be used with several different theories. Moreover, theoretical statements could be put forward, which makes corpus linguistics far more than a simple method. More importantly, Sardinha (2011) emphasizes that "collocation, semantic preference, semantic prosody, dimensions of register variation among others are all theoretical concepts that were either 'discovered' or made evident by means of electronic corpus analysis. These are theoretical categories which are universal to all languages justifying the theoretical status of Corpus Linguistics.

According to Leech (1992, p.105) corpus linguistics is "a methodological basis for doing linguistic research." Supporting this claim, Kennedy (1998) states that "corpus linguistics has come to embody methodologies for linguistic description in which quantification is part of the research activity" (p. 7).

Developments in hardware as well as software technology enabled corpus linguistics to be highly efficacious in computing and text processing procedures. As a result, the researchers have become equipped to compile much larger corpora, tag, annotate, and analyse the texts faster and at a higher accuracy rate. However, it is arguable that corpus linguistics is tool-driven (Rayson & Archer 2008, p. 78), in other words, users are "counting only what is easy to count" rather than what they would like to count (Stubbs and Gerbig, 1993).

In text analysis, concordancing, being a corpus tool, has not only proved to be an effective way "to mimic the effects of natural contextual learning" (Cobb, 1997, p. 314), but also its use and usefulness for error correction in foreign or second language writing have been highlighted by many studies (Bernardini, 2004; Chambers, 2005; Gaskell and Cobb, 2004; Gray, 2005).

The corpus research process involves three main stages: corpus compilation, annotation, and retrieval (Rayson & Archer, 2008). Firstly, a corpus needs to be compiled by transcription, scanning, or sampling from online sources. The second stage is

annotation which combines manual and automatic methods to add tags, codes, and documentation identifying textual and linguistic characteristics. Kennedy (1998) also states that there exist three stages to corpus compilation: “corpus design, text collection or capture, and text encoding or mark-up” (p. 70) and gives clues about the importance of categorization with safe backup and storage. Adolphs (2008) also menions three stages in the corpus compilation process: “data collection, annotation and mark-up, and storage” (p. 21).

In learner corpus, for example, meta-information about the learners such as; learners’ L1, age, sex, education history and number of years the learner has studied English for are also accommodated in each text. Creating sub corpora and representing certain subsets of learners such as speakers of certain mother tongues or certain error types i.e. content word combinations as well as the error coding may be prepared with this information (Kochmar, 2016).

Design criteria of compiling a learner corpus should be clearly defined in the beginning. Dagneaux et al., (1998) highlights that compiling learner corpora needs to be in accordance with general principles of corpus linguistics. Variables pertaining to the learner such as age, language background, learning context, etc. and the language situation i.e. medium, task type, topic, etc. should also be recorded and can subsequently be used to compile homogeneous sub corpora.

2.1.1. Text Retrieval

Text retrieval software has been a kind of computer program that achieved the most astonishing results. As Rundell and Stock (1992, p. 14) point out, text retrieval software enables linguists to “focus their creative energies to doing what machines cannot do”.

Text retrieval software such asWordSmith Tools does not only count words, but it also counts word partials and sequences of words, which it can sort into alphabetical and frequency order. Another valuable option is the ‘concord’ option since it depicts the collocates or patterns that learners use, correctly or incorrectly. Furthermore, another option called ‘compare lists’ makes it possible for researchers to carry out comparisons of items in two corpora and bring out the statistically significant differences. If native and learner language are represented in these two corpora, then, this kind of a comparison will enable the analyst to access to those items which are either under- or overused by learners.

In the 1990s, increased processing potentialities of computers led to another (third) generation of retrieval software systems such as WordSmith, MonoConc, AntConc, and Xaira. (Rayson, 2015, Chapter 2). These could deal with corpora of tens of millions of words, comprising of languages other than English, and included implementations of the other methods in sync rather than as separate tools. Corpus retrieval software has moved to web-based interfaces referred as the fourth generation. Pre-existing corpora rather than texts collected by the users are only permitted by most of the web-based interfaces. To exemplify, Mark Davies' 'corpus.byu.edu' interface allows access to quite large corpora which include 450 million words of the Corpus of Contemporary American English (COCA), 400 million words of the Corpus of Historical American English (COHA), 100 million words of the Time Magazine Corpus, and 100 million words of the British National Corpus. Other systems generally rely on the Corpus Query Processor (CQP) server (part of the Open Corpus Workbench) or Manatee server (Rayson, 2015, Chapter 2) and BNCweb (providing access to the British National Corpus), SketchEngine (aimed at lexicographers), and CQPweb (based on the BNCweb design but suitable for use with other corpora) are the well-known web-facing front ends. Similarly, other web-based tools are Intellitext (aimed at humanities scholars), Netspeak, and ANNIS. In addition, the web-based Wmatrix software (Rayson, 2008) permits the users to perform retrieval operations as well as annotating uploaded English texts with two levels of corpus annotation: part-of-speech and semantic field.

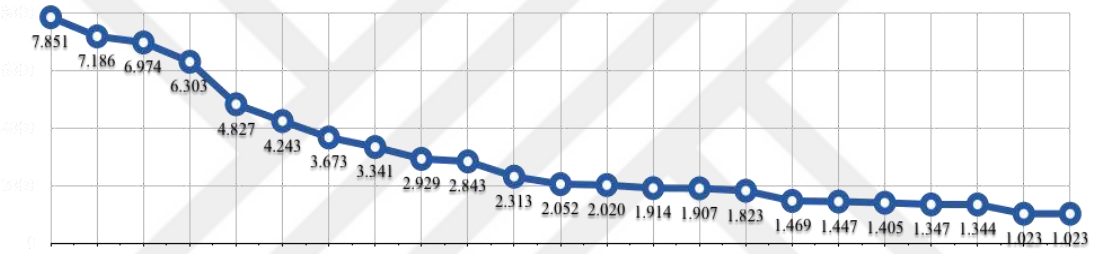
2.1.2. Annotation Process

At annotation stage, linguistic features investigated such as morphological, stylistic, discourse, lexical, pragmatic, syntax or semantic aspects may change forms (Rayson, 2008) and manual (human-led) and/or automatic (machine-led) methods could also be used at annotation. Adding annotation is crucial because the information encoded in annotations provide the researcher with linguistic information available in the corpus for later retrieval or extraction processes. Rayson (2008), referring to this process, states that "…Computational methods and tools for corpus annotation therefore take two forms. First, intelligent editors to support manual annotation and second, automatic taggers which apply a particular type of analysis to language data" (p. 39).

2.1.3. Part-of-speech Tagging

One of the good examples of fully automatic annotation is part-of-speech (POS) tagging. Each word is assigned a tag by a POS tagger in a corpus indicating its word class. SLA/FLT researchers use this to conduct selective searches of particular parts of speech in learner language, especially error-prone categories like prepositions or modals.

Part-of-speech taggers claim a high overall success rate along with many advantages including being fully automatic, widely available and inexpensive. The value of using POS-tagged and error tagged learner corpora is exemplified clearly in figure 1, which lists the most common error typologies found in the learner corpus of Turkish learners (Can, 2017). In that study, error-tagged CLC was analysed using SketchEngine in order to find the most common errors in the interlanguage of Turkish EFL learners.



Error Code	Freq.	Error Code	Freq.	Error Code	Freq.
RP: Replace Punctuation	7851	RN: Replace Noun	2843	AGN: Noun Agreement Error	1347
MD: Missing Determiner	7186	UD: Unnecessary Determiner	2313	RJ: Replace Noun	1344
MP: Missing Punctuation	6974	R: Replace Error	2052	SX: Spelling Confusion	1023
TV: Incorrect Tense Verb	6303	UT: Unnecessary Preposition	2020	MC: Missing Link Word	1023
RV: Replace Verb	4827	**FV: Wrong Verb Form**	1914		
RT: Replace Preposition	4243	FN: Wrong Noun Form	1907		
UP: Unnecessary	3673	**MV: Missing Verb**	1823		
MT: Missing Preposition	3341	**AGV: Verb Agreement Error**	1447		
W: Word Order Error	2929	RY: Replace Adverb	1405		

Figure 1. The most common interlanguage errors of Turkish EFL learners - A1-C2 (Can, 2017)

Here we see that preposition errors are among the most common error typologies. That's why, in our study, we aimed at analyzing PV errors of Turkish EFL learners across A1-C2 proficiency ranges according to CEFR.

2.2. The Role of Corpora in SLA Studies

SLA (Second Language Acquisition) research, which seeks to explore the mechanisms of language acquisition process, and FLT (Foreign Language Teaching)

research, the aim of which is to improve the learning and teaching foreign/second languages can be informed through a new kind of data provided by learner corpora (Granger, 2003). Collecting and storing learner data in large quantities and analysing this data automatically or semi-automatically with presently available linguistic software have been provided with technological developments.

Sinclair (1991), referring the value of 'authenticity' of data in SLA, states that in corpus driven SLA studies "all the material is gathered from the genuine communications of people going about their normal business 'unlike data gathered' in experimental conditions or in artificial conditions of various kinds" (p. 3). According to Sinclair (1991), applying to the foreign/second language field, this means that experimental data gathered through elicitation techniques cannot be qualified as learner corpus data.

Within the same line of reasoning, Atkins and Clear (1992) underline that learner corpora address two non-native varieties: EFL (English as a Foreign Language) and ESL (English as a Second Language). They also describe learner corpora as 'continuous stretches of discourse' which comprise both erroneous and correct use of the language.

Furthermore, Meyer (2002) emphasizes that various usages in both as theoretical and practical domains make corpora valuable resources for theoretical, descriptive, and applied discussions of language. As corpus linguistics is a methodology, in the study of language, corpora may be used by all linguists and even generativists. Corpora have also proved to be an indispensible source of reference in language teaching material development, understanding the process of language acquisition, studying language change and variation, and improving foreign and second language instruction.

Barlow (1996) notes, "the results of a corpus-based investigation can serve as a firm basis for both linguistic description and, on the applied side, as input for language learning" (p.32). Large general corpora have proven as an invaluable resource designing language teaching syllabi which extremely emphasise communicative competence (Hymes, 1992) and give prominence to those items likely to be encountered by the learners in real-life communicative situations.

Also, in the material development field, corpora have created a crucial impact on reference publishing and a new generation of dictionaries and grammar books have been produced. The Collins COBUILD English Course (CCEC), for example, was the first and probably most ground-breaking development in the context of corpora related English language teaching syllabi (Willis, 1989). The COBUILD dictionaries, grammars, usage guides, and concordance samplers based on authentic English compiled in accordance

with the needs of the language learners provide teachers and learners with more reliable information about the English language than traditional reference grammar books or previous generation non-corpus-based dictionaries (Goodale, 1995; Sinclair et al., 2001; Sinclair, 2004). Integrated corpus-derived findings on frequency distribution and register variation as well as containing genuine examples instead of invented examples based on intuitions are two major advantages of the COBUILD and other corpus-based reference works for learners.

According to Meyer (2002), corpora have become valuable resources for many disciplines of linguistics regarding creating dictionaries, understanding the process of language acquisition, studying language change and variation, and improving foreign and second language instruction. Meyer (2002) also explains as “for corpora to be most effective, they must be compiled in a manner that ensures that they will be most useful to those who are potential users of them” (p. 24).

2.3. Computer Learner Corpus

Learner corpus consists of a representation of learner language or “interlanguage” (Selinker, 1972). Interlanguage (IL) is a language system used by L2 learners and enables them to have approximations about L2 which indicate the learners’ current hypotheses about the target language. According to Selinker (1972), learners’ systematic and rule-governed use of language along with the errors in the process of language acquisition including transfer from the first language or overgeneralization of some L2 rules are presented through an interlanguage. From this point of view, learner language is not viewed as limited, but as a dynamic, non-linear, and evolving language system. As the linguists and researchers understand more about aspects of interlanguage, more informed L2 instruction may be provided to help learners produce native-like language.

The emergence of a new source of data for SLA research was brought to the life in the early 1990s: the computer learner corpus (Dagneaux et al., 1998). Computer learner corpus may be defined as a collection of machine-readable natural language data produced by L2 learners. Leech (1998) defines the learner corpus as an idea 'whose hour has come,' and "it is time that some balance was restored in the pursuit of SLA paradigms of research, with more attention being paid to the data that the language learners produce more or less naturalistically" (p.8). Computer learner corpus research differs from EA in some ways as "the computer, with its ability to store and process language, provides the means to

investigate learner language in a way unimaginable 20 years ago" (Leech, 1998, p. 8).

As explained in the introduction chapter (p. 1), learner corpora provide a new type of data which can alter thinking both in SLA (Second Language Acquisition) research, trying to realize the mechanisms of foreign/second language acquisition, and in FLT (Foreign Language Teaching) research, the aim of which is to improve the learning and teaching of foreign/second languages.

The corpus used in this study is the Cambridge Learner Corpus (CLC), the largest annotated test performance corpora which enables the investigation of the linguistic and rhetorical features of the learner performances in CEFR proficiency bands. The CLC is a 55,546,260-word corpus (16 million-words in 2003) of learner English collected by Cambridge University Press (CUP) and Cambridge English Language Assessment since 1993 (Nicholls, 2003). Essays produced by English language learners sitting examinations in English and written in response to examination prompts are included in CLC. The corpus comprises over 200,000 exam scripts from students coming from 148 different mother tongues and living in 217 different countries or territories. “The examination scripts have been transcribed retaining all errors, and a part of the corpus (a 25.5 million-word component currently, and a 6 million-word component in 2003) has been manually error-coded by two coders using an error annotation scheme containing 88 distinct error codes devised specifically for the CLC” (Nicholls, 2003, p. 575). The corpus and its error-annotated section have also been growing over time as an arising number of non-native speakers undertake language examinations every year. The error-coded component of the corpus currently contains 29,266,800 words.

Several key notions in the definitions of learner corpora could be worthy of further comment:

2.3.1. Authenticity

Sinclair (1996, p. 86) describes the default value for corpora as ‘authentic’: “All the material is gathered from the genuine communications of people going about their normal business” unlike data gathered “in experimental conditions or in artificial conditions of various kinds”. In the foreign/second language regard, purely experimental data coming from elicitation techniques does not qualify as learner corpus data. Nonetheless, in the case of learner language, the notion of authenticity is reasonably problematic. Even the most authentic data from non-native speakers is almost never as

authentic as native speaker data, especially in the case of EFL learners learning English in the classroom. For this very reason, learner corpora is considered as an important asset.

2.3.2. Textual Data

In order to arrive at valid conclusions, the language sample must consist of continuous stretches of discourse, not isolated sentences, or words in order to characterize as learner corpus data. Expressing 'corpora of errors' (James, 1998, p. 124) is therefore deceptive. The term "corpus" cannot be used to refer to a collection of erroneous sentences extracted from learner texts. Learner corpora are made up of continuous stretches of discourse containing both erroneous and correct use of the language at the same time. For this reason, utilizing an error tagged corpus like CLC yields relatively more reliable results to create the learner profile.

2.3.3. Explicit Design Criteria

As mentioned above, random collection of heterogeneous learner data cannot be characterized as a learner corpus. Learner corpora should be compiled according to strict design criteria, some of which are the same as for native corpora (Atkins & Clear, 1992), while others, relating to both the learner and the task, are specific to learner corpora. Figure 1 represents some of these CLC-specific criteria (Granger, 2002, p. 9).

Learner	*Task settings*
• Learning context • Mother tongue • Other foreign languages • Level of proficiency • [...]	• Time limit • Use of reference tools • Exam • Audience/interlocutor • [...]

Figure 2. CLC – specific design criteria

The usefulness of a learner corpus is related to the care maintained in controlling and encoding the variables. By this way only, we can carry out our analysis setting, the dependent and independent variables according to the purpose of our research.

2.3.4. ELT Purpose

A particular SLA or FLT purpose requires collecting a learner corpus. Testing or improving SLA theories such as confirming or disconfirming theories about L1 transfer, the order of acquisition, the production or contribution to FLT tools may be some of the purposes. In the current study, as we focus on prepositional verbs and related error typologies, the relevant samples are extracted in line with this purpose.

2.3.5. Standardization and Documentation

Various formats may produce a learner corpus. It can take the form of a raw corpus, i.e. a corpus of plain texts, or of an annotated corpus, i.e. a corpus enriched with linguistic or textual information, such as grammatical categories or syntactic structures. Standardized annotation should base on a software in order to ensure comparability of annotated learner corpora with native annotated corpora.

Thanks to the prominent language test performance corpora, such as the Cambridge Learner Corpus (CLC), containing the examples of texts in different score bands, data for investigation of the linguistic and rhetorical features of learner performances are provided (Granger, 2002).

Along with the small and sub-corpora corpus, a larger corpus was created with a view to being incorporated into the International Corpus of Learner English (ICLE) created by Sylviane Granger of the University of Louvain, Belgium. ICLE is a corpus of argumentative essays written by higher intermediate to advanced learners. Several academic teams' participation has created the corpus internationally as a teamwork. It contains over 2 million words of EFL writing from 16 different mother tongue backgrounds including Bulgarian, Chinese, Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana, Turkish (Granger et al., 2009).

The publication of the first version of the International Corpus of Learner English, ICLE (Granger et al., 2002) can be considered as the starting point in the development of large-scale learner corpora and a growing interest has increased in learner corpus research. Even though most of them are rather descriptive and/or pedagogical in nature, ICLE has been used in over 400 published L2 papers over the past few years (Granger, 2002). The first version of ICLE consists of 2.5 million words of argumentative essays written by university students with L2 English from several European countries, organized in

different sub-corpora divided according to L1: Spanish, Italian, French, Russian, etc. Such sub-corpora allow for a new type of analysis: Contrastive Interlanguage Analysis (CIA), i.e. the contrast of two (or more) interlanguage varieties, for example; L1 Spanish – L2 English vs. L1 Italian – L2 English. An expanded version of ICLE has been recently released (Granger et al. 2009) containing 3.7 million words from 16 mother-tongue backgrounds, including now Chinese, Japanese, Turkish and other L1s.

Two major categories comprise computer learner corpora: commercial CLCs, initiated by major publishing companies, and academic CLCs, compiled in educational settings. The Longman Learners' Corpus and the Cambridge Learner Corpus are two major commercial learner corpora both of which are quite large (10 million words for the Longman corpus and 16 million for the Cambridge corpus). The academic corpora, far more numerous, are extremely variable in size. The Hong Kong University of Science and Technology Learner Corpus contains 25 million words, while the Montclair Electronic Language Database only contains 100,000 words.

2.3.6. Studies on Computer Learner Corpus

Learner corpus studies have gained a great popularity for the last two decades. In this section, we will review some of the prominent studies in the field.

In the written English of native speakers and French learners' study of postnominal modifiers, Meunier (2000) clearly points to a lack of prepositional phrases and participial clauses in the learner writing and a significant overuse of relative clauses. In addition, there is also evidence from the Louvain error-tagged corpus of French learner writing that learners have persistent difficulty with relative pronoun selection.

Altenberg & Granger (2001), in their study, exphasize the importance of combining bilingual and learner corpus analysis techniques offering empirical evidence of L1 transfer in advanced L2 writing. In the study, the use of *make* in native English and advanced French and Swedish L2 writing, and the differences in the use of *make* in the native and learner groups have been compared. Comparisons have been investigated of original-version and translated written Swedish and English samples have been made using an aligned Swedish-English bilingual corpus with the hypothesis that overuse of causative *make* with adjective complements by Swedish L2 writers is due to L1 transfer.

Aijmer (2002) uses computer learner corpora to compare the range and frequency of some of the modal words in native and L2 writing of advanced level university students.

Although the primary focus of the study is Swedish L2 writers, she also represented comparisons from French and German L2 writers. Based on 52,000-word samples from each native and learner variety, the study takes into account not only modal auxiliaries but also a wider range of modal devices, such as adverbials and lexical verbs bearing modal meanings. The study reveals that modal auxiliaries by all the L2 writers are overused, a tendency which may be partly developmental, partly interlingual. Some evidences of 'register–interference', transferring patterns of use from spoken English into the written medium, were also revealed. In conclusion, Aijmer recommends providing students with a wider range of modal expressions including modal phrases, collocations and syntactically larger sentence patterns.

Housen (2002) carried out a corpus-based study into the acquisition of the basic morphological categories of the English verb system using annotated oral corpus data from learners grouped into four different levels of proficiency including native speaker baseline data. In the study, acquisition of these basic forms; stages of development in their acquisition; mapping these forms onto their appropriate temporal, aspectual and grammatical meanings, stages observed in the development of these form-meaning relations were elaborated. As a result, the study confirmed the general order of the formal morphological categories stated by previous studies, however, it revealed a significant variation at the level of individual learners fluctuating between overuse and underuse in building associations. In conclusion, Housen proposes studies based on large corpora of longitudinal data, to investigate the key issue of variation to be able to get a larger picture.

