

**REPUBLIC OF TURKEY
CUKUROVA UNIVERSITY
THE INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LANGUAGE TEACHING**

**THE TURKISH EFL LEARNERS' USE OF DETERMINERS: A LEARNER
CORPUS-BASED STUDY**

Abdulkadir ABDULRAHIM

MASTER OF ARTS

ADANA / 2020

**REPUBLIC OF TURKEY
CUKUROVA UNIVERSITY
THE INSTITUTE OF SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LANGUAGE TEACHING**

**THE TURKISH EFL LEARNERS' USE OF DETERMINERS: A LEARNER
CORPUS-BASED STUDY**

Abdulkadir ABDULRAHIM

**Supervisor : Prof. Dr. Cem CAN
Member of Examining Committee : Prof. Dr. Hasan BEDİR
Member of Examining Committee : Assoc. Prof. Dr. Pınar BABANOGLU**

MASTER OF ARTS

ADANA / 2020

To the Directorship of the Institute of Social Sciences, Çukurova University

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

Chair: Prof. Dr. Cem CAN
(Supervisor)

Member of Examining Committee: Assoc. prof. Dr. Hasan BEDİR

Member of Examining Committee: Assist. Prof. Dr. Pinar BABANOGLU

I certify that this thesis confirms to the formal standards of the Institute of Social Sciences.

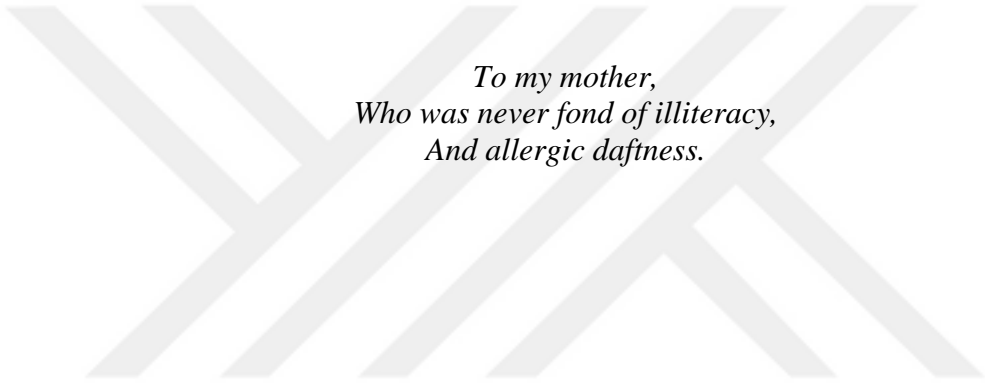
.... /.... /2020

Prof. Dr. Serap ÇABUK
Director of the Institute

Note: The uncited usage of the reports, charts, figures, photographs in this thesis, whether original or quoted from other sources, is subject to the law of Works of Art and Thought No: 5846.

Not:

Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafinin kaynak gösterilmeden kullanımı, 5846 Sayılı Fikir ve Sanat Eserleri Kanunu'ndaki hükümlere tabidir.



*To my mother,
Who was never fond of illiteracy,
And allergic daftness.*

ETİK BEYANI

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

- 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 çalışmada 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.

.... / / 20...

Abdulkadir ABDULRAHİM

ÖZET

YABANCI DİL OLARAK İNGİLİZCE ÖĞRENEN TÜRK ÖĞRENCİLERİN İNGİLİZCE TANIMLIKLARI KULLANIMI: ÖĞRENCİ DERLEMİNE DAYALI BİR ÇALIŞMA

Abdulkadir ABDULRAHİM

Yüksek Lisans Tezi, İngiliz Dili Anabilim Dalı

Danışman: Prof. Dr. Cem CAN

Aralık 2020, 98 Sayfa

Bu çalışma, yabancı dil olarak İngilizce öğrenen Türk öğrencilerinin belirleyicilerin kullanımını araştırmayı amaçlamaktadır. Çalışma, yabancı dil olarak İngilizce öğrenenlerin belirleyicilerin kullanımını çözümlmek için derleme temelli tekniklerden yararlanarak, Ortak Avrupa Çerçevesi (CEFR) doğrultusunda A1, A2, B1, B2, C1 ve C2 yeterlilik düzeylerindeki. Türk öğrencilerinden toplanan verileri incelemektedir. Bu amaçla kullanılan ve Cambridge Öğrenci Derleminin, Türk öğrenci altderlemi 1993- 2012 yılları arasında Cambridge İngilizce dil sınavlarına giren öğrencilerin yazılı anlatımlarından oluşmaktadır. Ayrıca, bu çalışmada, Türk EFL öğrenenleri alt derlemeleri kullanılan İngilizce belirleyicisinin sıklığının analiz edilmesinde çalışma olarak tanımlayıcı ve nicel yöntemler kullanılmıştır.

İngilizce belirleyiciler, adlara eşlik eden veya başka bir deyişle, genellikle ad tümceciklerinden önce beliren veya bir ad tanımlığı olarak görev üstelenen sözcüklerdir. McGregor (1997), tanımlık teriminin, belirlilik ve belirsizliğin en temel anlatımını sağlayan belirleyici özel alt ulamı olarak kullanıldığını belirtmektedir. Bu çalışmada, Türk EFL öğrencilerin derlemi incelenmiş, İngilizce belirleyicilerinin (tanımlıklar) sıklığı, olası fazlaşa veya az kullanımları Log-likelihood ve Type Token sayasal çözümlme yöntemleri ile belirlenmiştir, C1-C2 yeterlilik seviyelerinin A1-A2 ve B1-B2 CEFR yeterlilik seviyeleri ile karşılaştırılınca İngilizce tanımlıklar fazla kullandığı bulunmuştur.

Ayrıca, derlemede bulunan hata türleri göz önüne alındığında, Missing Determiner (MD), diğer üç belirleyici hata türünde (UD, FD ve RD) kalıcılık olduğu gözlemlenmiştir .

Anahtar Kelimeler: Yabancı Dil Olarak İngilizce Öğrenen Türk Öğrencileri, Derlem dilbilimi, Öğrenci derlemleri, İngilizce belirleyiciler (tanımlıklar).



ABSTRACT**THE TURKISH EFL LEARNERS' USE OF DETERMINERS: A LEARNER
CORPUS-BASED STUDY****Abdulkadir ABDULRAHIM****Master of Arts, English Language Department****Supervisor: Prof. Dr. Cem CAN****December 2020, 98 Pages**

This study intends to find out the Turkish EFL learners' use of determiners. The study will make use of the corpus-based techniques to analyse the use of determiners in the Turkish EFL learner written performance between A1, A2, B1, B2, C1 and C2 Common European Framework (CEFR) levels of proficiency available in the related corpora of the Cambridge Learner Corpora. The participants of the study are Turkish EFL learners from A1-C2 CEFR proficiency level, the sub-corpora consist of data from 1993 – 2012 written in different genres such as argumentative essay, letters, story, advice, application, et cetera. Moreover, in the present study, descriptive and quantitative methods are used as the study dwells on analysing the frequency of the English determiner used and errors made in the Turkish EFL learners' sub-corpus and seeks to obtain reasons behind those errors.

English determiners are words that accompany nouns or, in other words, are nouns modifiers which usually appear before noun phrases to specify or to serve as a reference of a noun. In addition, McGregor (1997), states that the term article is used for the special subcategory of determinatives that provide the most basic expression of definiteness and indefiniteness. In this study, the Turkish EFL learners' corpora were analysed, the use of English determiner (articles) frequency was determined, as the study dwelled on the over-/underuse of English determiners (articles), through the use of Log-likelihood ratio and Type-token ratio, it is found that C1-C2 proficiency levels overused the English determiner (articles) compared to A1-A2 and B1-B2 CEFR proficiency levels.

Moreover, regarding the error typologies found in the corpus, Missing Determiner (MD) is found to be more persistent than the other three determiner error typologies, these are, (UD, FD, and RD). At last, in the overall analysis of the English determiners

(articles), the presents that the use of 'the' is more frequent the corpus. Also, the main reason behind the learners' overuse and overuse of the determiners was the influence of the o the learners' native language.

Furthermore, in the implementation of corpus driving activities in teaching and learning English determiners (articles), it should experiment was carried out, the participants in the study were divided into two groups, and experimental group (corpus-driven techniques) and a control group (traditional method). A pre-test was done before the research process, after four weeks of teaching English determiners (articles) a post-test was conducted. it is realized that corpus-driven activities in teaching English determiners to Turkish EFL learners have a vital role in improving the learners' comprehension and the activities trigger them to produce more sentences to prove their understanding and use the target language beyond the classroom context.

Keywords: Turkish EFL learners, corpus linguistics, learner corpora, English determiners (Articles)

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Prof. Dr Cem CAN, for his perspicacious and illuminating comments and support during my studies. His advice has been priceless in sharpening the content and style of the whole thesis. This thesis has availed enormously from his remarks and advice; any shortcomings in this thesis remain mine.

My thanks go to the diligent, Prof. Dr. Hasan BEDİR, he is absolutely special person, I do not think I will ever forget the first I met him, and Doc. Dr. M. Pınar BABANOĞLU and Dr. Reyhan AGCAM for accepting to be my examination committee members and for their insightful comments at the end of my study.

I am grateful to every member of the English Language Teaching Department at Çukurova University starting from the Head of the Department Prof. Dr Zuhal OKAN to all for their assistance and during the different stages of my studies.

I would very much like to express my humble gratitude to Prof. Dr. Rana YILDIRM, Doc. Dr. Gulden TÜM, Doc. Dr. Nese CABAROGLU, Dr. Tugba SIMSEK and Sibel AYBEK (MA) for their support and friendly smiley faces which never goes away.

I would like to express my deepest gratitude to my family, "the people of Arjali", especially my mother who has always been there for me, I am so thankful for her unconditional love and kindness. And also, to my friends who have never lost faith in me. Ahmed Ali Mali Shuwadi, with whom destiny and search for the light (knowledge) joined us to become more than friends, I appreciate your support and motivation, and Hamidou Camara.

Abdulkadir ABDULRAHIM
Adana / 2020

TABLE OF CONTENTS

ÖZET	iv
ABSTRACT.....	vi
ACKNOWLEDGEMENTS	viii
LIST OF ABBREVIATION	xii
LIST OF FIGURES	xiii
LIST OF TABLES	xiv
LIST OF APPENDICES	xv

CHAPTER I INTRODUCTION

1. Introduction.....	1
1.1. Background of the Study	1
1.2. Statement of the Problem.....	3
1.3. Purpose of the Study	4
1.4. Significance of the Study.....	4

CHAPTER II LITERATURE REVIEW

2. Introduction.....	6
2.1. Corpus (Corpora)	6
2.2. Types of Corpora	6
2.2.1. General Corpora	7
2.2.2. Specialized Corpora	7
2.2.3. Historical Corpora	7
2.2.4. Diachronic Corpora.....	8
2.2.5. Synchronic Corpora	9
2.2.6. Parallel Corpora	9
2.2.7. Comparable Corpora.....	10
2.2.8. Monitor Corpora	10

2.2.9. Learner Corpora	10
2.3. Corpus Linguistics	12
2.4. Corpus Approaches	13
2.4.1. Corpus-based Versus Corpus-driven studies	14
2.5. Corpus and Language Teaching	17
2.6. Contrastive Analysis	20
2.7. Error Analysis	21
2.7.1. Traditional Error Analysis.....	24
2.7.2. Computer-aided Error Analysis.....	24
2.7.3. Interlanguage.....	25
2.8. English Determiners	25
2.8.1. English Articles	27
2.8.1.1. The definite article (the).....	28
2.8.1.1.1. Definiteness and Specificity	28
2.8.1.1.2. Uses of Definite Articles	29
2.8.1.1.2.1. Anaphoric Use.....	30
2.8.1.1.2.2. Associative Anaphoric Use	30
2.8.1.1.2.3. Immediate Situation Use	31
2.8.1.1.2.4. Larger Situation Use.....	31
2.8.1.2. The indefinite article (a/an).....	31
2.9. Turkish Determiners	32
2.9.1. Turkish Indefinite Article.....	32
2.10. English Articles in Language Acquisition Researches	33

CHAPTER III

METHODOLOGY

3. Introduction.....	36
3.1. Research Design	36
3.2. Turkish EFL Learners' Sub-corpora.....	37
3.3. Type-token Ratio	39
3.4. Log-likelihood	40
3.5. Participants of the Experimental and Control Group.....	42
3.6. Procedure for Using Corpus	42

3.7. Duration of the Study.....	44
---------------------------------	----

CHAPTER IV FINDINGS

4. Introduction.....	45
4.1. Overall Findings of the Study	45
4.2. Type-token Ratio Analysis Results.....	46
4.3. Log-likelihood Analysis Findings	47
4.4. Traditional and Corpus-driven Techniques Findings	50

CHAPTER V DISCUSSIONS

5. Introduction.....	55
5.1. Summary of the Study	55
5.1.1. Research Question 1:.....	55
5.1.2. Research Question 2:.....	57
5.1.3. Research Question 3:.....	59
5.1.4. Research Question 4:.....	64
5.1.5. Research Question 5:.....	66

CHAPTER VI CONCLUSION

6. INTRODUCTION	68
6.1. Conclusion	68
6.2. Implications of the Study in English Language Teaching.....	69
6.3. Recommendation for Further Study	70

REFERENCE.....	72
APPENDICES.....	81
CURRICULUM VITAE.....	96

LIST OF ABBREVIATION

ARCHER	: A Representative Corpus of Historical English Registers
CA	: Contrastive Analysis
CCEC	: Collins Co-build English Collocation
CEA	: Computer-aided Error Analysis
CEFR	: Common European Framework
CIA	: Contrastive Interlanguage Analysis
CLC	: Cambridge Learner Corpus
DDL	: Data-driven Learning
EA	: Error Analysis
EFL	: English as a Foreign Language
ELL	: English Language Learning
ELT	: English Language Teaching
ESL	: English as a Second Language
ESP	: English for Special Purpose
FL	: Foreign Language
IH	: Interface Hypothesis
IL	: Interlanguage
L1	: First Language
L2	: Second Language
L2L	: Second Language Learning
LL	: Log-likelihood
MT	: Mother Tongue
NNS	: Non-native Speaker
NP	: Noun Phrase
NS	: Native Speaker
SLA	: Second Language Acquisition
TL	: Target Language