2.4. Computer-aided Error Analysis

Methodological approach that is specific to learner corpora established the foundations of studies of error-annotated learner corpora which is formed as computer-aided error analysis (Granger, 2002). In FLT field, a thoroughly error-analysed corpus can be an invaluable resource to inform and constitute as a pedagogical tool (Granger, 2009, p. 24). Cambridge Learner Corpus (CLC) is a clear example of the development of Cambridge University Press course and remedial teaching materials.

As we stated in Significance of Study Section (p. 8), the nature of the data and associated error classifications used by Traditional Error Analysis (TEA) have been criticized by many. The methodological limitations in this regard have been presented as follows (Dagneaux et al., 1998, p. 164):

- TEA is based on heterogeneous learner data;
- TEA categories are fuzzy;
- TEA cannot cater for phenomena such as avoidance;
- TEA is restricted to what the learner cannot do;
- TEA gives a static picture of L2 learning.

The first two limitations are methodological. Ellis (1994), referring to such data, highlights "the importance of collecting well-defined samples of learner language so that clear statements can be made regarding what kinds of errors the learners produce and under what conditions" and critisizes that "many EA studies have not paid enough attention to these factors, with the result that they are difficult to interpret and almost impossible to replicate" (p. 49). The problem is combined by the error categories suffering from a number of weaknesses such as being ill-defined, resting on hybrid criteria and involving a high degree of subjectivity. "Grammatical errors" or "lexical errors" terms, for example, are almost never defined making results difficult to interpret, as several error types (i.e. prepositional errors) fall somewhere in between and it is usually impossible to know in which of the two categories they have been counted. Also, the description and explanation levels of analysis are mostly mixed by the error typologies. Scholfield (1995) illustrates this with a typology made up of the following four categories: spelling errors, grammatical errors, vocabulary errors and L1 induced errors. As spelling, grammar and vocabulary errors may also be L1 induced, there is an overlap between the categories, which is "a sure sign of a faulty scale" (Scholfield, 1995, p. 190).

As for the other three limitations found in the scope of TEA, TEA's particular focus is on overt errors, in other words, neither non-errors such as instances of correct use, nor non-use or underuse of words and structures are not considered. Referring to this point, Harley (1980) clearly stated that:

> [It] is equally important to determine whether the learner's use of 'correct' forms approximates that of the native speaker. Does the learner's speech evidence the same contrasts between the observed unit and other units that are related in the target system? Are there some units that he uses less frequently than the native speaker, some that he does not use at all? (p. 4)

Furthermore, "EA has too often remained a static, product-oriented type of research, whereas L2 learning processes require a dynamic approach focusing on the actual course of the process" (van Els et al., 1984, p. 66).

These weaknessess of TEA have created a need for a new direction in EA studies. One possible direction, grounded in the fast growing field of computer learner corpus research, is computer-aided error analysis.

According to this, TEA's particular focus is on overt errors, in other words, both non-errors such as instances of correct use, and non-use or underuse of words and structures are not considered. Learners' error production could only be identified and analysed using TEA and ignores possible errors that learners do not make as they avoid difficult structures (Schachter, 1974). Granger (2002) adds that computer-aided approach to learner corpora is quite different from TEA studies as they involve a higher level of standardization and errors are demonstrated in the full context of the text, along with non-erroneous forms.

In his study, Can (2017) also commented that corpus-based computer-aided approach to error analysis with its pedagogical significances has been considered as a new direction introduced to guide Foreign Language (FL) teaching completing the missing parts stemming from the limitations of TEA. Computer-aided error analysis, taking its basis from learner corpora, according to Gilquin and Granger (2015, p. 427) "represents a major improvement over traditional EA, not least because instead of considering errors in isolation, it sees them as part of a system which includes both correct and incorrect uses" combining the fundamental principles of IL and TEA. About the facilitating role of corpus-driven error analysis in the replicability and interpretability of studies in the field, Ellis (1994) states that:

> The importance of collecting well-defined samples of learner language so that clear statements can be made regarding what kinds of errors the learners produce and under what conditions. […] Many EA studies have not paid enough attention to these factors, with the result that they are difficult to interpret and almost impossible to replicate (p. 49).

Clarifying Ellis's statement in this regard, Granger (2003) details the following two methods which may be used in computer-aided error analysis: the first one requires

selecting an error-prone linguistic item such as a word, phrase, word category or syntactic structure, and then, scanning the corpus to find examples of misuse of that structure using standardised text retrieval software tools. This method is quiet fast which is an advantage, however, the disadvantage is that the analyst must pre-empt the issue, limiting the searches to the items which the researcher considers to be problematic.

The latter method takes longer time however, the researcher may be led to discover problematic items of learners of which he was not aware of before. In this method, standardised system of error tags and tagging all the errors or particular categories, such as verb complementation or modals in a learner corpus are formulated.

The advantages of the computerised learner data for researchers could be presented as follow (Rayson, 2008):

- First, they are more manageable, and therefore easier to analyse.
- Secondly, they support the raw data with extra linguistic information, using either automatic or semi-automatic techniques.

Kochmar (2016) adds that the analysis of occurring learner errors based on the texts generated by foreign language learners is only possible thanks to the computer-driven error analysis techniques applied over learner corpora compiled systematically. Referring to the certain limitations of TEA, this was not possible before since TEA was performed on actual texts and uses a biased practice of "analysing out the errors and neglecting the careful description of the non-errors," only occurring errors are conceived; not the "uncommitted" ones (Hammarberg, 1999, p. 186). "The potentially problematic cases that learners might be unsure of even if they manage to not commit an error" (p. 187). That's why, only potential errors that are never realised due to avoidance strategies could not be considered by TEA approach.

This is very crucial for the pedagogical point of view because such cases that are missed through EA are of high value (Kochmar, 2016). The pivotal aspect of keeping these issues in mind when drawing conclusions about the typical errors of language learners, was also highlighted by Leacock et al. (2014). They described the most frequent errors in the writing of U.S. college students (all native speakers) comprising punctuation and sentence structure-related errors. However, only a minor portion of errors is represented by these errors in texts written by language learners. Leacock et al. (2014, p. 19) claimed that "the reason for that is not that non-native speakers master these aspects

of writing more successfully than native speakers, but rather that they avoid using the structures in which such errors can occur since they are unsure of the correct use in these cases."

Overall, there exists a great necessity of research into the field of learner corpora to be able to raise awareness concerning the linguistic patterns of language use within interlanguage.

The crucial issues to be investigated using the error analysis techniques have been listed by the researchers (James, 1998; Ellis, 1994; Ziahosseiny, 1999; Keshavarz, 2003; Darus, 2009; Kazemian & Hashemi, 2014; Can, 2017) are presented below;

- Language transfer
- Overgeneralization
- Simplification
- Underuse
- Fossilization
- Lack of the knowledge of the rules
- Interference

In the following sections, each of these will be dealt under separate headings exemplifying the related theories and studies conducted.

2.4.1. Language Tranfer

The main argument of Contrastive Analysis (CA) is that "language transfer" is the main cause of errors in second language learning. According to this, second language learning is similar to first language acquisition and is guided by universal and innate mechanisms. Language transfer, as Romaine (2003) suggests, is a source of linguistic variability in second language interlanguage. Kellerman (1979; 1983) propounded a re-definition of the notion of "transfer" claiming that the learner is an active administrative about what linguistic structures may be transferable to the second language. And thus, transfer can be seen as a cognitive process dependent on psycholinguistic rather than linguistic understanding of markedness on the basis of psychological and perceptual complexity, not structural complexity. This stems from Eckman's (1977) Markedness Differential Hypothesis (MDH) claiming that it should be possible to predict difficulties

L2 learners have depending on a contrastive analysis of two languages (L1 and L2) and the inclusion of the concepts of typological markedness and cross-linguistic influence. The MDH asserts that L1 structures that are different from L2 structures and typologically more marked will not be transferred while L1 structures that are different from L2 structures and typologically less marked are more likely to be transferred by the learner.

2.4.2. Overgeneralization

According to Richards et al. (1989), in both first and second-language learning, overgeneralization is a common process, in which a learner extends the use of a grammatical rule beyond its accepted uses. To exemplify, a child may come up with the word *ball* to be able to refer to all round objects around. Thus, the deviant structures produced by the learner on the basis of his limited knowledge of L2 and exposure to other structures of the target language are referred to overgeneralization errors.

At the cognitive level, the learning network improves performance with many learning trials, which results in a gradual developmental process. In this process, overgeneralization is specified by linguistic experience coupled with the similarity of the exemplar being learned by others already stored, its consistency and salience, as well as by frequency. How grammatical representations are acquired can be predicted by such single-route mechanisms (Richards et al., 1989).

2.4.3. Simplification

Perhaps the most important process (i.e. strategy) "employed by learners as compensation for partial or incomplete acquisition is simplification" (Winford, 2003, p. 217). Non-realisation (i.e. reduction, omission or deletion) of morphological markings and function words are mostly addressed by simplification, a strategy designed to ease the learner's perception and production. Simplification is a process where an L2 learner replaces a complex form of the target language with a simpler one according to Davydova (2011). To exemplify, a German speaker using the simple past tense in a context where Standard English requires the present perfect can be conceived of as simplification (Davydova, 2011, p. 11):

(1) Mesolectal German English (HCNVE: GE08)

(2)

I ***visited*** *French lots of times*.

'I have visited France many times.'

Moreover, in contexts where a native speaker of English is very likely to employ the *HAVE*-perfect, non-native speakers tend to use the structurally and semantically simpler present tense according to Davydova (2011). Nonetheless, simplification may be considered as a very general strategy, where learners use the semantically and morpho syntactically simpler, i.e. "least marked" (Housen 2002, p. 160) or more "natural" (Davydova et al. 2011) variant in their L2 production where the target language requires a semantical and a morpho-syntactically more complex form.

2.4.4. Underuse

In her well-rounded definition, Granger (1998), uses ‘underuse’ as a neutral term, which merely reflects the fact that a word/structure is less used, while ‘avoidance’ entails a conscious learner strategy. And thus, “avoidance” is a term referring to the cases in which L1 speakers may elude the use of some L2 structures when they are considered to be difficult (Alonso, 2002). According to Edmondson (1999), one of the key concepts in SLA studies is avoidance strategies used to refer to a process, whereby learners tend to avoid producing those forms in their target language that they perceive as difficult.

From a similar perspective, Macaro (2010) sees avoidance as a term mostly used in the context of learner strategies and particularly as a communication strategy in that it is a conscious mental act with the goal of using target language. Two basic types of avoidance strategies are adopted by learners: avoidance of topic (the lack of cultural knowledge required for the topic to be discussed adequately may cause topic avoidance. In this context, it results from a conceptual deficiency) and avoidance of formulations (the choice of certain words, phrases or language structures to express ideas).

Scovel (2000), touching upon an important point, argues that when a particular linguistic structure in the L2 is very different from that which a speaker has acquired in the L1, avoidance of certain structures is likely to occur. Accordingly, through avoidance of structures, very few mistakes are made and, for this reason, evaluating instances of avoidance as well as having an apprehension of the learner’s current interlanguage profile is quite difficult.

To sum up, the reason learners “underuse” some certain structures are due to the fact that they have limited access to the target language forms and are therefore not familiar with those linguistic categories. According to the study Davydova (2011) carried

out with learners having different mother tongue background, learners simply avoid or replace structures they have difficulty in with the simpler forms. Underuse is also the most probably reinforced by teaching since pedagogical materials rarely put emphasis on some unmarked multi-word units. It is therefore not surprising that some certain structures are underused by almost all learner populations from various L1 backgrounds.

2.4.5. Fossilization

Vygotsky (1978) explains “fossilization” in terms of cognitive processes that undergo prolonged development. Since ‘‘they have lost their original appearance, and their outer appearance tells us nothing about their internal nature’’ (p. 64), he acknowledges that these processes are difficult to investigate.

Selinker (1972) first introduced fossilization in the second language learner’s language as a construct in SLA. It has become widely accepted as a psychologically real phenomenon of theoretical and practical importance. Since 1972, it has undergone several revisions and is now accepted in general as referring to both “the stabilized deviant linguistic forms in the learner/user’s interlanguage and to the underlying cognitive mechanisms which have produced deviant stabilized forms” (Han, 2004, p. 20). From an initial perspective, incorrect linguistic forms that seem to have become permanent in the second or foreign language of the learner are referred by fossilization. For instance, /l/ and /r/ errors in the speech of Japanese learners of English despite a long time of training or word order errors in English speakers learning German. For Selinker (1992), fossilization could be defined as persistent non-target-like structures in the learner/user’s interlanguage. From the second perspective, it is also used to characterize and explain a foreign language learning process which has come to a halt and thus, it may be described as an inability of a learner to gain or produce a native-like ability in the target language. Similarly, Mitchell and Myles (2004) define fossilization as a phenomenon when a learner's second language system seems to 'freeze', or becomes stuck, at some more or less deviant stage.

2.4.6. Lack of the knowledge of the rules

According to Keshavarz (2012), errors stemming from the lack of the knowledge of the rules reflect the learner's competence at a particular stage of second language development as well as illustrate some of the general characteristics of language learning.

As a matter of fact, such errors are similar to errors produced by monolingual children in some ways and result from the learner's attempt to be able to create advanced structures in the target language in spite of his/her limited experience with it. The learner's ignorance of the restrictions and exceptions of target language rules are some of the causes of these errors. In other words, language learning process is halted and needs to be scaffolded by remedial teaching.

2.4.7. Interference

Last but not least, interference proposes that, as explained by Granger (2008), learners' mother tongue backgrounds may prevent them from acquiring certain grammatical patterns and lexical properties of the second language. For instance, errors stemming from L1 interference in students' writing performances are basically related with semantic transfer. In his study, Zhang (2013) studied his own learners that persistently keep developing hypotheses of semantic equivalents between English and Chinese lexical items, i.e. Chinese learners tend to make word translations in their writing, thus leading to transfer errors. For example, the erroneous collocations like "variety choices", "bad behaviour students", and "travelling phenomenon" are the results of word-for-word translations; and similarly, the semantic redundancy in the erroneous collocations such as "human population", and "a race competition" is due to the fact that there are frequent uses of semantic repetitions in Chinese (Zhang, 2013).

2.5. Prepositional Verbs

2.5.1. Definition of Prepositional Verbs

Prepositional verbs (PVs) are classified under the category of lexical items called 'multiword items' (MWIs), or 'multiword expressions' (MWEs) (Baldwin, 2005). These lexical items include idioms, phrasal verbs, prepositional verbs, stock phrases, prefabrications, and formulaic sequences. Baldwin (2005) claims that *passivization* is one of the characteristics of PVs. Some PVs cannot be passivized (fixed PVs) and others can be passivized (mobile PVs). Providing a passivization test would help distinguish PVs from other similar structures as in following examples of fixed PVs:

1. a. He *came across* a problem
   b. *A problem was *come across* by him

2. a. She walked down the path
   b. *The path was walked down

Baldwin (2005) states that the object of the PV must follow directly after the PV. These two examples cannot be passivized concluding that fixed PVs cannot be passivized. Mobile PVs can be passivized which could be frequently seen in academic prose:

3. a. She *referred to* the book in her paper
   b. The book was *referred to* in her paper
4. a. Lung cancer *associated with* smoking
   b. Smoking was *associated with* lung cancer

Claridge (2000) describes that multi-word verbs have not been investigated sufficiently so far; multi-word verbs have been excluded by the linguists dealing solely with lexical matters, for instance word-formation, MWIs were seen from another perspective as not constituting 'words'. However, Claridge (2000) claims that MWIs constitute lexemes and are formed on the basis of regular patterns; "and yet multi-word verbs are neither exotic nor at the fringes of the English language; rather, they are a part of the mainstream development" (p. 1).

Biber et al. (1999) provide a clear definition of prepositional verbs proposing, "the verb + preposition functions as a single semantic unit, with a meaning that cannot be derived completely from the individual meanings of the two parts" (p. 414). These two parts (words) unit may usually be replaced by a simple transitive verb with a similar meaning as follows:

looks like *(the barrel)* → ***resembles*** *the barrel*
thought about *(it)* → ***considered*** *it*
asked for *(permission)* → ***requested*** *permission*
deal with *(parking problem)* → ***handle*** *parking problem* (p. 414)

Additionally, many of the combinations used as PVs could function as phrasal verbs or free combinations according to Biber et al. (1999). To exemplify, *come from*, apart from its use as a PV, is also common as a free combination in conversation and fiction.

Referring to the Longman corpus, Biber et al. (1999) clearly exemplifies from various genres that when the noun phrase following prepositions *in, on* and *from* mainly identifies a person or a thing, these same multiword sequences can be considered PVs (with *wh*-questions formed with *who* or *what*):

*Susannah York and Anna Massey **appear in** the thriller The Man from the Pru. (NEWS)*

*They are, however, widely **used in** the preparation of special cakes. (ACAD)*

*The first goal **came from** Tim Cliss. (NEWS)* (p. 406)

Biber et al. (1999) also provides two structural patterns of PVs, both of which are transitive:

Pattern 1: **Noun phrase (NP) + verb + preposition + noun phrase (NP)**

*It just **looks like** [the barrel].*

*I've never even **thought about** it.*

*Britannia said he had **asked for** permission to see the flight deck.*

*The regulations also **apply to** new buildings.*

Pattern 2: **NP + verb + NP + preposition + NP**

*No, they like to **accuse** women **of** being mechanically inept.*

*He **said** farewell **to** us on this very spot.*

*But McGaughey **bases** his prediction **on** first-hand experience.*

*They were cosmologists wrestling to **apply** quantum mechanics **to** Einstein's general theory of relativity* (p. 415).

PVs have two competing structural analyses. On the one hand, they can be treated as simple lexical verbs followed by a prepositional phrase functioning as an adverbial. According to this analysis, inserting another adverbial between the verb and the prepositional phrase is possible, such as:

*She **looked** exactly **like** Kathleen Cleaver.*

*I never **thought** much **about** it.*

PVs are relatively common in all four registers, occurring almost 5,000 times per million words being particularly common in fiction (Biber et al., 1999) corpus findings. PVs have higher frequency than phrasal verbs which have a limited set of adverbial particles usable for their formation, all referring location or direction. Nonetheless, PVs attract the full set of prepositions, including important forms referring "non-spatial

relations, such as '*as, with, for*, and *of*' (p. 415).

On the other hand, De Haan (1988) identifies PVs by determining whether the object of the preposition answers the questions *who* or *what* as may be seen in the examples:

5. a. He looked (at the dog)
   b. What did he look at?
   c. *Where did he look?
6. a. She was walking on the sidewalk
   b. Where was she walking?
   c. *What was she walking on?

The object of the preposition answers the question *what* in the first example assigning "look at" a prepositional verb function. However, in the second example, the object of the preposition answers the question *where,* which does not meet the criteria for a prepositional verb and the answer to the question "at the dog" is actually a free combination of prepositional phrases.

From cognitive perspective, PVs are built with two or more words in the mental lexicon enabling learners to sound more native-like as claimed by Gardner & Davies (2007); they are also difficult to acquire given the challenges posed by prepositional usage for most EFL learners. Tetreault & Chodorow (2008) claim that various linguistic functions served by prepositions cause the difficulty in prepositional usage and make it challenging for the learners. Once the argument of a predicate is marked by a preposition, such as an adjective, a noun, or a verb, preposition selection is restrained by the argument role that it marks, the noun which fills that role, and the predicate. In addition, alternations are indicated by many English verbs (Levin, 1993) in which an argument is sometimes marked by a preposition and sometimes not (e.g., "They loaded the wagon with hay" / "They loaded hay on the wagon"). Tetreault & Chodorow (2008) clarify this alternation as "when prepositions introduce adjuncts, such as those of time or manner, selection is constrained by the object of the preposition ("at length", "in time", "with haste"). The selection of a preposition for a given context also depends upon the intended meaning of the writer ("we sat at the beach", "on the beach", "near the beach", "by the beach")" (p. 865).

Following Quirk et al.'s definition (1985), Nesselhauf (2005) also states that prepositional verbs are formed when verbs govern a preposition either directly after the

verb (as in *wait for*) or after an object (as in *explain sth. to sb.* and also in more idiomatic cases such as *take care of*).

This particular study focuses on one specific type of MWIs, prepositional verbs. One of the important aspects needed to be discussed regarding to PVs is how they are similar and different from other MWIs. According to Biber et al. (1999), two main lexical verbs containing multiple words exist: phrasal verbs (verb + adverbial particle) and prepositional verbs (verb + preposition). PVs employ the full set of prepositions such as "be used in", "be based on", "look for", "be associated with". Additionally, the two parts consist of a "PP (prepositional phrase) argument is made up of a specified preposition head and NP (noun phrase) argument" (Baldwin, 2005, p. 116). According to this definition, PVs are transitive, and they take an object.

"Since none of the criteria for prepositional or phrasal-prepositional verbs are compelling, it is best to think of the boundary of these categories as a scale" (p. 1165) is how Quirk et al. (1985) elaborate on prepositional verbs.

A classic classification given by (Mitchell, 1958, p. 31) is as follows which has the advantage of great clarity and simplicity despite including only a small part of the possible combinations:

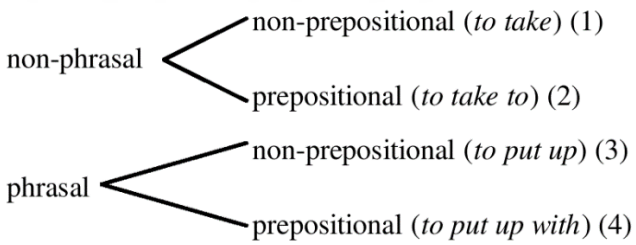


Figure 3. Classification of prepositional and non-prepositional verbs

His whole system is based on binary contrasts, as is inferred above. Four types of multi-word verbs are classified with (1) representing the simplex verb, namely prepositional verbs (2), phrasal verbs (3) and phrasal-prepositional verbs (4). Mitchell's scheme contributed a lot to the establishment of common terms for the three major and most common types of multi-word verbs. Claridge (2000) defines prepositional verbs as "verbs followed by a preposition in its clear prepositional use" (p. 39) adding that only one type, a transitive one, exists because if one construes verb and preposition as a unit, then the following noun (phrase) functions as its direct object.

Another reliable test to be able to differentiate PVs from phrasal verbs (Quirk et

al., 1972) is that the flexibility to insert the object of the preposition between the verb and preposition. Examples of phrasal verbs allowing the grammatical object to be inserted between the verb and the preposition:

7. a. John *called up* the man
   b. John called the man up
8. a. She *left out* that part in her paper
   b. She left that part in her paper out

In contrast to examples (7) and (8), in the following PVs, the object cannot be inserted between the verb and the preposition, as it is part of the prepositional phrase (PP).

9. a. John *called on* the man
   b. *John called the man on
10. a. She *put up* with the heat
    b. *She put with the heat up

On the other hand, Teschner and Evans (2000) differentiate PVs from phrasal verbs by adverb intrusion:

11. a. They called *frequently* on their teacher.
    b. They called on their teacher *frequently.*
12. a. They *frequently* called up their teacher.
    b. *They called *frequently* up their teacher.

In (11), *called on* is a PV and the adverb between its constituents is accepted. Nevertheless, in (12), the adverb cannot be inserted between *called up* because it is a phrasal verb, not a PV.

In this context, (Goyvaerts, 1968) identification of several subgroups of prepositional verbs falling into two broad categories, which are (i) purely prepositional verbs (e.g. look after, believe in) and (ii) verbs which need a preposition to introduce their object (e.g. consist of, rely on). Claridge (2000) takes both kinds as prepositional verbs despite some of their differing characteristics. With prepositional verbs, the preposition cannot be changed after the following noun phrase, while this is possible with phrasal verbs, and even obligatory in the case of a pronominal object (with certain idiosyncratic idiomatic exceptions). This may be regarded (Brown et al., 1976) as the main distinguishing criterion. In addition, Claridge (2000) elaborates that “in contrast to prepositional verbs, phrasal verbs do not quite as freely allow their sequence to be interrupted by manner adverbs (with certain monosyllabic exceptions and those only with

literal phrasal verbs)" (p. 64).

Claridge (2000) claims prepositional verbs as the most difficult to define, adding that they are also "the nearest the borderline of all possible multi-word verbs which stems from the fact that (most, if not all) sentences containing a verb-preposition sequence allow two different syntactic analyses, depending on the various possibilities of bracketing of constituents" (p. 57). 'SVA' which involves a simplex verb followed by an adverbial phrase is the first and more straightforward solution, i.e. in this case there is no prepositional verb according to Claridge (2000). Nevertheless, certain peculiarities of English point towards the likelihood of an SVO analysis as in (18b):

(18a) [All four books] [speak] [of a "someone other".]
(18b) [All four books] [speak of] [a "someone other".] (BNC A05 493) (p. 56)

The facts privileging (18b) are both structural/syntactic and semantic. The curious prepositional passive is involved in the first one rather than the others (cf. (19)) and the various preposition stranding possibilities (cf. (20)), e.g. in relative clauses and topicalized sentences, while the latter is represented by clearly idiomatic (mostly semantically opaque) verb-preposition combinations (cf. (21) vs. literal (22)) as explained by Claridge (2000, p. 56) from examples extracted from BNC:

(19) Her Birmingham background is <u>hinted at</u> ... (BNC A06 913)
(20) ... and never try to lift more than you can easily <u>cope with</u>. (BNC A0G 2204)
(21) "... and thus we <u>come by</u> those ideas we have of yellow, white, [etc.]" (BNC ABM 816)
(22) Soldiers from Stirling would have <u>come by</u> Crieff and Amulree, ... (BNC A0N2171)

All these point to a greater cohesion between verb and preposition than between preposition and following noun phrase, however, they do not completely avoid the SVA analysis except for the idiomatic cases.

2.5.2. Turkish Prepositional Verbs

Having an agglutinative morphosyntactic nature, prepositions exist as

“postpositions” in Turkish language, owning up to what prepositional verbs interpret in English language. According to Kornfilt (1997), postpositions can be grouped into two categories:

1. postpositions that do not bear agreement morphology with their objects;

2. postpositions that do exhibit (possessive) agreement morphology with their objects and can thus be analysed as nouns rather than genuine postpositions (p. 423).

Kornfilt (1997) describes that assigning a variety of cases to their objects or co-occurring with objects that are not overtly marked for case are subsisted in the first group consisting of postpositions as follows:

A. Postpositions that assign no overt case:

Üzere, üzre 'on; according to; for the purpose of'
mostly takes as object an infinitival clause,
but can also take a noun phrase as an object

(1470) Hasan [[Ankara -ya git -mek] **üzere**]
Hasan Ankara -Dat. go -Inf. **for the purpose of**
biz -den ayrıl -dı
we -Abl. part -Past
"Hasan left us so as to go to Ankara" (p. 423).

B. Postpositions that assign genitive case to all personal pronouns, to singular demonstrative pronouns and to the singular interrogative pronoun **kim** 'who', but which assign no overt case to all other nominal elements (including pronouns pluralized by -**1Ar**):

gibi **'like';**
ile **'with';**
kadar **'as much as';**
için **'for'**

(1472) Hasan bu sonat -**I** [[ben -**im**] **gibi**] çal -dı
Hasan this sonata -**Acc**. I -**Gen** **like** play -Past
"Hasan played this sonata like me" (p. 423).

C. Postpositions that assign dative case:

göre, **nazaran**	‘according to; suitable for’
doğru	‘towards’
karşı	‘against’
dair	‘concerning’
kadar, **-dek**, **değin**	‘as far as’
rağmen	‘in spite of’
nispeten	‘in comparison to; comparatively’

(1476) Hasan [köy -**e** **doğru**] yürü -dü
Hasan village -**Dat**. **towards** walk -Past
"Hasan walked towards the village" (p. 424).

(1477) Hasan tam [[ban -**a** **göre**] bir piyano]
Hasan exactly I -**Dat. suitable for** a piano
al -dı
buy -Past
"Hasan bought a piano precisely suitable for me" (p. 423).

D. Postpositions that assign ablative case:

evvel, önce	‘before’
sonra	‘after’
beri	‘since’
bu yana	‘since’
yana	‘as regards…’
dolayı, ötürü	‘because of’
başka	‘besides, apart from’
itibaren	‘starting from, with effect from’

(1478) Hasan sinema -ya [ben -den **önce**] git -ti
Hasan cinema -Dat. I -Abl. **Before** go -Past
"Hasan went to the movies before me" (p. 425).

The second group as Kornfilt (1997) describes, i.e. the "fake" postpositions that actually comprises of nouns with possessive suffixes, which are two subsets:

1. those that can be used with any possessive suffix and any case (the latter assigned to it by the verb),

2. those that can be used only with certain (frozen) cases, although the possessive suffixes can differ (p. 425).

Group 1:

alt ‘underside’
ara ‘interval, space’
arka, art ‘back’
baş ‘immediate vicinity’
dış, hariç ‘exterior’
etraf, çevre ‘surroundings’
iç, dahil ‘interior’
karşı **‘**opposite side’
orta ‘middle’
ön ‘front’
peş ‘space behind’
üst, üzer- ‘top’
yan **‘**side’

(1479) Hasan ben -im **arka -m -da** dur -uyor
Hasan I -Gen. **back -1.sg. -loc.** stand -Pr .Prog.
"Hasan is standing behind me" (p. 425)

Group 2: The examples are listed in their forms for third person singular:

hakk-ın-da 'concerning, about', in the locative
taraf-ın-dan 'by; through the agency of', in the ablative
yüz-ün-den 'because of', in the ablative
bakım-ın-dan 'from the point of view of', in the ablative
nam-ın-a 'in the name of; by way of', in the dative

(1481) Hasan [dilbilim **hakkında**] bir bildiri ver -di
Hasan linguistics **about** a paper give -Past
"Hasan gave a paper about linguistics" (p. 426).

In this context, the postpositions in the second group are like those exemplified under B, nevertheless, with the difference that these postpositions in the second group bear possessive suffixes and (frozen) case suffixes. These two cannot be seen on group B

postpositions according to Kornfilt (1997).

Group 3: This group is a version of group 2, in that the genitive case may mark the lexical objects of these inflected postpositions (even though they don't have to).

esna-sın-da 'in the course of', in the locative

zarf-ın-da 'during', in the locative

saye-sin-de 'thanks to', in the locative

uğr-un-a 'for the sake of', in the dative

yer-in-e 'instead of', in the dative

(1484) [Hasan -ın **sayesinde**] bugün iş -ten erken
Hasan -Gen. **thanks to** today work -Abl. early
çık -abil -di -m
leave -abil. -Past -l.sg.
"Thanks to Hasan, I was able to leave work early today" (p. 427).

Additionally, Kornfilt (1997) gives some examples of place expressions which are traditionally analysed as adverbs (rather than as nouns) such as:

içeri 'inside' dışarı 'outside'
yukarı 'up' aşağı 'down'
ileri 'forward' geri 'backward'
karşı 'opposite'

These expressions may be used as nouns demonstrated by attaching suffixes to them which are used as nominal inflectional markers, and by directing them into nominal phrases such as:

(1576) ev -in içeri -si
house -Gen. inside -3.sg.
"the inside of the house" (p. 452).

These adverbs can also be used in their bare form as well as in the locational cases, i.e. with the suffixes for dative, locative and ablative:

(1577) içeri gir -di dışarı çık -tı
Inside enter -past outside exit -past
"She went in" "He went out"

(1578)	içeri -**ye** gir -di	dışarı -**ya** çık -tı
	Inside -**Dat**. enter -Past	outside -**Dat**. exit -Past
	"She went in"	"He went out"
(1579)	içeri -**de** kal -dı	dışarı -**dan** gel -di
	Inside -**loc**. stay -past	outside -**Abl**. come -Past
	“He stayed inside”	“She came from outside” (p. 452).

Having a mother tongue with a heavily agglutinative structure, Turkish learners of English are assumed to have difficulty in learning the related functional category in the target language. This assumption has also been propounded by Hong et al. (2011) stating that many prepositional verb errors are interlingual. Supporting the view, Nesselhauf (2005) claimed that almost 50% of errors pertaining to the related error typology is caused by L1 influence. The following section aims at reviewing some of the studies carried out related with the prepositional verbs in English as an interlanguage.

2.5.3. Studies on Prepositional Verbs and Learner English

According to Römer (2008), some influence on the design of reference works and teaching materials that has been exerted from another branch of general corpora research is the area of phraseology and collocation studies. The importance of recurring word combinations and preassembled strings in a pedagogical context due to their great potential in nurturing fluency, idiomaticity and accuracy have been emphasized by many scholars like Biber et al. (1999), Hunston/Francis (2000), Kjellmer (1984), Lewis (2000), Meunier/Gouverneur (2007), Nattinger (1980). In this context, Hunston and Francis (2000) emphasizes as follow:

> Although corpus-based collocation dictionaries (Hill/Lewis 1997) are available, and although information on phraseology (i.e. about the combinations that individual words favour) is implicitly included in learners dictionaries in the word definitions and the selected corpus examples and sometimes even explicitly described, e. g. in the grammar column in the COBUILD dictionaries and in the COBUILD Grammar Patterns reference books (Francis/Hunston/Manning 1996); such information and exercises on typical collocations are as yet largely missing from LT course books (or they are inadequate). I see a necessity in and look

> forward to [more] information about patterns being incorporated in language teaching materials (p. 272).

Speaking, reading, and writing more fluently and native-like may be enabled by the acquisition of MWIs (Hong et al., 2011); however, lacking these items in a learner's interlanguage may be resulted as "foreign sounding" speech or writing (Cobb, 2003). Gardner & Davies (2007) point out that it is quite difficult how and what to include in second language curricula due to the idiosyncratic nature of these items. Laufer and Waldman (2011) describe the difficulty with collocations and learning MWIs could not be related to learners' particular mother tongues or years of L2 instruction variables.

As for this corpus-based study's specific topic, which is the use of PVs, Biber et al. (1999) claims that, for one reason it is crucial to examine PVs is that because they are commonly found across registers and genres (unlike phrasal verbs, which are the most specific to conversational discourse). Moreover, due to their complex behaviour, PVs display a challenge for EFL learners. Tetreault & Chodorow (2008) claim that one of the causes of such difficulty stems from the use of prepositions. Also, it is a set of forms that not most L2 learners find difficult according to Hong et al. (2011). Because of their patterning challenge, even the native speakers of English have difficulty in choosing the correct preposition in a cloze-test task, yielding a 75% success rate.

Moreover, according to Hong et al. (2011), one of the other reasons why EFL learners have difficulty in preposition-related collocations is that, despite the fact that L2 students are taught and encouraged to look up unknown words in a dictionary, they are not usually taught and encouraged to study the related lexical items in the context of their collocates. This causes unknown verbs being used with the inappropriate preposition combinations, which could not inevitably affect meaning. However, it certainly affects the EFL learners' proximity to the native-like quality in their interlanguages.

In order to gain an apprehension of EFL learners' knowledge and understanding of PVs, it is necessary to look into how these patterns are displayed in naturally-occurring language, for instance, using corpus data. PVs may be found in both spoken and written genres. Moreover, Biber et al. (1999), report that PVs are three to four times more common than phrasal verbs in their *Longman Grammar of Spoken and Written English* and also PVs are more common in academic prose. Causative (such as *call for*) and existence (such as *refer to, live with, rely on, stand for*) domains are two most commonly found domains in academic prose. Activity PVs, such as *deal with*, *go through*, and *engage in* and mental PVs such as *apply for*, *look after*, and *point to* are the other two

common domains (p. 414-418).

Because of the fact that accepted combinations of verbs with their prepositional collocates are not produced by the learners, their interlanguage turns out to be less "native" when these verbs do not contain all their necessary constituents, whether substituted or omitted, according to Wilcoxon (2014). The literature in this area highlights the importance of prepositional verbs in L2 learning (Hong et al., 2011); however, there is a lack of empirical studies of these verbs in ESL and EAP writing.

On the other hand, some other approaches deny the existence of prepositional verbs or at least conceive them superfluous. Götz & Herbst (1989), in their review of Quirk et al. (1985), criticize their view because of the fact that prepositional verbs lead to an increase in both the syntactical rules and the lexicon. According to Huddleston (1984), verb-preposition sequences cannot be seen as a single syntactic element, based on two things especially: "(i) the position of an adverbial, i.e. Ed relied steadfastly on the minister vs. *Ed relied on steadfastly the minister, and (ii) the possibility of coordination, or rather repetition of the preposition, i.e. Ed relied on the minister and on his solicitor" (p. 202).

According to some of the common verb-preposition sequences such as depend on, look after, wait for, or belong to, certain collocational restrictions (Palmer, 1975, p. 215) or even a collocational fixity appear in this situation. Chosen preposition should not be freely substitutable by another preposition, (Davison, 1980), stating the same condition for the verb. The importance of this condition is for the preposition, as it may be assumed that it is the verb that selects the accompanying preposition, and not vice versa. Additionally, the preposition can be substitutable in some cases e.g. besides *look after*; there also exists *look for, look into*, but the exchange in these cases is not 'free', instead, it is governed by different meanings that also affect '*look'* itself. Furthermore, this should not mean that the verb should never occur without the preposition, as, e.g., rely on, if it is to be a 'real' prepositional verb. '*Depend, look, wait, belong'* can all stand on their own, however, each involves different (shades of) meaning(s) or different syntactic environments, e.g. That depends / That depends what you mean. In a certain usage, the preposition is a fixed member of the sequence. If collocational fixity is taken one step further, it is the stage of lexicalization (Bolinger, 1977 p. 58) thinks it possible that "the great majority of prepositional verbs are lexicalized" (p. 59). His specific definition of lexicalization is not given, however, if Bauer's (1983) terminology is taken into account, for example, many prepositional verbs can certainly be called "institutionalized" and "a certain amount even semantically lexicalized" (p. 48). The latter contributes to the

criterion of meaning. That verb-preposition sequence is more likely to be a prepositional verb if it has a "unitary (i.e. usually idiomatic) meaning" (Palmer, 1975; Bolinger, 1977).

The studies conducted on L2 acquisition of prepositional verbs, and specifically on error analysis, is scarce. Among few, Hong et al. (2011) report and discuss common PV errors in their study about collocations in Malaysian English learners' writing. In the research, the authors used the Oxford Collocations Dictionary and the BNC to determine acceptable collocations with prepositions. Analysing the corpus from their Malaysian EFL learners, (written and spoken data of 35,931 words), they found the most general and common errors of verb-noun collocations and categorized them according to their frequency. They found out that these errors stem from overuse of the same prepositions. Some common reasons in the use of prepositions are provided by Hong et al. (2011), one of which is interlingual transfer in collocations which occurs between the learners' L1 and L2. An example of intralingual transfer errors includes **dropped into the river* instead of *fell into the river.*

As for the errors in this regard, Tetreault and Chodorow (2008) also claim that learners overused the same prepositions adding that there are two common types of collocation errors regarding PVs. They either select an incorrect preposition or simply omit it. They state that the most common incorrectly used prepositions are *in, to,* and *of,* which take place in the top ten list of most frequent prepositions.

In their study, Morris and Cobb (2004) found out that L2 learners commonly apply avoidance strategies when it comes to using MWIs. Laufer and Waldman (2011) share the same view, proposing that "EFL learners have a tendency to overuse a limited number of learned verb+preposition collocations formed from common verbs such as 'be, have and make'" (p. 652). Laufer and Waldman (2011) also agree with the influence of L1 on L2.

2.6. Data Driven Learning

That DDL is able to empower learners to learn by themselves, and that corpora have a great pedagogic potential have strenghtened the TaLC (Teaching and Language Corpora) approach according to Römer (2008). Studies on the teaching and learning of vocabulary by Cobb (1997), Cresswell (2007) and Stevens (1991) have proved the effectiveness and potential of DDL. Learner-centred activities with the teacher as a facilitator in the DDL method have not only been discussed with reference to English

language corpora and English language teaching but have also been applied in teaching other languages. The awareness-raising potential of DDL activities with Language for Specific Purposes (LSP) corpora has also been proven to be useful for DDL with learner corpora according to Ludeling and Kytö (2009). Meunier (2002) provides many examples and describes the advantages of DDL exercises with native and learner concordances adopting Granger's (1998) suggestion to combine data from native and learner corpora in the LT classroom. Nevertheless, according to Meunier (2002, p. 134) these kinds of exercises ought to "only be used, here and there, to complement native data and to illustrate […] universally problematic areas" (e. g. verb or noun complementation).