LIST OF FIGURES

Figure 1. Corpus-based and corpus-driven approaches to corpus linguistics.....	17
Figure 2. Contrastive Interlanguage Analysis Adapted from	21
Figure 3. Categories of English Determiners.....	26
Figure 4. Screenshot of Log-likelihood Calculator.....	40
Figure 5. English Determiner (articles) Frequency among the Levels	56
Figure 6. Overall Frequency of English Determiners (articles).....	58



LIST OF TABLES

Table 1. Turkish EFL learners' Sub-corpora Breakdown	38
Table 2. Contingency Table for Log-likelihood Calculation	41
Table 3. Corpus Size and English Determiner Frequency	45
Table 4. Comparative Frequency of English Determiners (articles) in the Corpora	46
Table 5. Log-likelihood Ratio of A1-A2 and B1-B2	48
Table 6. Log-likelihood Ratio of A1-A2 and C1-C2	49
Table 7. Log-likelihood Ratio of B1-B2 and C1-C2.....	50
Table 8. Experimental Group Participants' Test Scores	51
Table 9. Control Group Participants' Test Scores	52
Table 10. Descriptive Results of Pre-test and Post-test	53
Table 11. ANOVA Results	54
Table 12. Frequency in Missing Determiner Error Typology.....	59
Table 13. Frequency in Unnecessary Determiner (UD) Error Typology.....	60
Table 14. Frequency in Incorrect Determiner Form (FD) Error Typology.....	62
Table 15. Frequency in Replace Determiner (RD) Error Typology	63

LIST OF APPENDICES

APPENDIX I: List of English Determiners (articles) Used in A1-A2.....	81
APPENDIX II: List of English Determiners (articles) Used in B1-B2.....	82
APPENDIX III: List of English Determiners (articles) Used in C1-C2.....	83
APPENDIX IV: List of English Determiners (articles) Used Overall	85
APPENDIX V: Pre-test and Post-test Questions.....	87
APPENDIX VI: Lesson Plan 1.....	88
APPENDIX VII: Lesson Plan 2	89
APPENDIX VIII: Fill in Blanks Activity	90
APPENDIX IX: Error Correction	92
APPENDIX X: Picture Description.....	95

CHAPTER I

INTRODUCTION

1. INTRODUCTION

This chapter serves to give insights about the background of the study in which the definition and history or precisely the background of corpus linguistics and corpus studies are mentioned, then the statement of the problem, purpose of the study and at last the significance of the study are discussed at the end of the chapter.

1.1. Background of the Study

Before getting deeper into the subject matter, defining corpora (corpus) or shading knowledge about corpus would be relatively convenient. Corpus linguistics can best be defined as a linguistic methodology which is founded on the use of electronic collections of naturally occurring texts, viz. corpora (Granger, Hung, & Petch-Tyson, 2002), and added that "It is neither a new branch of linguistics nor a new theory of language, but the very nature of the evidence it uses makes it a particularly powerful methodology, one which has the potential to change perspectives on language" (Granger, Hung, & Petch-Tyson, 2002: p. 4).

In the preceding few decades, the interest in corpus studies has developed swiftly. Corpus linguistics could be seen as the fastest growing methodology among the various linguistics disciplines. Gries (2009), also highlighted that Corpus Linguistics is one of the fastest growing methodologies in linguistics, which is also a lively and increasingly multilingual subject. "The most striking development in the linguistics description over the past twenty years has been the use of the computer to collect and analyse vast corpora of actually occurring language data" (Widdowson, 2004).

Acknowledging the fact that the state of corpus linguistics and its components are yet broadly discussed among the linguistic circles, it seems that the controversy related to the status of corpus linguistics as a separate linguistic discipline or a methodology through which diverse linguistic investigations can be made in the world of linguistics and ELT, in general, is likely to continue. However, corpus linguistics presents the

profound changes in the way that we study, teach, and learn languages all over the world due to its enormous potential to present entirely authentic, genuine, qualitative, and quantitative findings related to the nature of language.

In addition, “It is neither a new branch of linguistics nor a new theory of language, but the very nature of the evidence it uses makes it a particularly powerful methodology, one which has the potential to change perspectives on language.” Corpus linguistics is the study of language data on a large scale – the computer-aided analysis of very extensive collections of transcribed utterances or written texts (McEnery & Hardie, 2012).

Corpus linguistics has been one of the more exciting methodological developments in linguistics since the Chomskyan revolution of the 1950s. It reflects changing attitudes among many linguists as to what constitutes an adequate “empirical” study of language, and it has drawn upon recent developments in technology to make feasible the kinds of empirical analyses of language that corpus linguists wish to undertake. As McEnery & Wilson (2001) states, corpus linguistics is a methodology which could be defined as a study of natural language on examples of real-life language use by means of a corpus, identified as a body of text representative of a particular variety of language which is stored on a computer.

According to McEnery & Wilson (2001), “ESP (English for Special Purpose) as a specialized area within language teaching and learning claims that corpora can be used to provide several different kinds of domain-specific material for language learning, as well as quantitative accounts of vocabulary and usage which meet students’ specific needs of a particular domain more directly than those taken from more general language corpora” (p. 121). Corpus linguistics, through the use of learner corpora, plays a vital role in linguistics, language teaching and learning, and second language acquisition, by providing a linear way of developing materials such as dictionary, textbooks, etc.

Learner corpora serve as to provide the real life, as in, authentic language use of the language learners in order to broaden linguistics studies and to develop corpus-based approaches in which language teachers can use as a convenient way of language teaching. “Learner corpora are a fairly recent phenomenon, as they only started to emerge in the 1990s, more than 30 years after native speaker corpora began to be compiled” (Nesselhauf, 2004, pg. 40).

Looking into the learner corpora helps us to determine the learners’ performance towards the target language; it also provides the natural sample of the language. Moreover, through looking into the authentic learner produced data we have the

opportunity to see that language progress and/or the difficulties faced by the learners in the language learning process, and the reason behind the failure. Briefly, there is need for more research in the landscape of learner corpora in terms of pedagogical and academic needs, in order to come up with a suitable room for discussing linguistics items in the interlanguage data provided by the language learners. Therefore, the present research serves to find out the issues pertaining to the use of determiners by the Turkish EFL learners in order to contribute to the educational and linguistic fields.

1.2. Statement of the Problem

Language education, learning and language research are still amongst the trending topics in the academia. The present study deals with the acquisition/learning of English as a foreign/second language, it focuses on the use of English determiners, which is amongst the most hardly acquired as of English grammar. Master (1987) states that “unlike the acquisition of English article by L1 learners, the English article system is one of the most difficult of aspects of English grammar for ELLs [English Language Learners] and one of the last to be fully acquired.” As also pointed by Cook (1994) “the most persistent recurrent problems in terms of slips and minor errors amongst all groups of the most sophisticated users of English as a second language involve confusion with the words *the*, *a* (*an*) and their proper necessary non-use.” Considering the above comments, it could be presumed that determiners are problematic in language learning, or language acquisition, especially in the English language. Thus, in the present study we would like to look into the Turkish ELF learners' written texts to find out the problems and the reason why such errors occur. Since there is an absence of the definite article in Turkish language, it presumed that the Turkish EFL learners face some difficulties in the process of learning or acquiring English articles. “Once L2 errors are analysed, the nature of problems and difficulties encountered by language learners will be identified. Identifying such difficulties can therefore help EFL/ESL teachers pinpoint their students' difficulties in this respect and hence revise their teaching methods and learning materials accordingly” (Al- Khresheh, 2011).