In addition, although there exists a shift in focus from learner errors to dealing with questions learners have about what they and their classmates have written, Seidlhofer (2002) also elaborates on "using learner corpora for learning" (p. 217). Motivation can be assured through the focus on these kinds of familiar texts (i.e. on texts the learners themselves have produced) and "the consideration of two equally crucial points of reference for learners: where they are, i.e. situated in their L2 learning contexts, and where they eventually (may) want to get to, i.e. close to the native-speaker language using capacity captured by L1 corpora" (p. 222) as put by Seidlhofer (2000). Mukherjee and Rohrbach (2006) also provided positive examples of pedagogical applications of such local learner corpora including their learners' not profiting only from the correction of their own mistakes but also their fellow-students' errors and corrections in their paper on error analysis in a local learner corpus. Seidlhofer's and Mukherjee and Rohrbach's local learner corpora may easily be compiled by a larger number of teachers and lecturers by simply collecting their students' writings in electronic format, which afterwards serve as an appealing source of data to inspire the creation of DDL materials.

DDL can be either teacher-directed or learner-led (i.e. discovery learning) depending on the student's levels, however, it is generally learner-centred which "gives the student the realistic expectation of breaking new ground as a 'researcher', doing something which is a unique and individual contribution" (Leech, 1997, p. 10). In this context, language teachers should be more equipped with the required skills in corpus analysis.

Seidlhofer (2002) suggests using learner data not in the context of data-driven learning but rather of learning-driven data. In her study, she collected the writings of a summary of a text and a short personal reaction to it and used these texts as the primary objects of analysis, to get the learners to work with and on their own output. Seidlhofer

stated the success of her study since learners played much more active and responsible roles in their learning. It seems possible that the approach may be more successful with advanced learners even though this remains to be tested.

Flowerdew (2009) carries out a study on signalling nouns which have functions as attitude, assistance, difficulty, endurance, process, reason, result etc. are thought to be problematic for learners. He collected a corpus of argumentative essays written by Cantonese L1 learners of English, presenting a taxonomy of error types and frequency data of the different error types in the use of signalling nouns. The study also compared the average number of signalling nouns used per essay with grades and the numbers of signalling noun errors according to grades. The findings confirmed the intuitive idea that the use of signalling nouns adds to the overall coherence of a text.

Furthermore, Chambers (2005) has helped in the creation of a one-million-word corpus of journalistic French (Chambers-Rostand Corpus of Journalistic French) and provided a number of illustrations of how DDL can be used in the context of teaching French and how it can facilitate the development of learner autonomy.

Data driven learning and teaching, although it still needs some improvements in tools and related teacher training curriculum, seems to be very promising in terms of learner authonomy, awareness raising and learning through cognitive associations.

CHAPTER III

METHODOLOGY

3.1. Introduction

Majority of the studies conducted on EFL grammatical development could be categorised into two as stated by Bardovi-Harlig and Bofman (1989):

(a) studies that examine formal features and
(b) studies that seek to gauge overall progress by a developmental index" (p. 18).

The present study could be inserted into the second category. It aims at detecting and analysing the interlanguage characteristics of developmental PV errors of Turkish EFL learners using a substantial written data available in CLC. As Biber, Conrad, and Reppen (1998) stress research studies focusing on large data sets may lead to significant improvements in the understanding of language acquisition. For this reason, by analysing Cambridge Learner Corpus (CLC), probably the largest annotated learner corpus (29,266,800 words), this study aims to contribute to the emerging body of research in learner-corpus-based error analysis and foreign language learning with its probe of the PV errors of Turkish EFL learners.

The CLC (Cambridge Learner Corpus) is a 52.5 million-word corpus (16 million-words in 2003) of learner English collected by Cambridge University Press (CUP) and Cambridge English Language Assessment since 1993 (Nicholls, 2003). Essays produced by English language learners sitting examinations in English and written in response to examination prompts are included in CLC. Kochmar (2016) claims that the CLC comprises over 200,000 exam scripts from students speaking 148 different mother tongues, living in 217 different countries or territories. As a comparison, in 2003, texts written by speakers of 86 different mother tongues were included in the corpus and more than 15 mother tongues were represented with more than 350, 000 words (Nicholls, 2003). "The examination scripts have been transcribed retaining all errors, and a part of the corpus (a 25.5 million-word component currently, and a 6 million-word component in 2003) has been manually error-coded by two coders using an error annotation scheme containing 88 distinct error codes devised specifically for the CLC" (Nicholls, 2003, p. 575). The corpus and its error-annotated section have also been growing over time as an

arising number of non-native speakers undertake language examinations every year. Meta-information about the learner such as; learner's L1, age, sex, education history and number of years the learner has studied English for are accommodated in each examination script.

3.2. Material and Error Tagging

The Turkish EFL learner subcorpus extracted from CLC consists of 275.969 words. For the purpose of the study, spoken part of the corpus data has been excluded from the subcorpus. The breakdown of the corpus is as follows.

Table 1

The Breakdown of Turkish EFL Learner Subcorpus in CLC

Corpus Size	**Exam**	**Years**	**Exam Format**	**Pass/Fail**	**Age**	**Register**
A1-A2: 31.277	BEC1	1993-2012	Article	Above level	10-50+	Formal
	BEC2		Composition/essay	At level		Informal
B1-B2: 143.191	BEC3		Informative/instructional text			Mixed
	BECH		Letter/reference			Neutral/unmarked
	BECP		Note/email/memo			
C1-C2: 101.501	BECV		Proposal			
	BULATS		Report			
Total: 275.969	CAE		Review			
	CPE					
	FCE		Story			
	ICFE		Survey/questionnaire/form			
	IELTS ac					
	IELTS gt					
	ILEC					
	KET					
	KETfS					
	PET					
	PETfS					
	SfLE					

Granger (2003) describes ideal error taxonomy to be used in error analysis as consistent, informative, flexible and reusable. As stated by Heift & Schulze (2007), most error taxonomies have the following problems:

• Coverage: some errors may be left unclassified when the taxonomy is applied to more texts or future analysis, due to the impossibility of error anticipation;

• Homogeneity: it is difficult to achieve a comprehensive and meaningful description of errors; and,

• Classification characteristics: many categories are language dependent or cannot be easily distinguished

In this study, the PV typology used is based on the one developed by CLC. Learner errors in CLC are tagged using the following convention (CUP, 2014, p. 14-15):

<#CODE>wrong word|corrected word</#CODE>

According to this tagging convention, when lexical/functional items or punctuation marks are missing, the error coding system still includes the bar (|), but with nothing before it. For example:

"*he explained me" would be coded as:
he explained <#MT>|to</#MT> me.

Almost all of the error codes are based on a two-letter coding system. In this system, the first letter represents the general error typology (e.g. wrong form, omission), while the second letter identifies the word class of the required word. According to this, general error typology, namely first letters of the tags, would be:

F wrong **Form** used
M something **Missing**
R word or phrase needs **Replacing**
U word or phrase is **Unnecessary** (i.e. redundant)
I word is wrongly **Inflected**
D word is wrongly **Derived**

M, R and U codes can be assigned to errors alone. In these cases, it means that no

more specific information can be given. Second part of the tagging includes the word classes:

A Pronoun (Anaphoric)
C Conjunction (linking word)
D Determiner
J AdJective
N Noun
Q Quantifier
T Preposition
V Verb (includes modals)
Y Adverb (...LY)

Following this convention, preposition errors were tagged, and the examples of the related errors made by Turkish learners are as follow:

- **DT derivation of preposition:** The preposition is not valid. It resembles the correct form but it has been incorrectly derived. In our study, derivational PVs have not been found across proficiency levels except for one:
  1. bought a bicycle *<#MP>* | , *</#MP>* I would n't ride it **during** the winters *<#DT>* *in case* | *because of* *</#DT>*<#UP> , | </#UP> the

- **MT missing preposition:** A preposition needs to be added to complete the phrase or sentence.

Examples:

1. And the restaurant did not have any special menu which would provide us <#MT> | with </#MT> a different taste

2. that <#TV> it 's <#S> definetly | definitely </#S>| it 's definitely been </#TV> worth waiting <#MT> | for </#MT> as I have some great and <#S> unbeliavable | unbelievable </#S>

3. I think it 's unfair to interfere <#MT> | with </#MT><#UD> the | </#UD>

nature

- **RT replace preposition:** A valid preposition has been used but it is not correct in the context.

Examples:

1. If you are interested <#RT> on | in </#RT> it , you can call me <#MT> | on </#MT> this number 0 786 5544 . Dear ... , I started my new

2. it 's very pretty and I can use it . I can look <#RT> at | in </#RT> the mirror and when I look <#RT> at | in </#RT> the mirror I 'll

3. I can look <#RT> at | in </#RT> the mirror and when I look <#RT> at | in </#RT> the mirror I 'll remember Jenny . <#L>

- **UT unnecessary preposition:** A valid preposition has been used but its inclusion is incorrect.

Examples:

1. write <#MT> | to </#MT> me soon . Dear Bart , I went <#UT> for | </#UT><#S> shoping | shopping </#S> yesterday and I bought <#RP> alot | a

2. tell you <#UC> that | </#UC> what happened . We went <#UT> to | </#UT> there <#RT> by | in </#RT> Sarah 's car . We stayed in a tent . It

3. <#AGN> part | parts </#AGN> of a school , because I believe <#UT> in | </#UT> the first impression <#M> | you make </#M> on

Using the error tags, we extracted all preposition errors and created a subcorpus. In the subcorpus compiled including all preposition errors, we needed to find all prepositions occurring in verb bundles for the purpose of the study. For this reason, we wrote a special syntax in Corpus Query Language (CQL) to extract the related structures:

[tag="IN"] within [tag="V.*"] [tag!="V.+"] [tag="IN"]

In the resulting extraction, we conducted our detailed data analysis focusing on

each preposition and the related error typologies excluding phrasal verb category manually.

3.3. Data Analysis Instruments

3.3.1. The Sketch Engine

Sketch Engine is a corpus manager and text analysis software available since 2003. Its purpose is to examine language behaviour by searching large text collections according to complex and linguistically motivated queries and algorithmas. It has three main integrants: a database management system called Manatee, a web interface search called Bonito and a web interface for corpus building and management called Corpus Architect. The corpus we investigate, CLC, is available in the Sketch Engine based on a special contract with Cambridge University Press. All corpus extraction and CQL search process have been carried out using this software.

3.3.2. Type / Token Ratio

In lexical frequency research, a series of lexical variation in both native and learner corpora are measured. The most common measure, the type/token (T/t) ratio, counts the number of different words in a text. It is computed by means of the following formula and the type/token results are used to draw conclusions on lexical richness in learner texts (Granger, 2002).

$$\text{T/t ratio} = \frac{\text{Number of word types} \times 100}{\text{Number of word tokens} \times 1}$$

In our study, we utilized the very same formula in our data analysis.

3.3.3. Log-Likelihood

When the differences between the frequencies of the items are found in a corpus, it is always possible that the difference we find is just a random, chance happening. In order to find out whether the differences are statistically significant, some tests are carried out. One of them is log-likelihood which helps us to normalize the sizes of the corpora we compare and reveal the significancies between frequencies. If results are significant,

we are reasonably certain (usually 95% certain, sometimes 99% certain) that these results are not due to chance.

When we obtain the frequencies for our corpus searches, we make the results comparable using frequencies by per cent, or by per million words. This is called normalising the frequencies. However, normalised scores do not mean that what we have is significant.

If the log likelihood for the result is greater than 6.63, the probability of the result- i.e. the difference between the two corpora - happening by chance is less than 1%. Thus, we can be 99% certain that the result actually means something. This is usually expressed as $p < 0.01$.

If the log likelihood is 3.84 or more, the probability of it happening by chance is less than 5% meaning 95% certainty of the result. This is expressed as $p < 0.05$.

3.4. Research Questions

1. How frequent are PVs in the written productions by Turkish EFL learners, as reflected in their sub-corpus of the CLC?
2. What are the most frequent PVs used by Turkish EFL learners across CEFR proficiency levels?
3. To what extent are PVs found in Turkish EFL learner's corpus well-formed across CEFR proficiency levels, i.e., verbs paired with appropriate preposition?
4. What PVs, if any, seem to be most problematic for Turkish EFL learners across CEFR proficiency levels?

Based on these research questions, we carried out our data analysis using the data analysis instruments mentioned above. The next chapter focuses on the data analysis procedure of our study.

CHAPTER IV

FINDINGS

4.1. Introduction

In the data analysis of the corpus data, in order to see the interlanguage system dynamics, we adapted a top-down approach, by gaining insight by analyzing the overall picture first. In other words, in the data mining and analysis process, we aimed at exploration and analysis of Turkish learner corpus data between A1-C2 proficiency band in order to discover meaningful prepositional verb patterns and pertaining errors by automatic or semi-automatic means.

Following this line of logic, as the first step PV errors data were extracted. We eliminated the preposition bundles which do not function as prepositional verbs. In the obtained data, prepositional verbs have been counted manually in order to determine and seperate *missing preposition, unnecessary preposition* and *replaced preposition* errors. After the learner corpus was analyzed in its entirety to give a bird-eye view, each of the subcorpus was analyzed separately to be able to make comparisons between the prepositional verb uses by the learners in each proficiency group. At this point, prepositions were counted according to their frequency levels and categorized across the levels.

4.2. Bird-eye View

In our PV subcorpus, the data includes 1441 use of PVs in total, 764 of which are erroneous in various typologies. A1-A2 proficiency level learners have 130 PVs used, B1-B2 level learners have 817 PVs and finally, 494 PVs are found in C1-C2 profiency range. Table 2 displays the cross tabulation of the total number of prepositional verb errors across the whole Turkish learner data.

Table 2

The Cross Tabulation of the Prepositional Verb Errors across the Whole Turkish Learner Data

Proficiency range	Corpus Size	PVs used	PV errors
A1-A2	31.277	130	93
B1-B2	143.191	817	478
C1-C2	101.501	494	193
TOTAL	**275.969**	**1441**	**764**

As we infer from the table, the learners in respective proficiency bands relatively underused PVs in comparison to the corpus sizes. The reason for this could be, as also Alonso (2002) observes with Spanish learners in his study, the avoidance strategy as the learners tend to avoid producing certain forms in their target language that they perceive as difficult. Turkish learners could prefer to underuse and/or avoid PVs as they feel insecure.

Table 2 displays the overall results without normalizing the corpus sizes. However, this might not give us a reliable idea about the use of errors in this regard. For this reason, we have calculated the log-likelihood scores across the proficiency levels normalizing the corpus sizes to reveal whether there are significant differences among the learners from A1 to C2. Table 3 displays the log-likehood results across levels.

Table 3

The Log-likehood Results across Levels

Group Comparison	LL	Significance
A1-A2 vs. B1-B2	1.07	0.301 -
B1-B2 vs. C1-C2	46.68	0.000 + ***

- indicates more prepositional verb errors in A1-A2 proficiency band relative to B1-B2 proficiency band

+ indicates more prepositional verb errors in C1-C2 proficiency band relative to B1-B2 proficiency band

The log-likelihood result between A1-A2 and B1-B2 level PV errors reveals that learners at B1-B2 level make more PV errors than A1-A2 and C1-C2 proficiency level learners. In this intermediary stage of English as a foreign language interlanguage, the learners start using more PVs both in terms of token and type (T/T= 478/817), and they make more mistakes. When we look at the type/token ratios which are T/T= 93/130 for A1-A2 band and T/T= 193/494 for C1-C2 band, we can infer this result validating the

finding.

Additionally, this phenomenon may be explained by U-shaped learning (Kellerman, 1985) as could be inferred from the the log-likelihood normalized results shown in the table. Erronous PVs seem to be the least at A1-A2 while B1-B2 level learners seem to have the most erroneous uses, this, again, decreases when reached C1-C2 proficiency range explaining that the learners seem to have learnt/acquired some PVs until they reach to C1-C2 proficiency range from A1-A2 level.

4.3. Analysis of Each Proficiency Band

4.3.1. A1-A2 Corpus Data

Table 4 displays the distribution of prepositional errors of the learners at A1-A2 proficiency range.

Table 4

The Distribution of Prepositional Errors of the Learners at A1-A2 Proficiency Range

A1-A2	to	*in*	for	of	with	about	on	at	from	Total
MT	42	0	3	3	2	2	1	0	0	53
RT	7	5	0	1	1	0	0	0	0	14
UT	25	0	1	0	0	0	0	0	0	26
T/T Ratio	**(74/96)**	**(5/5)**	**(4/15)**	**(4/5)**	**(3/3)**	**(2/2)**	**(1/3)**	**(0/1)**	**(0/0)**	**(93/130)**

As can be inferred from the table, the most common prepositional error typology is *to* which is followed by *in, for, of, with, about, on, at,* and *from* respectively according to their error frequencies. At A1-A2 proficiency range, 130 PVs are used, 93 of which are erroneous highlighting a significant importance in terms of type/token ratio. The most common error category was missing prepositions (MT) with the type of 53 errors out of 93 in total use. The most common prepositional error typology is *to* with a frequency score of 74. This preposition is missing whereas it is supposed to be used with the following verbs: *go to, come to, give to, listen to, talk to, write to, get to.* An example from the corpus is presented as follows:

- *<#SX> thinks | things </#SX>* . I wish you were here . I have been *<#MT>* | **to** *</#MT>* a lot of historic *<#AGN> place | places </#AGN>*

- you . If you want <#W> *to me* | *me to* </#W> , I will come <#MT> | **to** </#MT> your home . I want to see <#UT> *to* | </#UT> you . I am
- home <#MP> | , </#MP><#UT> *with* | </#UT> listening <#MT> | **to** </#MT> his <#RJ> *best* | *favourite* </#RJ> radio station

Second most common error typology is *for* and *of* which are each used for 3 times in the whole data. These errors are as in the examples:

- and phone number are <#S> *belows* | *below* </#S> : I am waiting <#MT> | **for** </#MT> your answer <#MP> | . </#MP> Best Regards
- <#MP> | . </#MP> I had a great time . Take care <#MT> | **of** </#MT> yourself.

While *with* and *about* PVs have only been missing for twice per each, other prepositions including *in, on, at* and *from* have not been used at all in A1-A2 data. Examples displaying PVs including *with* and *about* are:

- and phone number are <#S> *belows* | *below* </#S> : I am waiting <#MT> | **for** </#MT> your answer <#MP> |
- I can look <#RT> *at* | *in* </#RT> the mirror and when I look <#RT> **at** | **in** </#RT> the mirror I 'll remember Jenny . <#L> *Yours*

One of the most significant finding is that unnecessary preposition (UT) use is the second most common error across A1-A2 proficiency range. 26 errors out of 93 have been observed in PVs. What's more, only one of the errors is with *for* and the other 25 are unnecessary use of *to* preposition in PVs. Some of the samples of *to* error typology are *go to, come to, write to, listen to, give to* as exemplified below:

- They only cost £ 50 . Yours , Dear Ali , I went <#UT> **to** | </#UT> shopping for clothes yesterday.

As for the wrong uses (RT) of the PVs in A1-A2 level, only 14 out of 93 seems to be wrongly used by the learners. While 7 of these errors seems to be *to* typology that

should have been used with verbs such as *go to, come to, move to, write to, been to*; learners preferred to generally use *for* and *in*. *In* is the second most common wrong use in PVs bearing verbs like *look in* and *interested in;* some of which are as follows:

- hotel three times a day and are they included *<#RT>* **by | in** *</#RT>* the cost ? *<#DY>* *Finaly | Finally* *</#DY>* activities and
- everything about those two holidays . Look ***<#RT>* for | at *</#RT>*** all the details . Think where ***<#W>* *would you* | *you would* *</#W>***
- like *<#UP>* *it 's | its* *</#UP>* yellow flowers . You can call me *<#RT>* **from | on** *</#RT>* this number *<#RP>* *; | :* *</#RP>* xxxxx *<#MP>* | . *</#MP>* It is brown

4.3.2. B1-B2 Corpus Data

Table 5 indicates the distribution of prepositional errors of the learners at B1-B2 proficiency range.

Table 5

The Distribution of Prepositional Errors of the Learners at B1-B2 Proficiency Range

B1 - B2	to	in	about	for	on	with	at	of	from	Total
MT	109	21	29	26	15	19	8	8	5	240
RT	40	26	16	16	18	18	9	3	2	148
UT	35	12	7	7	11	4	7	5	2	90
T/T Ratio	**(184/333)**	**(59/96)**	**(53/70)**	**(49/119)**	**(44/48)**	**(41/58)**	**(24/31)**	**(16/44)**	**(9/19)**	**(478/817)**

As can be seen from the table, the most common erroneous typology is also *to* and the other erroneous error typologies are: *in, about, for, on, with, at, of,* and *from.* 817 PVs have been used by learners at the B1-B2 level. 478 PVs are wrong uses. 240 of the deviants belong to the missing prepositions which creates a remarkable difference compared to the wrong use (148) and unnecessary (90) prepositions. The most common PV error *to* has been used with some other prepositions instead of *attend to, take to, talk to, live up to, be to, invite to, go back to, look forward to, write to* such as;

- 13 June Dear <#UP> *Mr.* | *Mr* </#UP> James Felton I am writing <#MT> | **to** </#MT> you so as to underline that I am really
- *anyone* </#RA> . I took my music player and listened <#RT> **by** | **to** </#RT> my favourite song . It <#TV> *had calmed* | *calmed* </#TV> me down
- did n't explain the event which <#TV> *is* | *was* </#TV> related <#RT> **about** | **to** </#RT> the historical thing . They always display <#UD> *the* |

29 of these missing prepositions seem to belong *about* typology while 25 of them belong to *for* and 21 of them, *in* respectively. Other errorneous PVs include*, think about, go on, know about, look at/in, curious about, familiar with, fond of, get on with, angry with, catch up with, proud of, mention about, wait for*…etc. which have not been used by the learners. As can be seen from the PVs used, it is clear that learners had been using more various PVs compared to A1-A2 profiency range. Some of the PV samples extracted from B1-B2 level include:

- </#MD> lifestyle <#M> | which is </#M> typical <#MT> | of </#MT> these times </#W> . Children have <#MD> | their </#MD>
- <#UP> . Because | because </#UP> she never managed to get me out <#RT> from | of </#RT> the toy shop . As you see <#MP> | , </#MP> shopping is

When we compare Missing Preposition (MT category) data obtained from A1-A2 and B1-B2 proficiency bands using Log-Likelihood calculations based on the normalized corpus sizes, we get the following results as presented in Table 6.

Table 6

Log-Likelihood Results of Each Missing Prepositions between A1-A2 and B1-B2

Preposition	Freq. A1-A2	Freq. B1-B2	LL	Sig
to	42	109	8.91	0.003 ** +
in	1	21	3.60	0.058 -
for	3	26	1.30	0.255 -
of	3	8	0.58	0.445 +
with	2	19	1.17	0.279 -
about	2	29	3.50	0.061 -
on	1	15	1.88	0.170 -
at	1	8	0.32	0.572 -
from	1	5	0.01	0.935 -

+ indicates more missing prepositions in A1-A2 proficiency band relative to B1-B2 proficiency band,
- indicates more missing prepositions in B1-B2 proficiency band relative to A1-A2 proficiency band

As we can infer from the table, the most significant difference between the errors related with missing prepositions in PV patterns between A1-A2 and B1-B2 proficience bands appears to belong to the preposition *to*. B1-B2 level learners do not use this preposition in their related PVs. The other one is *of*. The use *of* preposition is also ignored in PVs by B1-B2 level learners. This does not mean that A1-A2 level learners are most successful in their use. B1-B2 level learners use more (both in terms of frequency and diversity) PVs in their sentences and, expectedly, make more errors with these particular prepositions in their learning process. As for the other prepositions, A1-A2 learners significantly make more errors due to the lack of knowledge probably stemming from less exposure to these patterns.

Second most common error type is the wrong use (RT) of PVs which are 148 in total, 40 of which are *to* typology. 26 are *in* errors while *about* and for have been wrongly used for 16 times and *on* as well as *with* have been used wrongly 18 times. Some of these wrongly used verbs are commonly used with *on, in, for, with* as in the examples*:*

- *anyone* </#RA> . I took my music player and listened <#RT> **by** | **to** </#RT> my favourite song . It <#TV> *had calmed* | *calmed* </#TV> me down
- and the national parks . It is nice to spend time <#RT> **for** | **on** </#RT> the environment .

- everything about those two holidays . Look <#RT> **for** | **at** </#RT> all the details . Think where <#W> *would you* | *you would* /#W>
- he left home . But he was <#S> *triembling* | *trembling* </#S><#RT> **of** | **with** </#RT><#UD> *the* | </#UD> excitement . First he <#S> *cought* |

When compared Wrong Preposition uses (RT category) between A1-A2 and B1-B2 proficiency bands using Log-Likelihood calculations based on the normalized corpus sizes, we get the following results presented in Table 7.

Table 7

Log-Likelihood Results of Each Wrong Preposition Uses between A1-A2 and B1-B2

Preposition	Freq. A1-A2	Freq. B1-B2	LL	Sig	
to	7	40	0.31	0.579	-
in	5	26	0.07	0.791	-
for	1	16	2.15	0.142	-
of	1	3	0.12	0.724	+
with	1	18	2.71	0.099	-
about	1	16	2.15	0.142	-
on	1	18	2.71	0.099	-
at	1	9	0.349	0.483	-
from	1	2	0.41	0.523	+

+ indicates more wrong preposition uses in A1-A2 proficiency band relative to B1-B2 proficiency band,
- indicates more wrong preposition uses in B1-B2 proficiency band relative to A1-A2 proficiency band

As seen in the table, the prepositions of and from have been relatively significantly used in a wrong way by B1-B2 level learners. They have difficulty choosing the correct preposition in related PVs in their written productions. As expected, A1-A2 level learners made more errors with the other prepositions. Considering the frequencies, these errors are rare but significant as they are underused.

In addition, 90 out of 478 of these PV errors across B1-B2 proficieny range belong to the unnecessary PV errors (UT) which had not been used by the learners at all. *To* typology, again, pioneer the most of these unnecessary errors following *in, on, about, for, at* and *from* respectively. Some extracted examples are as follows:

- tell you <#UC> *that* | </#UC> what happened . We went <#UT> to | </#UT>

there <#RT> *by* | *in* </#RT> Sarah 's car . We stayed in a tent.

- <#AGN> *part* | *parts* </#AGN> of a school , because I believe <#UT> in </#UT> the first impression <#M> | *you make* </#M> on
- would call you before he comes . I also apologise <#UT> for | </#UT> a hundred times because I broke the beautiful

When Unnecessary Preposition (UT category) data obtained from A1-A2 and B1-B2 proficiency bands using Log-Likelihood calculations based on the normalized corpus sizes is compared, we get the following results as presented in Table 8.

Table 8

Log-Likelihood Results of Each Unnecessary Preposition Uses between A1-A2 and B1-B2

Preposition	Freq. A1-A2	Freq. B1-B2	LL	Sig
to	25	35	18.27	0.000 +
in	1	12	1.13	0.288 -
for	1	7	0.18	0.675 -
of	1	5	0.01	0.935 -
with	1	4	0.01	0.905 +
about	1	7	0.18	0.675 -
on	1	11	0.90	0.343 +
at	1	7	0.41	0.675 -
from	1	2	1.50	0.523 +

+ indicates more unnecessary preposition uses in A1-A2 proficiency band relative to B1-B2 proficiency band,

- indicates more unnecessary preposition uses in B1-B2 proficiency band relative to A1-A2 proficiency band

The data suggests that the most common error in this category is preposition *to* in prepositional verbs followed by *in, on, for, at, about, of, with*, and *from* consecutively when A and B level students are compared. This might imply that the more prepositional verb patterns are learned, the more fluctuations in the performance we encounter in relatively more advanced learners in the B level transition stage towards C level of proficiency. Some examples include:

- 1,000 pesos ! <#RP> *your* | *Your* </#RP> friend , Dear Anna , I went <#UT> **to** | </#UT> shopping yesterday . I bought a purple shirt and
- <#AGN> *part* | *parts* </#AGN> of a school , because I believe <#UT> **in** |

</#UT> the first impression <#M> | *you make* </#M> on

- </#MP> which is not right <#MP> | , </#MP> I bet <#UT> **on** | </#UT> that there were <#RP> *5.000* | *5,000* </#RP> people or over <#RP> *5.000* |
- <#MP> | , </#MP> which was a slower way to contact <#UT> **with** | </#UT> people . On the other hand <#MP> | , </#MP> and <#R> *as a* | *in* </#R>

4.3.3. C1-C2 Corpus Data

The distribution of prepositional errors of the learners of C1-C2 proficiency levels is displayed on the table below.

Table 9

The Distribution of Prepositional Errors of the Learners at C1-C2 Proficiency Range

C1 - C2	to	with	for	on	in	about	at	of	from	TOTAL
MT	24	13	12	5	5	4	3	4	2	72
RT	8	10	8	11	11	7	7	3	0	65
UT	21	8	9	7	3	3	1	2	2	56
T/T Ratio	**(53/134)**	**(31/52)**	**(29/72)**	**(23/39)**	**(19/65)**	**(14/31)**	**(11/17)**	**(9/50)**	**(4/21)**	**(193/494)**

Frequencies of error typologies for *to, with, for, on, in, about, at, of,* and *from* prepositions are respectively presented in table 9. The most common error seems related to be the missing prepositions (72) which are followed by wrong use (65) and unnecessary preposition (56) uses. *To* is the foremost common error category of missing prepositions, as in the other proficiency levels which has been used in *write to, attend to, go back to, come back to, live up to* sequences. *With* and *for* prepositions including *keep up with, put up with, provide with, responsible for,* and *suitable for* are the other two significant PVs missed following *to* as can be seen from the examples below:

- *it 's* <#S> *definetly* | *definitely* </#S>| *it 's definitely been* </#TV> worth waiting <#MT> | **for** </#MT> as I have some great and <#S> *unbeliavable* | *unbelievable* </#S>
- Talk with Chris and write <#MT> | **to** </#MT> me soon so that I can make our reservations with the hotel.

- And the restaurant did not have any special menu which would provide us *<#MT>* | **with** *</#MT>* a different taste from Scotland.

When we compare Missing Preposition (MT category) data obtained from B1-B2 and C1-C2 proficiency bands using Log-Likelihood calculations based on the normalized corpus sizes, we get the following results as presented in Table 10.

Table 10

Log-Likelihood Results of Each Missing Prepositions between B1-B2 and C1-C2

Preposition	Freq. B1-B2	Freq. C1-C2	LL	Sig
to	109	24	33.47	0.000 +
in	21	5	5.85	0.016 +
for	26	12	1.58	0.208 +
of	8	4	0.34	0.562 +
with	19	13	0.01	0.922 +
about	29	4	13.74	0.000 +
on	15	5	2.38	0.123 +
at	8	3	0.96	0.327 +
from	5	2	0.50	0.479 +

+ indicates more missing prepositions in B1-B2 proficiency band relative to C1-C2 proficiency band,

- indicates more missing prepositions in C1-C2 proficiency band relative to B1-B2 proficiency band

As presented in the table, C1-C2 level learners make more errors in missing preposition category. In all prepositions displayed in the table, advanced level students significantly miss the required prepositions in their PV constructions. The variety of their use of verbs in PVs increase, so do the frequency of the missing prepositions in these constructions. They probably have difficulty in building these patterns although the variety increases.

As for wrong use of the prepositions (RT) at C1-C2 level, the most common wrong use of PVs seems to be *on* and *in* (11). After that, *with* (10), *to/for* (8), *about/at* (7) follow in order of frequency. Some extracted wrong uses are:

- They went *<#RT>* **in** | **to** *</#RT>* a café to talk about the new product they wanted to *<#S> launge* | *launch </#S>*

and <#MA> | *which* </#MA> unfortunately ended <#RT> **with** | **in** </#RT> great <#S> *dissappointment* | *disappointment* </#S>

- Scotland which I have been <#RT> **to**| **on** </#RT> and <#MA> | *which* </#MA> unfortunately ended <#RT> *with* | *in* </#RT>

Wrong Preposition uses (RT category) in the data obtained from B1-B2 and C1-C2 proficiency bands using Log-Likelihood calculations based on the normalized corpus sizes is presented in Table 11 below.

Table 11

Log-Likelihood Results of Each Wrong Preposition Uses between B1-B2 and C1-C2

Preposition	Freq. B1-B2	Freq. C1-C2	LL	Sig
to	40	8	13.69	0.000 +
in	26	11	2.19	0.139 +
for	16	8	0.67	0.412 +
of	3	3	0.18	0.674 +
with	18	10	0.39	0.532 +
about	16	7	1.20	0.274 +
on	18	11	0.15	0.697 +
at	9	7	0.03	0.854 -
from	2	1	0.08	0.772 +

+ indicates more wrong preposition use in B1-B2 proficiency band relative to C1-C2 proficiency band,

- indicates more wrong preposition use in C1-C2 proficiency band relative to B1-B2 proficiency band

When we consider the results in Table 11, we observe similar findings to the ones depicted in Table 10 presenting the missing preposition category. The findings are consistent in the sense that C1-C2 level learners either do not use the required preposition or choose a wrong one in their PV constructions. In their trial and error process phase of learning, the richness of the PVs increases, but they either wrongly determine the correct preposition or completely ignore in the related constructions.

Besides, unnecessarily used PVs are 56 out of 193 in total. Those generally comprise *thank for, wait for, agree with, contact with, spend on, provide with, to be keen on, knock on, shop for*…etc. A few of the samples are as follows:

- Once enlightened by the new technology , the population does not want to give up *<#UT>* **with** | *</#UT>* their curiosity about the new machine
- With this feeling of happiness , I went *<#UT>* **to** | *</#UT>* home immediately and took some fresh tea and I am looking forward to hearing from you . Dear Alison , I am writing *<#UT>* **about** | *</#UT>* some tips that you

The data on Unnecessary Preposition (UT category) use between A1-A2 and B1-B2 proficiency bands has been calculated using Log-Likelihood calculations based on the normalized corpus sizes, the results is presented in Table 12.

Table 12

Log-Likelihood Results of Each Unnecessary Preposition Uses between B1-B2 and C1-C2

Preposition	Freq. B1-B2	Freq. C1-C2	LL	Sig
to	35	21	0.37	0.543 +
in	12	3	3.13	0.077 +
for	7	9	1.41	0.235 -
of	5	2	0.50	0.479 +
with	4	8	3.09	0.079 -
about	7	3	0.56	0.453 +
on	11	7	0.05	0.823 +
at	7	1	3.23	0.072 +
from	2	2	0.12	0.731 -

+ indicates more unnecessary preposition use in B1-B2 proficiency band relative to C1-C2 proficiency band,
- indicates more unnecessary preposition use in C1-C2 proficiency band relative to B1-B2 proficiency band

The unnecessary / overuse use of prepositions in all prepositions except for *from, with* and *for* by C1-C2 level learners confirm the results displayed in Table 10 and Table 11. They seem to learn more verbs as lexical items and increase the variety in this regard but overgeneralize the use of the prepositions presented in the table.

One of the important finding is that some of the PVs are persistent across all the proficiency ranges such as *go to, come to, write to, give to, ask for, thank for, know about, look forward to, interested in, listen to, take care of, provide with*…etc. Even though verbs used have changed as proficiency increases, most of the preposition errors are similar. For instance:

- Talk with Chris and write <#MT> | **to** </#MT> me soon so that I can make our reservations with the hotel
- 13 June Dear <#UP> Mr. | Mr </#UP> James Felton I am writing <#MT> | **to** </#MT> you so as to underline that I am really
- as expensive as the Frene Hotel . Please write <#MT> | **to** </#MT> me as soon as possible so that we can reserve a

4.4. The Comparison of the PV Errors of the Learners across Levels

Our data consisted of 1441 PVs in total, 764 of which were erroneous. A1-A2 proficiency range had 130 PVs used, 93 of which are problematic; B1-B2 levels had 817 PVs, 478 of which are errorenous and finally, 494 PVs were in C1-C2 profiency range, 193 of these were wrong uses.

All the proficiency ranges have missing preposition errors as the foremost common error type. Those errors comprise almost half of the errors in total across 3 levels. Wrong use of prepositions is the second most common error typology except for A1-A2 proficiency range. A possible explanation for this might be the L1 transfer in this regard. This finding is in parallel with the previous research. L1 transfer hypothesis which argues that second language learning is a cognitive process guided by universal and innate mechanism and is similar to first language acquisition (Kellerman, 1983). Learners tend to transfer what is typologically marked and different structures than L2. Namely, typologically less marked ones are more likely to be transferred by the learner. The study investigating the acquisition of relative clauses by adult second language learners revealed how transfer is possible under some circumstances according to Gass (1979). According to her, some target language properties and language universals must be taken into account to understand how transfer works. For the current study, as prepositions do not exist in Turkish language, “postpositions” might be interpreted as PVs by the learners. Students may negatively transfer while constructing PVs in English. Some examples provided by Kornfilt are as follows:

(1577)	içeri gir -di	dışarı çık -tı
	Inside enter -past	outside exit -past
	“She went in”	“He went out”
(1578)	içeri **-ye** gir -di	dışarı **-ya** çık -tı
	Inside **-Dat**. enter -Past	outside **-Dat**. exit -Past

	"She went in"	"He went out"
(1579)	içeri **-de** kal -dı	dışarı **-dan** gel -di
	Inside **-loc**. stay -past	outside -**Abl**. come -Past
	"He stayed inside"	"She came from outside" (p. 452).

Some of the errors might result in the overuse of the same prepositions. Under this analysis, it would be the case that there is a markedness relation between the two structures, with being the unmarked member and the marked. In other words, in languages like Turkish, Korean, Japanese in which adpositions are always postpositional, these languages do not have such structures. Every language that does permit this either has both of the structures at issue (as in English and German) or only the prepositional one (as in French). The prepositions would thus be marked after verbs and the postpositional one unmarked. Within the same line of reasoning, some common reasons in the wrong use of prepositions are provided by Hong et al. (2011), one of which is interlingual transfer in collocations which occurs between the learners' L1 and L2. An example of intralingual transfer errors includes **dropped into the river* instead of *fell into the river.*

To start with, *to* is the most common preposition in A1-A2 proficiency range which has been generally followed by the verbs *go, come, give, listen, talk, write, get,* and *be.* These PVs are the most commonly used ones at this level. *To* as missing in PVs appears to be more than half of the PV errors in total. The second most common error typology is *in* which *look in, include in, to be interested in.* Other preposition errors are observed in *wait for, help with, call on, take care of, look at, think about, wait for,* and *to be interested in, arrive at, help with,* and *call on* PVs. Unnecessary PV errors (26) ranks as the second after missing prepositions category which is specific to A1-A2 level, 25 of these errors generate from *to* usage. Some extracted examples are as follows:

- I *<#TV>* ***loved*** | ***love*** *</#TV>* this because *<#UP>* **,** | *</#UP>* it 's full *<#RT>* **with** | **of** *</#RT>* memories .
- like *<#UP>* ***it 's*** | ***its*** *</#UP>* yellow flowers . You can call me *<#RT>* **from** | **on** *</#RT>* this number *<#RP>* **;** | **:** *</#RP>* xxxxx *<#MP>* | **.** *</#MP*
- If you want *<#W>* ***to me*** | ***me to*** *</#W>* , I will come *<#MT>* | **to** *</#MT>* your home . I want to see *<#UT>* ***to*** | *</#UT>* you. I am

Secondly, there exist 478 wrong formations out of 817 PVs in total across B1-B2

levels. Half of these errors belong to the missing preposition errors. Wrong use of prepositions comes second with 148 usages out of 478 and lastly, unnecessary preposition usage comes as the last with the frequency of 90 errors. The most erroneous verb and preposition combinations is *to* with 184 usages out of 333 in total, some of which are *write to, invite to, look forward to, be to, talk to*…etc. *In* is the second most common PV error (59) across this level used with the following verbs; *to be interested in, believe in, include in, come in*…etc. *About* (53) is another preposition error with almost the same frequency as *for* (49) which have been used with *complain about, know about, apologise for, to be happy about.* Other PV typologies are *on* (44), *with* (41), *at* (24), *of* (16), *from* (9) respectively according to their frequencies. Some of these erroneous prepositional verb combinations are as follows: *spend on, agree with, tremble with, get out of, think of/about, know about, include in, look forward to, be to, go on*…etc. Some errors in the patters in this respect are exemplified below:

- This holiday will help me to think *<#MT>* | **about** *</#MT><#MA>* | *it </#MA>* and prepare it . I 'd prefer
- would call you before he comes . I also apologise *<#UT>* **for** | *</#UT>* a hundred times because I broke the beautiful
- *my </#DD>* opinions *<#MC>* | *and </#MC>* I hope you 'll agree *<#MT>* | **with** *</#MT><#AGA> it* | *them </#AGA>*. Because I 've *<#RV> done* | *told </#RV><# MA>* | *you </#MA>*

When compared with A1-A2 level, being the most common error typology, missing preposition category is approximately half of the total wrong prepositional patterns. Prepositions to be replaced with the correct ones appears to be the second most common errors in B1-B2 levels while A1-A2 proficiency range has wrong use of PVs in the third place. Unnecessary PV errors are the second most common PV errors for A1-A2 levels.