The present study aims to analyse the errors pertaining the use of determiners by Turkish EFL learners across 6 different proficiency levels, these are A1, A2, B1, B2, C1 and C2, in accordance with the Common European Framework of reference for language.

1.3. Purpose of the Study

The aim of the study is to find out the problems in the use of determiners (articles) in the written data of the Turkish EFL Learners between A1, A2, B1, B2, C1 and C2 CEFR proficiency level as shown in the Cambridge Learner Corpora.

The research dwells on answering the following research question.

1. What are the most common errors pertaining to the use of English Determiners along with the proficiency range A1 – C2 according to CEFR?
2. Which determiners are over-/under used by the Turkish EFL learners in various levels of proficiency range?
3. What are the persistent errors of Turkish EFL learners in their use of English determiners along with their proficiency levels?
4. Do the errors relate with the use of determiners stem from Turkish L1 or imply an interlanguage property?
5. Do corpus-driven techniques assist the learners in their acquisition of determiners?

1.4. Significance of the Study

This study aims at identifying the errors Turkish EFL learners make and find out the elements behind those errors related with the use of determiners, Ellis (1994) highlighted that collecting well-defined samples of learner language so that clear statements can be made concerning what sorts of errors the learners produce and under what condition.

In this case, investigating collected data of the students provides evidence of discovering the Turkish EFL learners' progress toward the target language learning. Looking into the learner corpora helps us to determine the learners' performance toward the target language, it also provides the natural sample of the learner language. Moreover, due to looking into the authentic learner-produced data we have the opportunity to see how language progresses and what difficulties faced by the learners during language learning, and the reason behind their errors.

Concisely, there is a need for more research in the aspect of learner corpora in terms of pedagogical and academic needs, in order to produce a suitable room for discussing linguistics items in the interlanguage data provided by the language learners.

Therefore, the present research serves to find out the issues pertaining to the use of determiners by the Turkish EFL learners in order to contribute to the educational and linguistic disciplines.

One explanation of second of second language errors is that the learner is carrying over the habits of his mother tongue into the second language. “This is called interference and the implication of this term can only be that his mother tongue habits prevent him in some way from acquiring the habits of the second language” (Corder, 1981, p.24).



CHAPTER II

LITERATURE REVIEW

2. INTRODUCTION

The current chapter will provide information on the meaning or the definition of Corpus linguistics, some background information about learner corpora and corpus analysis. Moreover, this chapter presents information on description and some explanations about determiner (English Articles), and the review of the previous studies conducted in the field of Learner corpora related to English articles.

2.1. Corpus (Corpora)

Some dictionaries define corpora as necessarily consisting of structured collection of text specifically compiled for linguistic analysis, that they are large or that they attempt to be representative of language as a whole (Kennedy, 1998, p. 3).

“Corpora have been introduced into other linguistics disciplines as well and have succeeded in opening up new areas of research or bringing new insights to traditional research questions” (Meyer, 2004, p. 11). Claridge (2008, p. 242) “in all other respects, the defining characteristics of a corpus apply: it is a finite electronic collection of texts or parts of texts by various authors which is based on well-defined and linguistically relevant sampling criteria and aims for some degree of representativeness.”

Moreover, doing a corpus analysis will always involve work— more work than sitting in one’s office or study and making up the data for a particular analysis – but doing a corpus analysis properly will always have its rewards and will help us advance the study of human language, an area of study that linguists of all persuasions would agree we still know relatively little about (Meyer, 2004).

2.2. Types of Corpora

As corpora does not only depend on the collections of written and spoken sentences, it also deals with the “genuine theoretical resources, which implement in a series of applied fields of research, such as language teaching, translation studies or even machine translation and dedicated processing software (spelling, grammar, and style)”

(Otat, 2016, p. 41). Thus, due to the vast research fields covered by corpus studies, different sorts of corpora could be used. The types of corpora are as follows.

2.2.1. General Corpora

Large general corpora are sometimes referred to as reference corpora because they are often used as a baseline against which judgements about the language varieties held in more specialized corpora can be made (Evans, 2004). Hunston (2002, p. 15) infers that reference corpora are also attributed as general corpora which can be based on both spoken or written registers, because these kinds of corpora can be utilized in language learning and translation studies in the norm of a native corpus which is much larger than specialized corpora.

Cheng (2012, p. 212) defines the general corpora as a large corpus which attempts to represent the general usage of a language through the careful selection of a cross-section of texts. As a result of this aim the corpus is very large (containing some 100 million words) and contains a balance of texts from a wide variety of different domains of spoken and written language (Evans, 2004).

2.2.2. Specialized Corpora

Specialized corpora contain texts from a particular genre or register or a specific time or context. They may contain a sample of this type of text or, if the dataset is finite and of a manageable size, for example all of Shakespeare's plays, be complete (Evans, 2008).

The Michigan Corpus of Spoken English (approximately 1.7 million words of spoken data collected from a variety of different encounters at the University of Michigan), the International Corpus of Learner English (20,000 words taken from essays of students learning English as a foreign language) and the Nottingham Health Communication Corpus (Ludeling & Kytö, 2008, p. 245).

2.2.3. Historical Corpora

Historical corpora have enhanced our ability to study the linguistic development of English: such corpora allow corpus linguists not only to study systematically the development of particular grammatical categories in English but to gain insights into how

genres in earlier periods differed linguistically and how sociolinguistic variables such as gender affected language use (Meyer, 2004). These corpora are commonly used to carry on a study related to language used in the past.

The common examples of historical corpora are Helsinki Diachronic Corpus of English Texts (containing 1.5 million words written between 700 and 1700) and the ARCHER (A Representative Corpus of Historical English Registers) corpus (1.7 million words covering the years 1650 to 1990) (Meyer, 2004). “A secondary type of historical corpus is the combination of clone corpora covering discrete periods with an identical setup, e. g. the 1990s and (prospective) 1930s clones of the 1960s LOB and Brown corpora.” (Ludeling & Kyto 2008)

There are several aspects which have generally restricted the size of the historical corpora:

1. The manual process of compilation
2. The availability of suitable material (historical/cultural factors)
3. The accessibilities of material
4. Copyright

Ludeling & Kyto (2008, p. 245)

Like modern corpora, historical corpora can be designed for rather general purposes, enabling a wide range of linguistics investigation, or for a more narrowly focused research agenda, e.g., specific socio-pragmatic concerns.

(... longer periods providing a satisfactory amount of material are necessary and, in fact, historical corpora are usually concerned with time spans of one hundred years or longer).

Ludeling & Kyto (2008, p. 245)

2.2.4. Diachronic Corpora

A diachronic corpus is a type of corpora which provides information about how language systematically changes over time. Rizzo (2010) defines diachronic corpus as the dynamic and systematically embraces different periods in order to study language

changes and development. Diachronic corpora deal with the time difference of the use of language. Diachronic data covers historical data and may include words and usages that are now considered archaic or have fallen out of use altogether (Walter, 2010: p. 430).

McEnery & Hardie (2012) briefed that much work on diachronic variation has been done using the general methodology of comparable corpus samples across time (using datasets such as the Brown Family, the Helsinki Corpus, or ARCHER (A Representative Corpus of Historical English Registers)). The study of synchronic variation, by contrast, differs substantially depending on whether social variables or register variables are the focus of the investigation (p. 118).