As for C1-C2 proficiency range, the error categories according to their frequencies are missing, wrong use and unnecessary preposition usages respectively. The most common type is missing prepositions with 72 errors out of 193 total errors. *To* is also the most common error typology among all of the PVs like the other two proficiency ranges. *To* is followed by *with* (31), *for* (29), *on* (23), *in* (19), *about* (14), *at* (11), *of* (9), and *from* (4) in error ranking. In addition to the PVs with *to* across the other levels, not any different

PVs occur in C1-C2 levels, however, as for the other PVs, *thank for, wait for, agree with, ask for, think of/about, spend on, shop for, contact with, keen on, interested in,* and *provide with* are some of the samples, which are the most common PVs found at this level. Some of the extracted instances are as follows:

- This is absolutely not true , because we only promised *<#UT>* **for** | *</#UT>* 35 stalls and we kept that promise .
- Talk with Chris and write *<#MT>* | **to** *</#MT>* me soon so that I can make our reservations with the hotel .
- And the restaurant did not have any special menu which would provide us *<#MT>* | **with** *</#MT>* a different taste from Scotland

Some of the PV errors made across 3 proficiency ranges are in line with examples Wilcoxon (2014) provided in her study with PVs in emergent academic writing such as *listen to, pay for, to be based on, give to, think about, agree with, depend on, know about, spend on, take care of.* When the most frequent PVs are considered, it is as follows:

Table 13

The Most Frequent Prepositional Verbs Used by Turkish EFL Learners

A1-A2	*go to, come to, listen to, give to, get to, wait for, think of/about, take care of, look at/in, interested in, arrive at*
B1-B2	*attend to, take to, live up to, be to, invite to, look forward to, write to, think about, go on, know about, look at/in, talk to, mention about, wait for, include in, ask about, call on, help with*
C1-C2	*wait for, thank for, ask for, spend on, keen on, provide with, knock on, agree with, think of/about, help with, write to, attend to, go back to, come back to, live up to*

The most frequent PVs used by Turkish EFL learners are as indicated in the table. Most of them are similar across levels which may be an instance of fossilization in this regard as those seem to become permanent across three levels.

Lastly, although learners seem to produce similar PVs across all the profiency levels as could be seen from the examples, there is an increase in the diversity of the verbs

in the use of correct PV patterns usages in C1-C2 levels such as *make use of, live up to, concentrate on, take part in, come up with, exert over, prevent from,* and *take account of.* As the proficiency level increases, learners seem to learn the related verbal patterns as chunks rather than learning verbal elements as lexical and prepositions as functional categories respectively. This might be leading the correct use of the prepositional verbs above without any fluctuations in the performance.

CHAPTER V

DISCUSSION

5.1. Introduction

In this section, under the light of the findings as a result of the data analysis, we would like to discuss the answers to our research questions.

5.2. Research Question 1

How frequent are PVs in the written productions by Turkish EFL learners, as represented in their sub-corpus of the CLC?

In the related literature, it has been consistently emphasized that the frequency of PVs across registers and genres is too common and many. Statistically speaking, Biber et al. (1999) indicate that PVs are found in all registers and genres, almost four times more common than phrasal verbs.

Such relatively frequent use of PVs has also been claimed to be causing one of the most frequent errors in learning English as a foreign language. This is also depicted in Figure 1 (Page 14). In that figure, the most common interlanguage errors of Turkish EFL learners is displayed. In our study, Table 2 (Page 52) displaying the cross tabulation of the errors also indicates the type and token ration (T/T: 764/1441) in this regard revealing that almost more than half of the PVs used are errorenous, which supports the findings shown in Figure 1 in Can (2017). As for the log-likelihood result across levels, PV errors reveal that learners at B1-B2 level make more PV errors than A1-A2 and C1-C2 proficiency level learners. More PVs, both in terms of token and type (T/T= 478/817), are used in this intermediary stage of English, by the learners. When we look at the type/token ratios which are T/T= 93/130 for A1-A2 band and T/T= 193/494 for C1-C2 band, we can infer this result validating the finding. While A1-A2 level learners make the least of errors, it increases the top when reached at the intermediary level which might be an instance of U-shaped learning (Kellerman, 1985) in this regard.

At A1-A2 proficiency range, 130 PVs were used, 93 of which are erroneous highlighting a significant importance in terms of type/token ratio. 817 PVs had been used by learners at the B1-B2 level. 478 of these PVs have been erroneous. 240 of the incorrect uses belong to the missing prepositions which creates a remarkable difference compared

to the wrong use (148) and unnecessary (90) prepositions. Finally, 494 PVs are used in C1-C2 profiency range and 193 of these are incorrect uses.

5.3. Research Question 2

What are the most frequent PVs used by Turkish EFL learners across CEFR proficiency levels?

Some certain PVs including *go to, come to, listen to, give to, get to, wait for, think of/about, take care of, look at/in, interested in, arrive at* are used as the most frequent PVs by A1-A2 level learners.

At B1-B2 proficiency range, *to* is the most remarkable preposition in the prepositional verb patterns among all other prepositions in missing, wrong use and unnecessary preposition error categories. The most frequent PVs used at this level are *attend to, take to, live up to, be to, invite to, look forward to, write to, think about, go on, know about, look at/in, talk to, mention about, wait for, include in, ask about, call on, help with*. Most of these PVs are similar to the ones which are problematic in A1-A2 level.

Last but not least, C1-C2 level learners use *wait for, thank for, ask for, spend on, keen on, provide with, knock on, agree with, think of/about, help with, write to, attend to, go back to, come back to, live up to* as the most frequent PVs.

All in all, *go to, come to, write to, listen to, wait for, think of/about, ask for, call on, to be interested in, come out of, get back to, get on with, agree with, get to, include in, look forward to, talk to,* seem to be the most frequent PVs used across all of proficiency ranges as can be seen from the examples as well as preposition frequencies in the findings.

As for the errors in this regard, learners overused the same prepositions and there are two common types of collocation errors regarding PVs. Learners either select an *incorrect preposition* or simply *omit* it. The most common incorrectly used prepositions are *in, to,* and *of*, which take place in the top ten list of most frequent prepositions which is in line with the findings found by Tetreault and Chodorow (2008).

5.4. Research Question 3

To what extent are PVs found in Turkish EFL learner's corpus well-formed across CEFR proficiency levels, i.e., verbs paired with appropriate preposition?

As can be seen from Table 13, type and token ratio for the correct PVs is 664/1441 in total. A1-A2 proficiency range learners use 37 correct PVs out of 130 PV uses in total

having a 37/130 type and token ratio. 22 of these uses belong to the preposition *to* while other 11 uses are related with the preposition for, when compared with the other proficiency ranges. It seems that PVs are not well formed across A1-A2 level.

Table 14

Correct Uses of Prepositional Verbs

Proficiency range	Total PVs	Correct PVs	Percentage (%)
A1-A2	130	37	28.46%
B1-B2	817	339	41.49%
C1-C2	494	288	58.29%
TOTAL	**1441**	**664**	

Table 13 displays that, for B1-B2 level, type and token ratio for the correct PV uses is 339/817. 149 of these prepositions belong to *to* PV patterns, and then *for* (70), *in* (37), *of* (28), *about* and *with* (17), *from* (10) and *on* (4) follows in order which has been mentioned in the findings. Compared to A1-A2 level, intermediate learners seem to be using more and various correct prepositions in PVs.

Lastly, 288 correct prepositions had been produced by C1-C2 level learners, 81 of which are *to* prepositions. *In* (46), *for* (43) and *of* (41) follow respectively. which is the same sequence as the B1-B2 level revealing a remarkable finding.

One of the interesting findings has been that all the profiency range learners use *to* in PVs both as the correct uses and as the most common wrong patterns. Correct PV uses increase as the proficiency increases.

When looked at the percentages between proficiency ranges, it can be inferred that C1-C2 level learners used PVs more correctly than other levels even though they may still struggle using correct certain PVs when necessary. Additionally, this phenomenon may be explained by U-shaped learning (Kellerman, 1985) depending on the log-likelihood normalized results as could be inferred from the percentages shown in the table. Erronous PVs seem to be the least at A1-A2 while B1-B2 level learners seem to have the most erroneous uses, this, again, decreases when reached C1-C2 proficiency range explaining that the learners seem to have learnt/acquired some PVs until they reach to C1-C2 proficiency range from A1-A2 level.

One of the reasons why wrong uses of prepositions is more common than correct

uses in Turkish learners' written production might be due to the fact that L2 students are taught and encouraged to look up unknown words in a dictionary, they are not usually taught and encouraged to study the related lexical items in the context of their collocates, as also stated in Hong et al., (2011). This causes unknown verbs being used in wrong patterns, which could not inevitably affect meaning.

In their study, Morris and Cobb (2004) state that L2 learners commonly apply avoidance strategies when it comes to using MWIs. Laufer and Waldman (2011) share the same view, proposing that "EFL learners have a tendency to overuse a limited number of learned verb+preposition collocations formed from common verbs such as 'be, have and make'" (p. 652). This may be the reason that PVs Turkish EFL learners used in their production are not varied and some of the typologies are quite similar across all the levels.

5.5. Research Question 4

What PVs, if any, seem to be most problematic for Turkish EFL learners across CEFR proficiency levels?

Turkish EFL learners seem to have difficulty in determining the correct preposition for the prepositional verb patterns. This can easily be seen from the most problematic PV type which is missing preposition (MT) across all proficiency ranges. As they have difficulty in finding the appropriate preposition, they could be avoiding using one or choosing not to use a preposition.

A1-A2 level learners use some certain PVs including *go to, come to, listen to, give to, come to, get to, wait for, think of/about, take care of, look at/in, interested in,* and *arrive at* which prove to be the most problematic PVs at this level. Preposition *to* is usually used in a wrong way when necessary, which seems to be the most problematic PVs at this level along with unnecessary preposition usage except the wrong use of prepositions. This level is the only one which unnecessary verb use excels the wrong use of preposition. The reason for this might be the overextension of the use of them in PVs due to the limited knowledge of the prepositions.

Learners tend to transfer what is typologically marked and different structures than L2. Typologically less marked are more likely to be transferred by the learner (Benson, 1986). For all the profiency ranges, learners may be negatively transferring while producing PVs in English since postpositionals exist in Turkish language.

Also, the lack of the knowledge of the rules could cause the learner's competence on the restrictions and exceptions of target language rules. Learners may not have been exposed, taught, or scaffolded related with the use of prepositions, especially certain PVs, in the target language. They surely have the lack of the knowledge of the rules about these at all the proficiency ranges. The number of unnecessary preposition usages across A1-A2 level might be an instance of this phenomenon. Also, the number of errors across B1-B2 level also indicate this phenomenon as learners are still learning at this intermediary stage, adopting some trial and error strategies.

As for B1-B2 proficiency range, *to* is the most remarkable component among all other prepositions in *missing, wrong use* and *unnecessary* preposition usages. The most problematic PVs seem to be *attend to, take to, live up to, be to, invite to, look forward to, write to, think about, go on, know about, look at/in, talk to, mention about, wait for, include in, ask about, call on,* and *help with.* Most of these PVs are similar to the ones which are problematic in A1-A2 level.

Lastly, C1-C2 level learners have difficulty in using PVs such as *wait for, thank for, ask for, spend on, keen on, provide with, knock on, agree with, think of/about,* with *write to, attend to, go back to, come back to, live up to, take part in, make use of, take into account, put up with, ashamed of, get rid of, capable of* and *blame for.* Among all, *to* had been the foremost common error in all missing, wrong use and unnecessary preposition use error categories at this level. For this level, *fossilization* could explain one of the causes of errors which can be defined as persistent non-target-like structures in the learner/user's interlanguage or inability of a learner to produce native-like performance in the target language. For some of the learners, fossilization may be a reason in this respect

When the type and token ratios are considered, which is 93/130 for A1-A2, 478/817 for B1-B2 and 193/494 for C1-C2 levels, it could be inferred that even though learners seem to be using similar PVs, they simply omit the required preposition. The reason for this is that they are not sure about what prepositions to use or simply avoid using those structures because they do not feel confident as mentioned in the discussion chapter. Even though verbs used varied as proficiency level increases, prepositions in PVs used are rarely change across levels.

Some of the most common PV errors in each one of the registers are in line with Longman Grammar of Spoken and Writen English's (Biber et al., 1999) most common PVs list which consist of *go to, write to, listen to, wait for, think of/about, call on, to be*

interested in, come out of, get back to, get on with, come up with, look forward to, talk to and *ask for.* In fact, the most frequent PVs used by Turkish EFL learners across all the profiency levels prove to be the most problematic ones at the same time revealing that as usage increases, error frequency also appear to be increasing

Even though PVs vary as profiency increases, errors also increase. That some of these PV errors are persistent across all the proficiency ranges along with being the most problematic PVs for Turkish EFL learners. These PVs include:

come to, write to, give to, thank for, know about, go to, write to, listen to, wait for, think of/about, wait for, call on, to be interested in, come out of, get back to, get on with, come up with, look forward to, talk to, ask for, provide with, knock on, agree with. Shop for, contact with and *keen on.*

One of the possible explanations might be that Turkish EFL learners overuse some of the certain PVs in their written productions. The reason for this might be the limited exposure to the varied use of such patterns in their materials and in learning environments.

On the other hand, another reason might be owing to Brown's (1973) order of morphemes acquisition which claims that between 12 and 26 months, children are expected to have a certain number of MLUm's (mean length of utterance measured in morphemes) and as they acquire more language, they gradually increase this number. They move from first stage to second stage where they learn to use "-ing" endings on verbs, "in", "on", and "-s" plurals. This is where the first prepositions are learnt in first language acquisition which may also be a possible order for the learners' second language acquisition giving another possible reason for the learners' errors on *in* and *on* prepositions across levels.

5.6. Discussion of the Findings

Some causes of these PV errors have been mentioned in review of literature (p. ?). To start with, one of the main causes of errors in second language learning would be language transfer which argues second language learning is a cognitive process guided by universal and innate mechanism and is similar to first language acquisition (Kellerman, 1983). Learners tend to transfer what is typologically marked and different structures than L2. Typologically less marked are more likely to be transferred by the learner. For the current study, as prepositions do not exist in Turkish language, rather, "postpositions" have the same interpretations as PVs, students may negatively transfer while producing

PVs in English.

The incorrect structures produced by the learner could be explained by their limited knowledge of L2 and exposure to other structures of the target language, which might also have referred as *overgeneralization* errors. Learners could have a tendency to overgeneralize the PVs where they are not sure what to use while forming the L2 structures as this can be seen from the most common error typology *to* which has been overused across all the proficiency ranges. It may have been overgeneralized by the learners.

Another cause of errors might stem from simplification which is a process L2 learner replaces a complex structure of the target language with a simpler one. As can be seen from Davydova's (2011) study, learners tend to use simpler tenses and clauses. Similarly, Turkish EFL learners may have wrongly simplified while producing PVs in English omitting certain prepositions.

In addition, *underuse* and/or *avoidance* could be another cause of the errors which learners tend to avoid producing certain forms in their target language that they perceive as difficult (Alonso, 2002). Learners usually underuse and/or avoid certain structures when they feel not confident. Turkish EFL learners, in this case, may have avoided using some of the prepositions including *about, of, from, with,* and *at* since these prepositions appear to be very few across proficiency ranges compared to other prepositions such as *to, in, for.* In their study, Morris and Cobb (2003) also similarly found out that L2 learners commonly apply avoidance strategies when it comes to using MWIs.

On the other hand, some of those PV errors might be stemming from interlanguage properties which can be defined as persistent non-target-like structures in the learner/user's interlanguage or inability of the learners to produce native-like patterns in the target language. In turn, fossilization occurs to prevent them to use PVs in an efficient and right way.

Learners may be transferring some of the features from L1 as can be expected. As Turkish have postpositions, learners' mother tongue may interfere with their correct PV usages in language production. For instance:

(1577)	içeri gir -di	dışarı çık -tı
	Inside enter -past	outside exit -past
	"She went in"	"He went out"

(1478) Hasan sinema -ya [ben -den **önce**] git -ti

Hasan cinema -Dat. I -Abl. **Before** go -Past
"Hasan went to the movies before me" (p. 425).

Furthermore, simplification is a process when L2 learner replaces a complex structure of the target language with a simpler one. Turkish EFL learners may have simplified some of the structures while producing PVs in English because of the fact that they had used the similar PVs unnecessarily more than needed such as *go to, come to, write to, listen to, ask for, thank for,* and *interested in* across all the profiency levels.

Lastly, those errors may have stemmed from underuse and/or avoidance strategies which are some other interlanguage properties that learners tend to avoid producing certain forms in their target language that they perceive as difficult (Alonso, 2002). Learners might prefer underuse and/or avoid certain structures as they are afraid of making errors. Turkish EFL learners could be avoiding using some of the prepositions including *about, of, from, with, at* as these prepositions appear to be very few across proficiency ranges compared to other prepositions such as *to, in, for.*

Hong et al. (2011) explicitly underlined that according to both the early error analyses and the more recent corpora studies, many prepositional verb errors are interlingual (i.e., errors induced by L1 influence). Also, Nesselhauf (2005) reported that almost 50% of errors are caused by L1 influence. In this case, a most likely cause may be transfer from the first language according to Hong et al. (2011) and Laufer and Waldman (2011).

The findings are consistent with that of Laufer and Waldman (2011) who propose that "EFL learners have a tendency to overuse a limited number of learned verb+preposition collocations formed from common verbs such as 'be, have and make'" (p. 652).

Furthermore, *fossilization,* especially with C1-C2 level, could be one of the causes of errors which can be defined as persistent non-target-like structures in the learner/user's interlanguage or inability of a learner to produce native-like performance in the target language. For some of the learners, fossilization may be a reason in this respect.

CHAPTER VI

CONCLUSION AND SUGGESTIONS

6.1. Conclusion

In this study, the aim was to contribute to our understanding of how Turkish EFL learners use PVs in their written productions and describe possible pedagogical implications. As the research on this area suggests, it is imperative for English learners to acquire and use multi-word items so as to use them in written and spoken production in a more fluent and effective way (Laufer & Waldman, 2011; Hong et al., 2011). The literature in this area highlights the importance of prepositional verbs in L2 learning (Hong et al., 2011); however, there is a lack of empirical studies of these verbs in ESL and EAP writing, and studies carried out in this area are scarce.

In this study, it has been also emphasized that the acquisition of PVs is difficult for L2 learners because of the fact that as Biber et al. (1999) claim PVs are commonly found across registers and genres. Also, PVs display a challenge for EFL learners owing to their complex behaviour. Tetreault & Chodorow (2008) claim that one of the causes of such difficulty stems from the use of prepositions. Possible combinations as well as limitless preposition choices create such difficulty. Because of their patterning challenge, even the native speakers of English have difficulty in choosing the correct preposition in a cloze-test task, yielding a 75% success rate.

Moreover, according to Hong et al. (2011), despite the fact that L2 students are taught and encouraged to look up unknown words in a dictionary, they are not usually taught and encouraged to study the related lexical items in the context of their collocates causing unknown verbs being used with the inappropriate preposition combinations. Studies similar to the current one carried out here could help to make more focused and adequate classroom decisions regarding type and amount of instruction and feedback provided to students in relation to multi-word items.

In this study, it has been indicated that one of the most common errors of Turkish EFL learners have proven to be use of PVs across levels. Learners have difficulty in finding the appropriate prepositions in their written productions. They have errors as to missing preposition; wrong use of preposition and unnecessary preposition uses. Missing prepositions have proven to be the foremost common error across all the profiency ranges. The reason might be underuse and/or avoidance phenomenon as learners tend to avoid

producing certain forms in their target language once they perceive as difficult. Turkish EFL learners may have avoided using some of the prepositions as they are afraid of making errors.

A1-A2 level learners used unnecessary prepositions as the second foremost common error type. Learners, at this stage, might not be sure about whether they should use any preposition or not. That is why, they may be ending up using any prepositions. The lack of the knowledge of the rules might cause the restrictions and perceiving the exceptions of target language rules. Learners may not have known, learnt, or acquired prepositions, especially certain PVs, in the target language. Learners may be using those wrong PV patterns due to the lack of knowledge of the rules at all the proficiency ranges. The number of unnecessary preposition uses across A1-A2 level might be an instance of this phenomenon.

Second most common error type is wrong use of prepositions for B1-B2 and C1-C2 level learners. At the intermediary levels, students are at the stage of learning and this may be the reason why they produce wrong use of prepositions more than the other proficiency ranges. At this intermediary level, students are at the stage of learning and this may be the reason why they produce wrong use of prepositions more than the other proficiency ranges.

To preposition has been the foremost common error type in all the levels across missing, wrong use and unnecessary preposition uses. With this in mind, learners also use some of the certain PVs across all the profiency ranges. It can also be inferred that certain PVs are used across all the ranges. However, only a few different PV correct usages could be seen from C1-C2 levels such as *make use of, live up to, concentrate on, take part in, come up with, exert over, prevent from,* and *take account of.* In this context, EFL learners seem to have a tendency to overuse a limited number of learned verb+preposition collocations formed from common verbs as can be clearly seen from the findings.

Even though PVs used vary as the profiency increases, some of the problematic PVs across levels appear to be the similar.

Language transfer could be another explanation for all these errors. Since prepositions do not exist in Turkish language, rather, “postpositions” have the same interpretations as PVs, students may negatively transfer while producing PVs in English.

Analysing learner errors with computer-aided error analysis, as in this study, gives insights about how language is acquired/learnt, what stages learners experience while learning/acquiring their language. It also gives teachers practical ideas about how to teach

certain structures, what to focus directly and indicate the similarities as well as differences between mother tongue and target language when necessary. Lastly, these findings could be integrated into the classroom setting which would contribute to the instructional purposes of the local institutions.

6.2. The Contribution of Corpus Study to EFL Teaching

The use of learner data in the classroom exercises such as comparing learner and native speaker data and analysing errors in learner language has been suggested to raise awareness (Meunier, 2002). All these help students become aware of gaps between their interlanguage and the language they are learning.

Studies that have been carried out so far indicate that corpora help to decide what to teach (indirect uses) than how to teach (direct uses) (McEnery and Xiao, 2010). Indirect uses of corpora are well established. However, direct uses of corpora in teaching are mostly restrained to advanced levels (i.e. higher education). Corpus-based learning activities almost never takes place in teaching English as a Foreign Language (EFL) classes at lower levels (i.e. secondary education) since teachers are not trained necessarily and they cannot access to appropriate corpus resources. To bridge the gap between corpora and ELT, "creating corpora that are pedagogically motivated, in both design and content, to meet pedagogical needs and curricular requirements so that corpus-based learning activities become an integral part, rather than an additional option, of the overall language curriculum" (McEnery and Xiao, 2010, p. 372-375).

While designing such corpus-based learning activities, learners' age, experience and level as well as their integration into the overall curriculum need to be taken into account. When the teachers are trained accordingly and their access to appropriate resources are provided, corpora can be implemented as a learning tool. Corpora are not only to help teachers to decide what to teach, they are also resources from which learners may learn directly in an autonomous manner (Gavioli and Aston, 2001).

Mauranen (2004) argues that for corpora to fit in language teaching, they need to be integrated into teacher education as well as published teaching materials, which are the backbone of most foreign language teaching. To make a serious contribution to language teaching, corpora must be adopted by ordinary teachers and learners in ordinary classrooms providing ordinary teachers with initiation and practice along with materials they can easily apply in their everyday teaching.

In their studies, Seidlhofer (2002) and Mukherjee and Rohrbach (2006) used a "local learner corpus" containing students' writings and they led students to solve questions about their own or classmates' writings, analyse and correct errors in such familiar writings. This has resulted in a significant awareness and had an impact on autonomy.

Glisan and Drescher (1993) claim quite convincingly that "[…] if grammar is to be taught for communicative purposes, the structures presented should reflect their use in current-day native speaker discourse" (p. 24). This is most conveniently possible by corpus-driven teaching.

Römer (2004) also emphasize the lack of studies analyzing EFL teaching materials. She comments on these books about their simplified, non-authentic kind of Englishes. Besides, pupils are generally provided with invented sentences, which probably do not take place in any natural speech situation.

In her study, she focused on the potential of construction and use of an electronic corpus of EFL textbook texts. This case study on if-clauses in spoken English (BNC) and "school" English (211 if-clause samples in the two GEFL TC subcorpora) proved that corpora can lead to valuable insights for teachers and how they may help to improve teaching materials. Discovering the kind of English, we teach in the classrooms requires more analytical and comparative studies of textbook language and its differences from "real" English, and the status of authenticity in ELT (Römer, 2004). Corpus-informed comparisons of authentic English and "school" English comparison studies may provide invaluable insights for linguists, language teachers, and language learners.

6.3. Implications for English Language Teaching as a Foreign Language

The importance of multiword units, in general, has been indicated in most of the studies (Gardner & Davies, 2007; Levin, 1993; Tetreault & Chodorow, 2008; Claridge, 2000) and the importance of prepositional verbs, in particular, have been emphasized by many researchers (Römer, 2008; Biber et al., 1999; Hong et al., 2011; Baldwin, 2005).

Prepositional verbs have proved to have difficulty for language learners including advanced learners. One of the reasons are caused because of virtually endless verb combinations of prepositional verbs according to Baldwin (2005), even though there are 34 prepositions alone. That learner language corpora could provide an invaluable source of data on language performance have been realized.

As the research on this particular area suggests, it is extremely necessary for English learners to acquire and practice multi-word items in order to use written and spoken English in a more fluent and effective manner according to Laufer & Waldman (2011).

Hong et al. (2011) explicitly underlined that according to both the early error analyses and the more recent corpora studies, many prepositional verb errors are interlingual (i.e., errors induced by L1 influence). Also, Nesselhauf (2005) reported that almost 50% of errors are caused by L1 influence. In this case, a most likely cause may be transfer from the first language according to Hong et al. (2011) and Laufer and Waldman (2011).

Nesselhauf (2005) believes that in some of the instances where a particular prepositional verb was not produced whereas it would have been appropriate, one prepositional verb and one collocation in particular appeared to be difficult for the learners.

Studies on remedial pedagogical solutions (Anderson, 2006) discusses the notion of semantic priming as a way a 'priming' word may provoke a particular 'target' word. As an instance, previously given the word body, a listener will recognise the word heart more quickly than once s/he had previously been given an unrelated word such as trick; in this regard, body primes the listener for heart. The word body sets up a word association with heart, while the word trick does not (Hoey, 2005).

What Lexical Priming proposes a radical theory of the lexicon is how words are used in the real world. According to classical theory, grammar is generated first and then words are dropped into, however; Hoey's theory argues that lexis is complexly and systematically structured and that grammar is an outcome of this lexical structure (Hoey, 2004). The notion of 'collocation', the property of language whereby two or more words seem to appear frequently in each other's company (e.g., 'inevitable' and 'consequence'), gives a clue about how language is organized. "Using concrete statistical evidence from a corpus of newspaper English, but also referring to travel writing and literary texts, the author argues that words are 'primed' for use through our experience with them, so that everything we know about a word is a product of our encounters with it. This knowledge explains how speakers of a language succeed in being fluent, creative, and natural" (Hoey, 2005, p. i).

When encountered a word, the mind subconsciously keeps a record of the context and co-text of the word. As we re-encounter the word (or syllable or combination of

words), we build up a record of its collocations. We are primed by each encounter, therefore; while using the word (or syllable or combination of words), the contexts we previously encountered are replicated (Hoey, 2005).

With the theory of priming, Hoey (2004) also explains "the only explanation that seems to account for the existence of collocation is that each lexical item is primed for collocational use" (p. 386). Collocation is not only explained for lexical priming, but it also describes everything we know about a word. That the mind has a mental concordance of every word which it encountered, a concordance that is richly annotated for social, generic, physical, discoursal and interpersonal context (Hoey, 2005). According to him, these are all connected to word association games.

Furthermore, priming is not simply between words, multi-word expressions, for instance, may have collocations that are distinct from those of the words that make them up. Sinclair (2004) exemplifies with his well-known analysis of the expression *naked eye* to illustrate how phrases and collocations can combine in texts. In this context, the words within the multi-word expression are primed to co-occur, and the co-occurring combination is also primed for particular collocational patterns. Hoey (2005) labels nesting to identify this phenomenon which simplifies the memory's task according to Krishnamurty (2003). Primings nest and combine. For example, winter collocates with *in*, producing the phrase *in winter* (Hoey, 2005), however, this phrase has its own collocations, which are separate from those of its components. Therefore, *in winter* collocates with several forms of the word BE (i.e. is, was, are, were, etc.). In this case, an instance of a nesting may be presented as;

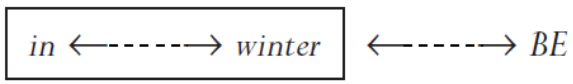


Figure 4. An instance of a nesting (p. 11)

Another instance would be the word collocates with *say, say a word*; in turn collocates with against, and say a word against collocates with won't. (i.e. I won't say a word against) (Hoey, 2005, p. 11). In this way, lexical items (Sinclair, 1999; 2004) and bundles (Biber et al., 1999) are created.

As lexical priming implicates, our experiences of language and primings coming out of these experiences are totally different from one another. As communication takes

place, harmonising principles try to ensure that our priming is not too different from one another. To exemplify, education and self-reflexivity are two of these harmonizing elements (Hoey, 2005).

Similarly, Hopper & Thompson (1980) also points out grammar as an output of what he calls 'routines', collocational groupings and the repeated use of what is learnt. It may be called as 'emergent grammar' and each individual's grammar differs since each person's experience and knowledge of routines are different. In this context, he underlines the importance of priming.

Hoey (2005) claims the transfer of primings from the first languages to later languages is unavoidable, so long as the learner does not learn in an immersive environment and never attempts to word-for-word translation (a rare set of circumstances indeed). Some of the learning practices taken place in the classrooms are not appropriate, such as the learning of vocabulary in lists (i.e. stripped of all its primings), On the other hand, the collocations, colligations, semantic associations of the language can be preserved by only authentic data and only complete texts and conversations preserve the textual associations and colligations.

Learners (or more accurately, types of learning) are recognized their being native or non-native however, they are recognized by the way their primings come into existence. When a speaker is surrounded by evidence (i.e. complete L2 exposure environment), in marked contradiction of early claims made by Chomsky (e.g. 1965) – the primings get built up inductively at variable speeds. Nonetheless, if the speaker is not so surrounded, then, other learning strategies need to be used.

Willis (2003) attempts to integrate grammar and vocabulary teaching in a complete natural manner to the learner by seeking to teach the collocations of the most common words in an integrated way with grammar and lexis. Since most of the language teaching materials cause unhelpful primings, as McEnery (2003) mentions that a textbook's overemphasis on certain features of the language or on its fabricated illustrations of grammatical points are some of the causes of unhelpful primings. These will eventually lead to cracks in the priming once the learner exposes to the authentic language, which would result in insecurity and distrust of what has been learnt in the classroom environment (Hoey, 2005; McEnery, 2003).

Muller-Hartmann and Schocker-von Ditfurth (2004) note that "lexical patterns become the major category in learning and teaching discourse" (p. 93) criticizing traditional approach of dealing with vocabulary and grammar in separate ways. They also

point out that “when we view language as a tool that you use to create meaning, then it is more appropriate to look at the different sub-systems, such as words, grammar, and sounds as a coherent whole under the notion of discourse” (p. 98).

6.4. Suggestions for Teaching Multiword Items

Research suggests that one of the possible alternative ways for learners to acquire both prepositional verbs and all MWIs is through *direct instruction* (Gardner & Davies, 2007; Hong et al., 2011; Cobb, 2003). As for the acquisition of these items, Cobb (2003) states that “even advanced learners are unlikely to discover very quickly on their own all of the relevant features of a second language that make it native-like” (p. 418-419). That’s why, one possible instructional approach may be to focus on those problematic PVs across levels and to address these forms directly through instruction. Another alternative way would be to highlight and indicate specific similarities as well as differences between prepositions used with certain PVs in English and their equivalents in the students’ L1.

Offering a remedy in pedagogical terms for situations such as L1 transfer, what students would benefit most may be *form-focused instruction* according to Laufer and Waldman (2011) including methods such as; eliciting the collocations, completing collocations from memory, matching the words together, selecting the missing word to teach collocations and MWIs. Raising learners’awareness to the areas they have difficulty in, communicative and task-based teaching should be supplemented by form-focused instruction, either preplanned Focus on Form or Focus on Forms, separating the target items and providing learners practice these in an authentic and communicative context.

Noticing, that is noting, observing or paying special attention to a particular language item, is generally a requirement for learning (Schmidt 1990, Skehan, 2001).

Collocations and MWIs would be taught by eliciting the collocations, matching the words together, selecting the missing word, and completing collocations from memory according to the degree of difficulty, from the most to the least difficult (Hong et al., 2011; Laufer and Waldman, 2011).

In lexical priming context, one of the jobs of teachers should be priming the learners’ vocabulary appropriately. Asserting whether the vocabulary of the discipline has been applied accordingly (primed appropriately) is dependent on the examination system.

Language teaching materials and language can facilitate priming by several ways such as usage notes, drilling exercises, texts or tapes with repeated instances of a word

sequence, collocational observations and illustrations or just drawing a class's attention to a feature. Language learners should be exposed to authentic data wherever and whenever possible, and the data should both reinforce existing priming (i.e. should be in line with previously primed element) and lead new primings to occur. Once some important effect or emotions are created during these primings, it will facilitate priming in an enormous degree. As producing is one of the main key reinforcing elements, the learner should speak and write as much and often as possible (Hoey, 2005).

Flowerdew and Mahlberg (2009) summarize that language is a social phenomenon and meaning may be considered as use of which patterns become visible in corpora and meaning and form are associated with corpus evidence. Corder (1967), as with error analysis in general, the results of this study are significant in ways such as serving a pedagogic purpose by showing what learners have not mastered yet, serving a research purpose by providing evidence about how languages are learned, and serving a learning purpose by making learners discover the rules of the target language (i.e. by obtaining feedback on their errors) (Flowerdew, 2004).

Liu and Jiang (2009) suggests quiet practical ideas as well as studies they have carried out with their own learners. One of the effective activities that would help students to make them check their work on lexicogrammatical exercises against corpus data. In the textbook used at the Chinese university, filling in blanks with appropriate verb particles and selecting the right synonyms exercises were many however, students commented very negatively on these exercises as they could not check their own accuracy.

However, checking corpus data, they are able to check their own accuracy easily. In addition to this, in the composition writing at the south-central U.S. university, the instructors marked their students' lexicogrammatical errors and then instructed students to check these errors themselves checking how the given lexicogrammtical items were used in the BNC. This way, students also could correct their own errors.

As in the studies mentioned above, once learners are exposed to authentic materials as well as tools that can be used in and out of the classroom where they actively take part in, especially correcting their own mistakes and comparing their language productions with existing available corpus tools, they will raise their awareness about certain lexical items such as multiword items. They will notice the similarities as well as differences between their mother tongue and English language. Lexical priming will occur as they actively keep on diving into the data. Additionally, it is not daunting and overwhelming for the teachers to use any of these corpus-based activities in their

classrooms. An example classroom activity, which is prepared by the researcher, is provided in the appendices with the screenshots of a corpus-based activity for teaching prepositional verbs at high school level learners.

Last but not least, multiword items are neither mentioned to the learners in Turkish educational settings in any of English lessons or courses nor paid any adequate attention. In most cases, only the words are taught directly as a word list to be memorized. Learners are supposed to memorise the verb “go” instead of “go to” in this context. Even the coursebooks provided by Ministry of National Education in Turkey do not indicate verbs with their prepositions as can be clearly seen Kutlu’s (2014) study. Learners are not exposed to any authentic language, either. That is why, learners are expected to make errors related with multiword items. However, activities with available and practical tools and materials will provide learners with an invaluable opportunity to prime and internalize multiword items.

Finally, this study is limited with Cambridge exam data across all the profiency levels according to CEFR. The conclusions could be supported with grammatical judgement tasks, translation and elicited imitation tasks for further research.

REFERENCES

Adolphs, S. (2008). *Corpus and context: Investigating pragmatic functions in spoken discourse*. John Benjamins Publishing Company. Amsterdam / Philadelphia.

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. John Benjamins Publishing.

Anderson, W. J. (2006). *The phraseology of administrative French, a corpus-based study*. Amsterdam-New York: Editions Rodopi B.V.

Allerton, D.J., Tschichold, C., Wieser, J. (Eds.), (2004). *Linguistics, Language Learning and Language Teaching.* Schwabe, Basel, Switzerland.

Alonso, M. R. (2002). *The Role of Transfer in Second Language Acquisition*. *Servicio de Publicacións.* Vigo: Universidade.

Altenberg, B. (1998). "On the phraseology of spoken English: the evidence of recurrent word combinations". In Cowie, A.P. (ed.). Phraseology. Theory, Analysis, and Applications. *Oxford: Oxford University Press*, 101-122.

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of *make* in native and non-native student writing. In Chambers, A. and Rostand, S. (eds.) *Le Corpus Chambers-Rostand de Français Journalistique*. (Oxford Text Archive) Oxford: University of Oxford.

Aston, G. (2000). *Corpora and Language Teaching.* In: Burnard/McEnery.

Aston, G., Bernardini, S., Steward, D. (2004) (eds.). Corpora and language learners. *Language*, *83*(4).

Atkins, S., & Clear, J. (1992).Corpus design criteria. *Literary and Linguistic Computing*, 7(1), 1–16.

Baldwin, T. (2005). Looking for prepositional verbs in corpus data. *Proceedings of the 2005 Meeting of the Association for Computational Linguistics*, 115-126.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of Syntactic and Morphological Accuracy by Advanced Language Learners. Studies in Second Language Acquisition, 11, 17-34.

Barlow, M. (1996). Corpora for theory and practice. In: *International Journal of Corpus Linguistics 1(1)*, 1-37.

Bauer, L. (1983). *English Word-formation*. Cambridge: CUP.

Behrens, H. (2008). Core Morphology in Child Directed Speech: Crosslinguistic Corpus Analyses of Noun Plurals. In *Corpora in Language Acquisition Research* (pp. 25-

60). Amsterdam / Philadelphia: John Benjamins Publishing Company.

Bernardini, S. (2002). *Exploring New Directions for Discovery Learning.* In: Kettemann/Marko, 165-182.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In *J. Sinclair (Ed.), How to use corpora in language teaching.* Amsterdam: Benjamins, 15–36.

Biber, D., S. Conrad & R. Reppen (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan. (1999). *Longman Grammar of Spoken and Written English.* New York: Longman.

Bolinger, D. (1977). Transitivity and spatiality: The passive of prepositional verbs. In: Adam Makkai, Valerie Beckker-Makkai, Luigi Heilmann (eds.), *Linguistics at the Crossroads.* Padova/Lake Bluff/Ill.: Liviana Editrice/Jupiter Press, 57-78.

Brown R., C. Fraser & U. Bellugi. (1964). "Explorations in grammar evaluation." *The Acquisition of Language: Monographs of the Society for Research in Child Development 29*, ed. by U. Bellugi & R. Brown, 79-92.

Brown, R. (1973). *A first language: The early stages*. London: George Allen & Unwin.

Cambridge International Dictionary of English. (1995). Cambridge: CUP.

Can, C. (2017). A Learner Corpus-Based Study on Verb Errors of Turkish EFL Learners. *Journal of Education and Training Studies*, *5*(9), 167.

Chambers, A. (2005). Integrating Corpus Consultation in Language Studies. *In: Language Learning and Technology 9(2)*, 111-125.

Chomsky, N. (1965). *Aspects of the theory of syntax.* The Mit Press, Institute of Technology Cambridge, Massachusetts

Claridge, C. (2000). Multi-word Verbs in Early Modern English: A Corpus-based Study. *Language and Computer 32*, 315.

Cobb, T. (1997). Is There Any Measurable Learning from Hands-on Concordancing? In: System: *An International Journal of Educational Technology and Applied Linguistics 25(3)*, 301-315.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review*, 59(3), 393-424.

Collins COBUILD English Language Dictionary. 1999. London: HarperCollins.

Corder, S. P. (1967). *The significance of learner's errors.* [Washington, D.C.]: ERIC Clearinghouse.

Corder, S. P. (1973). *Introducing applied linguistics.* Harmondsworth [Eng.]: Baltimore: Penguin Education.

Corder, S.P. (1974). Error Analysis, In Allen, J.L.P. and Corder, S.P. (1974). *Techniques in Applied Linguistics.* Oxford: Oxford University Press.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment.* Cambridge, U.K: Press Syndicate of the University of Cambridge.