Meyer (2002) highlighted that “with diachronic corpora (i.e., corpora used to study historical periods of English), the time-frame for texts is somewhat easier to determine, since the various historical periods of English are fairly well defined” (p.46).

2.2.5. Synchronic Corpora

Synchronic data is contemporary or recent material, used for general, contemporary dictionaries (Walter, 2010, p. 430). Most corpora contain samples of speech or writing that have been written or recorded within a specific timeframe. Synchronic corpora (i.e., corpora containing samples of English as it is presently spoken and written) contains texts created within a relatively narrow timeframe (Meyer, 2004).

2.2.6. Parallel Corpora

This is the type of corpora that contains text in two or more different languages, it is mostly used in translation studies and in building dictionaries. Evans (2004) added that “these are similar to comparable corpora in that they hold two or more collections of texts in different languages (Meyer, 2004). The main difference lies in the fact that they have been aligned so that the user can view all the examples of a particular search term in one language and all the translation equivalents in a second language.”

The typical example of parallel corpora is “the English–Norwegian Parallel Corpus contains samples of English and Norwegian fiction and non-fiction that are 10,000–15,000 words in length” (Johansson and Ebeling, 1996).

2.2.7. Comparable Corpora

This is the type of corpora that contains text from two different languages, and is used to compare languages. Teubert (1996), defines comparable corpora as two or more languages with the same or similar composition. Two or more corpora constructed along similar parameters but each containing a different language, or a different variety of the same language can be regarded as comparable corpora (Evans, 2004).

2.2.8. Monitor Corpora

Monitor corpora is a larger corpus which often contains constantly updated and constantly growing texts, used to track language changes in a different period of time. “A monitor corpus is one that is ‘topped up’ with new texts on a regular basis” (Evans, 2004). Monitor corpora deals with the changes that happen in languages. A monitor corpus aims to track language change by constantly adding new material and comparing it to a stable ‘reference corpus.’

[...some have argued that the only reliable way to study lexical items is to use what is termed a “monitor” corpus, that is, a large corpus that is not static and fixed but that is constantly being updated to reflect the fact that new words and meanings are always being added to English.]

(Meyer, 2002; p. 15)

Monitor corpora could be said to balance any need to be precise about the composition of a corpus against sheer size – as the corpus grows, we might assume that any skew in the data naturally self-corrects since there is no consistent skew in the data input (see McEnery & Hardie, 2012; p. 6-7).

2.2.9. Learner Corpora

Nesselhauf (2004b, p. 40) defines learner corpora as “systematic computerized collections of texts produced by language learners” Learner corpus research is a fairly young but highly dynamic branch of corpus linguistics, which began to emerge as a discipline in its own right in the late 1980’s/early 1990’s. Learner corpora, which can be roughly defined as electronic collections of texts produced by language learners, have

been used to fulfil two distinct, though related, functions: they can contribute to Second Language Acquisition theory by providing a better description of interlanguage (e.g. transitional language produced by second or foreign language learners) and a better understanding of the factors that influence it; and they can be used to develop pedagogical tools and methods that more accurately target the needs of language learners (Granger, 2008).

“The language learners whose language is covered by learner corpora are to be understood as foreign language learners, i.e. e. speakers who learn a language which is neither their first language nor an institutionalized additional language in the country where they live. (“Granger Learner corpora 2008”) Learner language is influenced by a wide variety of linguistic, situational and psycholinguistic factors, and failure to control for these factors greatly limits the reliability of findings in learner language re- search.”

(Granger, 2008, p. 125).

Corpus Linguistics can be defined as a systematic study of language computational authentic text produced by the native and non-native users for the purpose of language studies. Thus, Learner Corpora refer to the collection of authentic spoken or written data produced by the second/foreign language learners. Learner Corpora is natural production by the language learner, and are used in redesigning and/or reshaping the materials in which second/foreign language learners use in learning process, such as producing dictionaries, textbooks etc.

In addition, Learner Corpora are used in the study of language or language learners progress toward the target language learning, and learner corpora are also used to carry out the contrastive interlanguage analysis studies, by comparing the L1 and/or L2 students authentic spoken or written text, e.g. (Aijmer, 2009) ‘Modality in Advance Swedish Learners’ Written Interlanguage’. Moreover, Learner Corpora are used in Error Analysis through which language overuse, underuse and misuse by the non-native users are investigated, has become popular very recently, e.g., Khansir (2012) Error Analysis and Second Language Acquisition.

Granger (2004) states, learner corpus research opens up exciting pedagogical perspectives in different areas of pedagogy: materials design, syllabus design, language testing, and classroom methodology. Mukherjee (2006, p. 20) states, Learner Corpus

research is an area which is immediately relevant to the teaching and learning of languages, but its full potential has not yet been exploited in language pedagogy.

As corpora plays a vital role in linguistics community, it also provides sources from the daily usage of language to contribute to the development of language learning and making language teaching and learning more convenient, for instance; corpus contributes to constructing language learning materials, such as textbooks, dictionaries, etc. Meunier (2011) Cambridge University Press offers a 'real English guarantee' to the buyers and users of their materials, Longman assures its readership that [they] 'only see real English, as it is really used'. As for Macmillan, the use of their World English Corpus is described as 'a unique modern database of over 200 million words revealing fresh information on how words are used and natural examples of English as it is written and spoken now!'

2.3. Corpus Linguistics

Kennedy (1998) introduced corpus in language science as a body of written text or transcribed speech which can serve as a basis for linguistic analysis and description. He tirelessly added that "corpus linguistics did not begin with the development of computers but there is no doubt that computers have given corpus linguistics a huge boost by reducing much of the drudgery of text-based linguistics description and vastly increasing the size of the databases used for analysis. It should be made clear; however, corpus linguistics is not a mindless process of automatic language description. Linguists use corpora to answer questions and solve problems" (Kennedy, 2008, p. 2)

Before getting occupied with the significance of corpus linguistics in the field, it would be an innovative idea to start with the definition of the terminology "corpus". Taylor & Francis (2009) explained that "the notion of "corpus" refers to a machine-readable collection of (spoken and written) texts that were produced in a natural communication setting, and the collection of texts is compiled with the intention a) to be representative and balanced with respect to a particular linguistic variety or register or genre and b) to be analysed linguistically." (McEnery & Hardie, 2012).

Corpus linguistics today is often understood as being a relatively new approach in linguistics that has to do with the empirical study of "real life" language use with the help of computers and electronic corpora (Lüdeling & Kytö, 2008).

McEnery & Hardie (2012) define corpus linguistics as the study of language data on a large scale – the computer-aided analysis of very extensive collection of transcribed utterance and written texts. In other words, “Corpus linguistics is a linguistic methodology which is founded on the use of electronic collection of naturally occurring texts, viz. corpora” (Granger, 2012, p.4.). Gries (2009) also highlighted that Corpus Linguistics is one of the fastest growing methodologies in linguistics, which is also a lively and increasingly multilingual subject. McEnery and Wilson (2001) see corpus linguistics as the study of language based on examples of ‘real life’ language use. Moreover, A corpus can be seen, according to (McEnery & Hardie, 2012;1), as an area which focuses upon a set of procedures, or methods, for studying language (although, as we will, at least one major school of corpus linguists does not agree with the characterization of corpus linguistics as a methodology). Furthermore, corpus linguistics is not a monolithic, consensually agreed set of methods and procedures for the exploration of language. McEnery & Hardie, (2012, p. 1.) added that “corpus linguistics could be reasonably defined as dealing with some set of machine-readable texts which is deemed an appropriate basis on which to study a specific set of research questions.”