Cowan, R., Choi, H. E., & Kim, D. H. (2003). Four questions for error diagnosis and correction in CALL. *CALICO Journal,* 20(3), 451-463.

Cresswell, A. (2007). Getting to 'Know' Connectors? Evaluating Data-driven Learning in a Writing Skills Course. In: *Hidalgo/Quereda/Santana 2007*, 267-288.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. Elsevier, 26(2), 163-174.

Darus, S. (2009). *Error analysis of the written English essays of secondary school students in Malaysia: A case study*. Eur. J. Soc. Sci., 8(3), 483-495.

Diaz-Negrillo, A. Ballier, N. and Thompson, P. (2013) (eds.), *Automatic treatment and analysis of learner corpus data*, 151–168. Amsterdam: John Benjamins.

Eckman, F. (1977). On the explanation of some typological facts about raising. In: F. Eckman (ed.) *Current Themes in Linguistics*. New York: Halsted Press, 195-214.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, (27),* 315-330.

Eckman, F. (1996). A functional-typological approach to second language acquisition theory. In: W. Ritchie and T. Bhatia (eds). *Handbook of Second Language Acquisition*, 195-211. San Diego: Academic Press.

Eckman, F., Moravcsik, E. & Wirth, J. (1989). Implicational universals and interrogative structures in the interlanguage of ESL learners. *Language Learning, (39),* 173-205.

Edmondson, W. (1999). *Twelve Lectures on Second Language Acquisition.* Tubingen: Gunter NarrVerlag.

Ellis, R., (1994). The Study of Second Language Acquisition. Oxford University Press, Oxford.

Ellis, R. (2008). *The study of second language acquisition*. Second Edition. OUP, 5.

Faerch, C. & Gabriele, C. (1986). Cognitive dimensions of language transfer. In: Eric Kellerman Michael Scharwood Smith (eds). *Crosslinguistic Influence in Second*

Language Acquisition, 49–65. New York: Pergamon.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy, a critical evaluation. *International Journal of Corpus Linguistics*, *14*(3), 393-417.

Flowerdew, J., & Mahlberg, M. (2009). *Lexical cohesion and corpus linguistics.* John Benjamins Publishing, Amsterdam / New York.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly,* 41, 339-359.

Gaskell, D. & Cobb, T. (2004). Can Learners Use Concordance Feedback for Writing Errors? In: System: *An International Journal of Educational Technology and Applied Linguistics 32(3)*, 301-319.

Gavioli, L., & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT Journal*, *55*(3), 238–246.

Gilquin, G. (2007). To err is not all. What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift für Anglistik und Amerikanistik*, *55*(3), 273-291.

Gilquin, G., & Granger, S. (2015). From design to collection of learner corpora. *The Cambridge Handbook of Learner Corpus Research,* 3(1), 9-34.

Glisan, E. and V. Drescher. (1993). "Textbook grammar: Does it reflect native speaker speech?" *The Modern Language Journal,* 77 (1), 23–33.

Goodale, M. (1995). *Collins COBUILD Concordance Samplers 4: Tenses.* London: HarperCollins.

Goyvaerts, D. L. (1968). Towards a Theory of the Expanded Form. *La Linguistique 2*, 118-124.

Götz, D., & Herbst, T. (1989). Language description and language teaching: The London School and its latest grammar. In: *Die Neueren Sprachen*, *88*, 220-235.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. Languages in Contrast: Text-based cross-linguistic studies, Lund Studies in English 88, *Lund University Press*, 37– 51.

Granger, S. & S. Tyson (1996). Connector Usage in Native and Non-native English Essay Writing. *World Englishes 15(1)*, 17-27.

Granger, S. (Ed.). (1998a). *Learner English on Computer.* London: Longman.

Granger, S. (Ed.). (1998b). The computer learner corpus: a versatile new source of data

for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3–18). London: Longman.

Granger, S. (2002). A bird's eye view of learner corpus research. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Benjamins: Atlanta and Amsterdam, 3–33.

Granger, S., Dagneaux, E., and Meunier, F. (2002). *The International Corpus of Learner English: Handbook and CD-ROM.* Louvain-la-Neuve: Presses universitaires de Louvain.

Granger, S., Hung, J., & Petch-Tyson, S. (2002). Computer learner Corpora, Second Language Acquisition and Foreign Language Learning. *Language Learning & Language Teaching*, 258. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Granger, S. (2003). The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly, 37(3).*

Gray, B. E. (2005). Error-specific Concordancing for Intermediate ESL/EFL Writers.

Hammarberg, B. (1974). The insufficiency of error analysis. *IRAL: International Review of Applied Linguistics in Language Teaching*, *12(1–4)*, 185–192.

Hammarberg, B. (1999). *Manual of the ASU Corpus – A longitudinal text corpus of adult learner Swedish with a corresponding part from native Swedes.* Stockholm University, Department of Linguistics.

Han, Z. (2004). Fossilization: five central issues. *International Journal of Applied Linguistics*.

Harley, B., (1980). lnterlanguage units and their relations. *Interlanguage Studies Bulletin* 5, 3-30.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning parsers and pedagogues*. New York: Routledge.

Hong, A. L., Rahim, H. A., Tan, K. H., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *3L; Language, Linguistics and Literature, The Southeast Asian Journal of English Language Studies.*, *17*(special issue), 31-44.

Hoey, M. (1997). *From concordance to text structure: New uses for computer corpora. In Practical Applications in Language Corpora.* (eds), 2–23. Łódź: Łódź University Press.

Hoey, M. (2004). “Lexical priming and the properties of text”. In Aston, G., Bernardini, S., Steward, D. (2004) (eds.). Corpora and language learners. *Language*, *83*(4).

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Hoey, M., Mahlberg, M., Stubbs, M., & Teubert, W. (2007). *Text, Discourse and Corpora: Theory and Analysis* (Vol. 28).

Hopper, P. J. & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56* (2), 251-299.

Housen, A. (2002) The development of Tense-Aspect in English as second language and the variable influence of inherent aspect. *The L2 Acquisition of Tense-Aspect Morphology,* Rafael Salaberry, and Yasuhiro Shirai (eds.), 155-197. Amsterdam: Benjamins.

Huddleston, R. (1984). *Introduction to the Grammar of English*. Cambridge: CUP.

Hunston, S. & Francis, G. (2000). *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

Hymes, D. (1992). The Concept of Communicative Competence Revisited. In: *Pütz Martin (ed.), Thirty Years of Linguistic Evolution.* Studies in Honour of Rene Dirven on the Occasion of his Sixtieth Birthday. Amsterdam: John Benjamins, 31-57.

James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis.* London: Longman.

Jimenez Catalan, R. M. (1996). Frequency and variability in errors in the use of English prepositions. *Miscelánea: A Journal of English and American Studies*, 17, 171-187.

Johns, T. (1986). Microconcord: A Language-learner’s Research Tool. In: *System 14(2),* 151-162.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal, 4*, 1-16.

Johns, T. (1994). From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning. In: *Odlin, Terence (ed.), Perspectives on Pedagogical Grammar*. Cambridge: *Cambridge University Press*, 27-45.

Johns, T. (1997). Contexts: The Background, Development and Trialling of a Concordance- based CALL Program. In: *Wichmann et al.,* 100-115. Johnson, M. (2004). *A Philosophy of Second Language Acquisition*. London: Yale University.

Juozulynas, V. (1994). Errors in the Compositions of Second-Year German Students: An

Empirical Study for Parser-Based ICALI. *CALICO Journal,* 12(1), 5-17.

Kazemian, B., & Hashemi, S. (2014). A contrastive linguistic analysis of inflectional bound morphemes of English, Azerbaijani and Persian languages: A comparative study. *Journal of Education & Human Development, 3(1),* 593-614.

Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition*, *2(1),* 37–57.

Kellerman, E. (1983). Now you see it, now you don't. In Susan M. Gass & Larry Selinker (Eds.), *Language Transfer in Language Learning,* 112–134. Rowley, MA: Newbury House.

Kellerman, E. (1985). If at first you do succeed... In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 345-353). Rowley, Mass.: Newbury House.

Kennedy, G. (1992). Preferred ways of putting things with implications for language teaching. In J. Svartvik (Ed.), Directions in Corpus Linguistics. *Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991* (pp. 334–373). Berlin: Mouton de Gruyter.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics.* London: Longman.

Keshavarz, M. H. (2003). *Contrastive analysis and error analysis.* Tehran: Rahnama Publications.

Keshavarz, M. H. (2012). Sources of Errors. In *Contrastive Analysis and Error Analysis* (pp. 117-128). Rahnama Press.

Kettemann, B. & Marko, G. (2004). Can the L in TaLC stand for literature? In Aston, G., Bernardini, S., Steward, D. (2004) (eds.). Corpora and language learners. *Language*, *83*(4).

Kochmar, E. (2016). *Error detection in content word combinations.* University of Cambridge.

Kjellmer, G. (1984). Some Thoughts on Collocational Distinctiveness. In: Aarts, Jan/Meijs, Willem (eds.), Corpus Linguistics. *Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Rodopi, 163-171.

Kornfilt, J. (1997). *Turkish*. London: Routledge.

Krashen, S. D. (1981) *Second Language Acquisition and Second Language Learning.* Oxford: Pergamon.

Kutlu, Ö. (2014). *An investigation of verb islands in Turkish and English language coursebooks used at primary school level.* (Doctoral dissertation, Çukurova

University, Adana, Turkey).

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning,* 61(2), 647-672.

Leacock, M. Chodorow, M. Gamon, and J. Tetreault. (2014). Automated grammatical error detection for language learners. *Morgan & Claypool Publishers*, (2).

Leech, G. N. (1991). The state of the art in Corpus Linguistics. K. Aijmer & B. Altenberg (eds), In *English Corpus Linguistics. Studies in Honor of Jan Svartvik,* 8–29. London: Longman.

Leech, G. N. (1992). Corpora and theories of linguistic performance. In Directions in Corpus Linguistics. *Proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, J. Svartvik (ed.), 105–122. Berlin: Mouton de Gruyter.

Leech, G. (1997). Teaching and Language Corpora: A Convergence. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–23). London: Longman.

Leech, G. N. (1998). Learner corpora: what they are and what can be done with them. In: Granger, S. (Ed.). *Learner English on Computer.* Addison Wesley Longman, London.

Leech, G. N. (2002). *Corpora. In The Linguistics Encyclopedia.* K. Malmkjær (ed.), 84-93. London: Routledge.

Levin, B. (1993). English verb classes and alternations: a preliminary investigation. *University of Chicago Press.*

Lewis, M. (1993) *The Lexical Approach*. Hove: LTP.

Lewis, M. (2000). *Teaching Collocation. Further Developments in the Lexical Approach.* Hove: Language Teaching Publications.

Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, *93*.

Lorenz, G. (1999). *Adjective Intensification – Learners vs Native Speakers. A Corpus Study of Argumentative Writing.* Amsterdam & Atlanta: Rodopi.

Lüdeling, A., & Kytö, M. (2008). An International Handbook. *Handbooks of Linguistics and Communication Science,* (1).

Lüdeling, A., & Kytö, M. (2009). *Corpus Linguistics An International Handbook*. Handbooks of Linguistics and Communication Science.

Macaro, E. (2010). A Compendium of Key Concepts in Second Language Acquisition. In

Continuum Companion to Second Language Acquisition (pp. 29-101). London: Continuum International Publishing Group.

McEnery, T. and A. Wilson. (1996). *Corpus Linguistics.* Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based Language Studies.* London: Routledge.

McEnery, T., & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (Vol. 2, pp. 364-380). London & New York: Routledge.

Mahlberg, M. (2005). *English General Nouns: A corpus theoretical approach*. John Benjamins Publishing: Amsterdam

Meunier, F. (1998). Computer Tools for the Analysis of Learner Corpora. In*: Granger, S. (ed.), Learner English on Computer.* London/New York: Addison Wesley Longman, 19-37.

Meunier, F. (2002). The Pedagogical Value of Native and Learner Corpora in EFL Grammar Teaching. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (Vol. 2, pp. 364-380). London & New York: Routledge.

Meunier, F. & Gouverneur, C. (2007). The Treatment of Phraseology in ELT Textbooks. In: *Hidalgo/Quereda/Santana 2007*, 119-140.

Meyer, C. F. (2002). English Corpus Linguistics: An Introduction. *2002*, *41*(1), 765.

Mitchell, T. F. (1958), 'Syntagmatic relations in linguistic analysis', in Claridge, C. (2000). Multi-word Verbs in Early Modern English: A Corpus-based Study. *Language and Computer 32*, 315.

Mitchell, R., & Myles, F. (2004). *Second Language Learning Theories* (eds.). London: Hodder Arnold.

Mitton, R. (1996). *English spelling and the computer*. Harlow, Essex: Longman Group.

Morris, L. and Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System 32*: 75–87.

Mukherjee, J., & Rohrbach, J. (2006). Rethinking Applied Corpus Linguistics from a Language-pedagogical Perspective: New Departures in Learner Corpus Research. In: *Kettemann, Bernhard/Marko, Georg (eds.). Planing, Gluing and Painting Corpora*, 205- 232.

Muller-Hartmann, A. & Schocker-von Ditfurth, M. (2004). *Introduction to English Language Teaching*. In Flowerdew, J. & Mahlberg, M. (2009). Lexical Cohesion and Corpus Linguistics. Stuttgart: Klett.

Nattinger, J. R. (1980). A Lexical Phrase Grammar for ESL. In: *TESOL Quarterly 14(3)*, 337-344.

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. John Benjamins Company. Amsterdam / Philadelphia.

Ndiaye, M., & Faltin, A. V. (2003). A Spell Checker Tailored to Language Learners. *Computer Assisted Language Learning,* 16(2-3), 213-232.

Nicholls, D. (2003). The Cambridge Learner Corpus: Error Coding and Analysis for Lexicography and ELT. In: Archer, D./Rayson, P./Wilson, A./McEnery, T. (eds.). *Proceedings of the Corpus Linguistics 2003 Conference.* UCREL, Lancaster University, 572-581.

O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching.* Cambridge University Press.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Rayson, P. (2003) *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. (Lancaster University).

Rayson, P. & Archer, D. (2008). *Key domain analysis: Mining text in the humanities and social sciences*. Paper presented at the workshop on "Text mining and the social sciences," 4th International Conference on e-Social Science, University of Manchester.

Rayson, P. & Baron, A. (2011). *Automatic error tagging of spelling mistakes in learner corpora.* In A Taste for Corpora. In Honour of Sylviane Granger, F. Meunier, S. De Cock, G. Gilquin & M. Paquot (eds), 109–126. Amsterdam: John Benjamins.

Reppen, R. Fitzmaurice, S. M. and Biber, D. (2002). Using Corpora to Explore Linguistic Variation, 275.

Richards, J., C. Platt, J., & Weber, H. (1989). *Longman dictionary of applied linguistics.* London: Longman.

Romaine, S. (2003). Variation. In Doughty, C.J. and Long, M.H. (eds), *The handbook of second language acquisition.* Oxford: Blackwell Publishing, 409-35.

Römer, U. (2004b). Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In: *Aston/Bernardini/Stewart 2004*, 151-168.

Römer, U. (2008). Corpus research and practice:What help do teachers need and what can we offer? In*:* Aijmer, Karin (ed.), *Teaching and Corpora. Amsterdam: John Benjamins.*

Rundell, M., & Stock, P. (1992). The corpus revolution. *English Today*, *8*(4).

Sardinha, B. (2011). *Metaphor and corpus linguistics*. RBLA, Belo Horizonte.

Schachter, J. (1974). An error in error analysis. Language Learning, 24, 205-214.

Schmidt, R. (1990). "The role of consciousness in second language learning". Applied Linguistics 11(2):129–158.

Scholfield, P., (1995). Quantifying Language. *Multilingual Matters*, Clevedon.

Scott, M. (1999). *WordSmith Tools Help Manual. Version 3.0.* Mike Scott and Oxford University Press.

Scott, M. (2000). 'Focussing on the text and its key words.' In L. Burnard and T. McEnery (eds). *Rethinking Language Pedagogy from a Corpus Perspective.* Frankfurt: Peter Lang, pp. 103–22.

Scott, M. (2001). 'Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs.' In M. Ghadessy, A. Henry and R. L. Roseberry (eds) *Small Corpus Studies and ELT: Theory and Practice.* (pp. 47– 67). Amsterdam: John Benjamins.

Scott, M. & Thompson, G. (eds.) (2001). *Patterns of Text: In Honour of Michael Hoey*. Amsterdam: John Benjamins.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, (*20)*, 213–223.

Seidlhofer, B. (2002). Pedagogy and Local Learner Corpora: Working with Learning-driven Data. In: *Granger/Hung/Petch-Tyson 2002*, 213-234.

Selinker, L. (1972). Interlanguage. *IRAL - International Review of Applied Linguistics in Language Teaching*, *10*(1-4).

Sinclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1991a). *Shared knowledge. In J. Alatis (Ed.), Georgetown University Round Table in Language and Linguistics. Linguistics and Language Pedagogy: The State of the Art (pp. 489–500).* Washington DC: Georgetown University.

Sinclair, J. (1991b). *Corpus, Concordance.* Collocation. Oxford: Oxford University Press.

Sinclair, J. (1997). Corpus Evidence in Language Description. In: *Wichmann et al. 1997*, 27-39.

Sinclair, J. (2001). (eds.) *Collins COBUILD English Dictionary for Advanced Learners*. London: HarperCollins.

Sinclair, J. (2004). How To Use Corpora in Language Teaching. *Studies in Corpus Linguistics*, *12*, 317.

Skehan, P. (2001). *"The role of a focus on form during task-based instruction".* In Trabajos en lingüistica aplicada. C. Muñoz, M.L Celaya, M. Fernández-Villanueva, T. Navés, O. Strunk and E. Tragant (eds), 11–24. Barcelona: Univerbook SL.

Stevens, V. (1991). Concordance-based Vocabulary Exercises: A Viable Alternative to Gap- fillers. In: Johns, Tim/King, Philip (eds.), *Classroom Concordancing (ELR Journal 4),* 47-63.

Stubbs, M. and Gerbig, A. (1993). Human and inhuman geography: On the computer-assisted analysis of long texts. In M. Hoey (ed.), *Data, description, discourse: Papers on the English Language in honour of John McH. Sinclair on his sixtieth birthday,* 64–85. London: HarperCollins.

Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems*. Cambridge University Press.

Tetreault, J. & M. Chodorow (2008). The Ups and Downs of Preposition Error Detection in ESL Writing. *In Proceedings of the 22nd International Conference on Computational Linguistics (COLING-08),* 865–872. Manchester, UK: Association for Computational Linguistics.

Van Els, T., Bongaerts, T., Extra, G., van Os, C. Janssen-van Dieten, A.M. (1984). *Applied Linguistics and the Learning and Teaching of Foreign Languages.* Edward Arnold, London.

Viana, V., Zyngier, S., & Barnbrook, G. (2011). Principles and Applications of Corpus Linguistics. In *Perspectives on Corpus Linguistics* (pp. 155-169).

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Cambridge: Harvard University Press.

Wilcoxon, E. M. (2014). *A corpus based study of the use of prepositional verbs in second language emergent academic writing.* Master of Arts, The University of Texas, El Paso.

Willis, D. J. (1989). *Collins COBUILD English Course*. London: HarperCollins.

Willis, D. (2003) *Rules, Patterns and Words: Grammar and Lexis in English Language Teaching*. Cambridge: Cambridge University Press.

Winford, D. (2003). *An Introduction to Contact Linguistics.* Maiden: Blackwell Publishing, 217.

Wulff, S. (2017). What learner corpus research can contribute to multilingualism research. *International Journal of Bilingualism*, *21*(6).

Zhang, R. (2013). A Corpus-based error analysis of students' writing in graded teaching classes. *Journal of Convergence Information Technology, 8(10),* 551-557.

Ziahosseiny, S. M. (1999). *A contrastive analysis of Persian and English and error analysis.* Tehran: Nashr-e Vira.

CURRICULUM VITAE

PERSONAL INFORMATION

Name & Surname **:** Sibel Aybek
Date of Birth : 24.07.1993
E-mail **:** saybek@cu.edu.tr

EDUCATIONAL BACKGROUND

2016 – 2018 **:** MA, Çukurova University, Institute of Social Sciences, English Language Teaching Department

2011 – 2016 **:** BA, Çukurova University, Faculty of Education, English Language Teaching Department

WORK EXPERIENCE

2018 - **:** Research Assistant, Çukurova University, English Language Teaching Department, Adana