[...corpus linguistics is not a branch of linguistics on a par with these other branches, since ‘corpus’ does not tell you what studies is, but rather that a particular methodology is used. Corpus linguistics is thus a methodology, comprising a large number of related methods which can be used by scholars of many different theoretical leanings. On the other hand, it cannot be denied that corpus linguistics is also frequently associated with a certain outlook on language. At the centre of this outlook is that the rules of language are usage-based and that changes occur when speakers use language to compete with each other.] (Lindquist, 2009, p. 1)

2.4. Corpus Approaches

According to Fantinuoli & Zanettin (2015), Corpus linguistics has become a major paradigm and research methodology in translation theory and practice, with practical applications ranging from professional human translation to machine (assisted) translation and terminology. A corpus-based approach is seen as ‘a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study’

“A corpus-based approach facilitates the comparison of source and target segments and can help explore the extent of reflection on different socio-cultural realities and attitudes and the effect on the linguistic choices of subtitles” Fantinuoli & Zanettin (2015). Corpus-based studies dedicated to the analysis of various phenomena involving translations, (e.g., Teich 2003; Steiner 2004; Neumann 2013), etc. concentrate on the analysis of human translations only (Lapshinova-Koltunski, 2015, p. 93).

The corpus-based approach is ‘deductive’ because the reasoning works from the more general to the more specific, which is a ‘top-down’ approach (Tognini-Bonelli, 2001: 10–11). As stated in Fantinuoli & Zanettin (2015), A corpus-based approach allows for that range to be extended, by collecting texts into corpora and developing techniques for analysis. It has also allowed for different approaches to stylistic analysis to be developed. Two main approaches to the stylistic use of corpus linguistics currently being undertaken are corpus annotation and the analysis of collocation.

2.4.1. Corpus-based Versus Corpus-driven studies

Corpus-based studies have shed light on a number of distinctive linguistic features of academic discourse as compared with other genres (Gilquin, Granger, & Paquot, 2007), and added that “Corpus linguistics plays vital role in the linguistics, language teaching and learning field.” However, there are some contradictions whether the studies are corpus-based or corpus-driven studies, to make it clear, (McEnery & Hardie, 2012) differentiate the above classification of corpus studies as; Corpus-based studies typically use corpus data in order to explore a theory or hypothesis, typically one established in the current literature, in order to validate it, refute it or refine it. Meanwhile, Corpus-driven linguistics rejects the characterizations of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language (McEnery & Hardie, 2012).

Corpus linguistics serves as a remedy to language misuse (in terms of error analysis) and it provides the representation of authentic language usage, thus, corpus-based studies are studying that shade light as an evidence of the language existence and how the language is being used by the members of the community or is being used by a language learner. Moreover, Corpus-based studies give some overviews towards language progress, the future of a language, information about the languages and the feature of the language used in the past such as diachronic and synchronic studies and

used in terms of academic fields as in learners' learning progress known as learner corpora.

The distinction is not unimportant because, as we shall see, the position one takes is likely to influence the approach adopted in a corpus linguistic study. Simply put, those who see corpus linguistics as a methodology (e.g., McEnery et al., 2006, p. 7–11) use what is termed the 'corpus-based approach' whereby they use corpus linguistics to test existing theories or frameworks against evidence in the corpus. Those who view corpus linguistics as a discipline e.g., (Tognini-Bonelli, 2002). Biber (2009), use the corpus as the starting point for developing theories about language, and they describe their approach as 'corpus-driven'.

(Cheng, 2012: p. 6)

According to Widdowson (2004, p. 357), examines that “the most striking development in linguistic description over the past twenty years has been the use of the computer collected and analyse vast corpora of actually occurring language data” and he made mention of the “abundance dictionaries and grammatical descriptions which are corpus-based, and which chart the patterns of the contemporary usage of English”.

Moreover, corpus-based studies do not only rely on linguistics analysis or study of language development in general, Cheng (2012, p. 40) emphasizes that “in addition to tracking English language changes and production of dictionaries, textbooks and other learning and reference materials, a major application of corpora is to study different aspects of linguistics, including lexis, multi-word units, grammar, discourse, pragmatics, discourse intonation, registers (e.g., academic, business, social, scientific and legal) and genres (e.g., university textbooks, financial reports, conversation, plays and contracts).”

Unlike the corpus-based studies, a corpus-driven study is the process whereby corpus data is used as a tool of the study. Data is used as a support or a resource to shape the way, technique, or methods of language teaching. “Corpus-driven linguistics are said to be strictly committed to ‘the integrity of the data as a whole’ (Ibid.: 84) and therefore, in this latter approach, it is claimed that ‘[t]he theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus’ (ibid.: 85)” (McEnery, Xiao & Tono, 2006).

When the negative fact about corpus-driven approaches is put into consideration (McEnery, Xiao & Tono, 2006) made a mention of a certain problem (e.g., frequency).

[... problem for the corpus-driven approach relates to frequency. While it has been claimed that in corpus-driven approach corpus evidence is exploited fully, in reality frequency may be used as a filter to allow the analyst to exclude some data from their analysis] (p. 9).

The corpus-driven approach has famously been adopted by Johns (1991a, 1991b) to become a powerful language learning methodology: data-driven learning (DDL). In DDL, the language learner is also a language researcher and, to better learn the language, the learner accesses corpora to explore and make sense of patterns of language use. The approach is captured very succinctly by Johns (1991a: 30): ‘research-then-theory’ (Cheng, 2012).

“In a corpus-driven approach the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus . . . The theory has no independent existence from the evidence and the general methodological path is clear: observation leads to hypothesis leads to generalization leads to unification in a theoretical statement.” (Tognini-Bonelli, 2001: 65–6).

It appears, then, that the corpus-driven approach is not so different from corpus-based approach – while the latter allegedly insulates theory from data or standardizes data to fit the theory, the former filters the data via apparently scientific random sampling, though there is no guarantee that the corpus is not explored selectively to avoid inconvenient evidence.

However, in some cases, “Corpus-driven linguistics rejects the characterizations of corpus linguistics as a method and claims instead that the corpus itself should be the sole source of our hypotheses about language” (McEnery & Hardie, 2012).

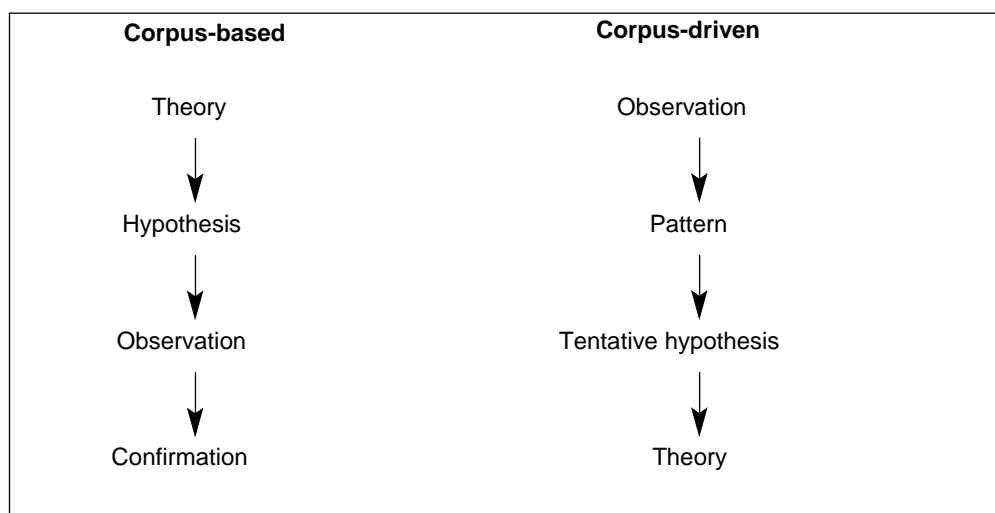


Figure 1. Corpus-based and corpus-driven approaches to corpus linguistics

Source: Cheng, 2012; p. 187

Theories are proven through corpus-based approaches, while corpus-driven form theories the hypothesis for the observation of language usage.

"The idea that data and theory become in some way co-terminal is sometimes expressed in slightly different terms – such as the corpus is the sole source of a corpus linguist's theory of language – as part of the so-called corpus-driven approach (a label we will discuss shortly)" (McEnery & Hardie, 2012). Sinclair (2004) states "that such an approach eschews prior linguistic categorization, with the implication being that such theories are simply a distraction from what we might term the 'pure' theory hidden within the data" (p. 191).

"Gries (2010b: 328–9) also treats corpus-driven as a methodological term synonymous with 'bottom-up', and, going even further than Biber, argues that all or nearly all self-identified corpus-driven studies cannot be considered 'corpus-driven' in this sense."

(cited in McEnery & Hardie, 2012; p. 187)

2.5. Corpus and Language Teaching

Over the past two decades, corpora (i.e., large systematic collections of written and/or spoken language stored on a computer and used in linguistic analysis) and corpus evidence have not only been used in linguistic research but also in the teaching and learning of languages probably a use that "the compilers [of corpora] may not have foreseen" (Johansson 2007).

When we talk about the application of corpora in language teaching, this includes both the use of corpus *tools*, i. e. the actual text collections and software packages for corpus access, and of corpus *methods*, i. e. the analytic techniques that are used when we work with corpus data. In classifying pedagogical corpus applications, i.e., the use of corpus tools and methods in a language teaching and language learning context, a useful distinction (going back to Leech 1997) can be made between direct and indirect applications (Römer, 2008).

As Barlow (1996) notes, “[t]he 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). As cited in Römer (2008) “Large general corpora have proven to be an invaluable resource in the design of language teaching syllabi which emphasize communicative competence (cf. Hymes 1972, 1992) and which give prominence to those items that learners are most likely to encounter in real-life communicative situations” (p. 114).

The contents of this new, corpus-driven “lexical syllabus” are “the commonest words and phrases in English and their meanings” (Willis 1990, p. 124). With its focus on lexis and lexical patterns, the CCEC responds to some of the most central findings of corpus research, namely that language is highly patterned in that it consists to an immense degree of repeated word-combinations, and that lexis and grammar are inseparably linked (cf. Hoey 2000; Hunston 2002; Hunston/Francis 2000; Partington 1998; Romer 2005a, 2005b; Sinclair 1991; Stubbs 1996; Tognini Bonelli 2001).

The immediate accessibility of authoritative information about the language is also a major advantage for language *teachers* who decide to interact with a corpus. As a recent survey on teachers’ needs has shown (cf. Romer forthcoming), teachers often require native-speaker advice on language points. (“Corpus and Teaching | Language Education | Linguistics”) Computer corpora have been described as “tireless native-speaker informant[s], with rather greater potential knowledge of the language than the average native speaker” (Barnbrook 1996, p. 140), which can provide excessive help in dealing with the corpus data.

Römer (2006, p. 121) however states that “despite the progress that has been made in the field of corpus linguistics and language teaching, the practice of ELT has so far been largely unaffected by the advances of corpus research.”

“In the unintentional case, teachers are often not aware of the possibilities offered by corpora or not aware of the changes that corpus methods have brought to materials that they are sometimes actually using. This may be due to a lack of information at the pre- or in-service teacher training levels and/or to the sometimes too vague statements found in the introductions to teaching materials, which might refer to the corpus-informed nature of the materials but not explicitly list the pedagogical implications of this corpus-based nature. As for the intentional negation of the benefits of corpus research in L2 learning, it may be caused by the oft-cited ‘ivory tower effect’, i.e., the perception by teachers that linguists work in their offices at university and have no idea of what teaching is about - and this despite the fact that some of those linguists are also teachers.”

Meunier (2011, p. 461)

Another reason explaining the lack of uptake of corpus methods in instructional settings (and one that is especially worth taking into account by corpus linguists advocating the applied relevance of their research whilst not being directly involved in teaching) is that the importance of using authentic, corpus-based descriptions of the target language is only one line of thinking among many other influential ones in L2 language learning.

More practical and methodologically oriented views have however also been put forward against the use of a native speaker model in L2L: it might set goals that are unachievable, unrealistic, and unnecessary for the actual needs of the learners, and might not be appropriate to the socio-cultural conventions of the groups acquiring the L2. Another key issue that has attracted controversy in L2 language learning and teaching circles is the issue of frequency which is unmistakably entwined with corpus linguistic research.

Meunier (2011, p. 463), highlighted that “the last factor accounting for the lukewarm reception of corpora in the classrooms is the lack of empirical studies exploring the actual impact of corpus methods on the learning outcomes”.

In the case of analysing data learner corpus research and L2 language learning, Meunier (2011), infer that “the fact that L2L has remained largely unaffected by the advances of native corpus research is even truer when one looks at the L2 teaching/learning applications of learner corpus research” (p. 464). And he added that “It would be wrong to state that learner corpora have not been used by ELT publishers, but

it must be acknowledged that they have been used to a much smaller extent than native corpora.”

“Another reason accounting for the limited impact of learner corpora in instructional settings is put forward by Flowerdew, who states that in most studies of learner corpora “the implications for pedagogy are not developed in any great detail with the consequences that the findings have had little influence on [...] syllabus and materials design” (1998, p. 550).’ Cited in Meunier (2011, p.465). corpus compilers will urgently have to address the learner’s needs for what Braun (2005) calls ‘pedagogical relevance’ or what Belz and Viyatkina (2008) refer to as ‘authentication’.

2.6. Contrastive Analysis

Contrastive analysis was originally developed by Charles Fries (1945), expanded, and clarified by Robert Lado (1957), and demonstrated by innumerable dissertations and, at its best, in a still widely used series of studies under the editorship of Charles Ferguson (Spolsky, 2014). To investigate L2 learners' errors, Contrastive Analysis theory (CA) has first appeared, and it was the well-known theory in the SLA or L2 learning fields. The central assumption of this theory was that L2 learners' errors are due to negative interlingual interference from their mother tongue (MT) (Hamad Al-Khresheh, 2016; p. 49). In another study Al-Khresheh (2015; p. 123), states that "the main criticism of CA was that interlingual interference from first language is not the only reason for the occurrence of errors in SLA.”

CA was an effective theory and famous for its ability to compare between the structures of two languages (L1 & TL) in order to identify the areas of similarities and differences between them (Al-khresheh, 2013). In other words, “CA is founded on the assumption that L2 learners will tend to transfer the formal features of their L1 to their L2 utterance. Simply put, this notion of 'transfer' means carrying over the habits of the MT into the L2” (Al-Khresheh, 2016b).

According to Gass and Selinker (2001, p. 72), "contrastive analysis is a way of comparing languages in order to determine potential errors for the ultimate purpose of isolating what needs to be learnt and what does not need to be learnt in an SLL situation". Federation et al. (2014), stated that contrastive analysis assumed that many of the mistakes made by learners are caused by differences between the native and target

languages, and led to a large number of extremely valuable language descriptions and pedagogical grammars.

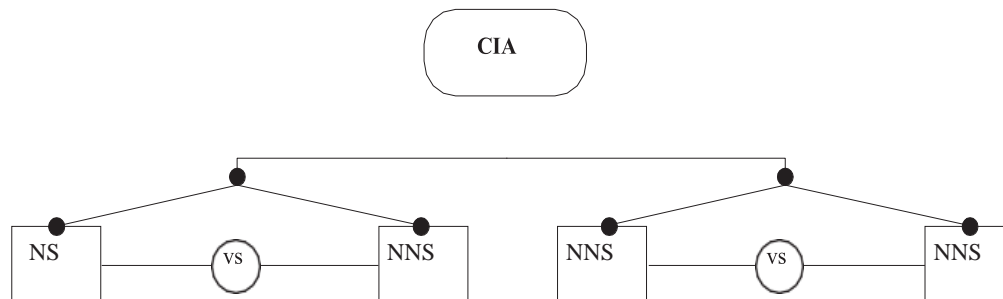


Figure 2. Contrastive Interlanguage Analysis Adapted from

Source: Granger et al., 2002; p. 12

Contrastive Interlanguage Analysis (CIA) is an approach which has been used in carrying out corpus studies to compare the retrieved data from one corpus to another, (Akban, 2018). According to Granger (2003), comparison in CIA is of two kinds: 1, The native and learner language (L1 vs L2) and 2, Different varieties of interlanguage (L2 vs L2).

2.7. Error Analysis

In the field of SLA, EA was first established by Corder et al. in the late 1970s and became a very popular approach for describing L2 errors. Starting with Ellis definition of error, Ellis (1997) defines error "as reflection of learner's knowledge and it occurs because the learner does not know the correct one". Error analysis in learner corpora serves as an instrument used in detecting learner errors and learner progress toward language learning. James (1998:1) defined Error Analysis as "the process of determining the incidence, nature, causes and consequences of unsuccessful language". Mahmoodzadeh (2012), EA could be defined as a procedure used to identify, categorize, and explain the errors committed by FL/L2 learners. EA, however, deals with "the learners' performance in terms of the cognitive processes they make use of in recognizing or coding the input they receive from the TL (Target Language)" (Erdogan, 2005, p. 263).

As a language learner it is a norm to make errors while using the target language. “Learning has generally been taken to be an inductive process of trial and error” (Leuc et al., n.d. p. 845). Errors are essential parts of human learning processes just as error correction is an indispensable part of the teaching process and, unlike mistakes, systematic (Can, 2017).

Corder (1975), in their study mentioned that “traditionally the attitude to errors was that they were a sign that the learner had not yet mastered the rules he was taught and that they were therefore to be dealt with by repeating the explanations until they disappeared.”

Coder and James (1998) explained that error analysis is of two crucial aspects, which are Error analysis tells the teacher what needs to be taught, and it tells the researcher how learning proceeds. Error analysis (EA) is an everlasting endeavour for the simple reason that first, second or foreign language learner, Mahmoud (2013, p. 2). Moreover, Mahmoud (2013) added that “EA is indispensable in language teaching and it is assuming more importance as an index of language learning and communication strategies” (p. 2). In other word, Sridhar (1980, p. 101), inferred that “the best way to discover transfer is through error analysis.” Altenberg and Granger (2002,), “it [error analysis] is a key aspect of the process which takes us towards understanding interlanguage development and one which must be considered essential within a pedagogical framework” (p. 14).

“Analysis of second language learner’s errors can help identify learner’s linguistic difficulties and needs at a particular stage of language learning. (“Error Analysis | Second Language Acquisition | Language ...”) In general, Error analysis has several implications for the handling of learner’s errors in the classroom as follows:

1. Devising remedial measures
2. Preparing a sequence of target language items in classrooms and textbooks with the difficult items coming after the easier ones.
3. “Making suggestions about the nature or strategies of second language learning employed by both first and second language learner.” (Khansir, 2012, p. 1029)

Analysing learner errors is a useful way of finding out the stage of learning learners have arrived at (Flowerdew & Mahlberg, 2009).

Error analysis, deriving essentially in its recent importance from the seminal paper by Corder (1967), showed how systematic errors are interpretable as giving evidence of the learning process and so brought research in second language acquisition into theoretical connection with work in first language acquisition (Spolsky, 2014; p. 251).

Hamad Al-Khresheh (2016), stated that “according to EA, a considerable number of errors made by FL learners are similar regardless of their MT. Such errors are caused due intralingual interference or transfer.” Moreover, Schachter and Celce-Murcia (1977), defined interlingual error as "those [the errors] caused by the influence of the learner's MT on production of the TL in presumably those areas where languages clearly differ" (p. 443).

According to Corder (1973), there are two main objectives of EA: one theoretical and the other being known to be applied. The theoretical objective checks the validity of the theories such as the theory of transfer. In other words, this objective can help in understanding how and what an FL learner learns whilst studying an FL. On the other hand, the applied objective, "concerns pedagogical purposes" (Mahmoodzadeh, 2012, p. 735), as the learner uses schemata to learn or acquire a target language.

Dulay et al. (1982), noted two main reasons why error analysis is being into existence.

1. To provide data from which interferences about the nature of the language learning process can be made.
2. Indicates to teachers and curriculum developers which part of the TL students have most difficulty producing correctly, and which error types detract most from a learner’s ability to communicate”

(Dulay et al. 1982, p. 138)

According to Ellis (1994), EA is carried out in four consecutive stages as stated by Ellis (1994, p. 48). These stages are as:

1. Collection of a sample of learner language,
2. Identification of errors,
3. Description of errors,
4. Explanation of errors.

Many error analyses have revealed linguistic differences between the sentences produced in a second language by second-language learners and corresponding sentences produced by native speakers, (Tarone, 2006).

2.7.1. Traditional Error Analysis

Error Analysis (EA) has a longest tradition. Yet, until recently a typical EA went little beyond impressionistic collections of "common" errors and their taxonomic classification into categories (mistakes of agreement, omission of articles, etc.) (Sridhar, 1980, p. 21). Traditional error analysis is carried manually and systematically to study the errors made by the FL learner during TL productions. Sridhar (1980) added that "little attempt was made either to define "error" in a formally rigorous and pedagogically insightful way or to systematically account for the occurrence of errors either in linguistic or psychological terms. Hence it is substantially correct to say that traditional EA was an ad hoc attempt to deal with the practical needs of the classroom teacher" (p. 20).

Commenting on these setbacks, Dagneaux et al., (1998) list five limitations pertaining to the nature of the data and related error typologies used by traditional EA and:

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

Cited in (Can, 2018; p. 78)

An error analysis of learner corpus data only lifts a corner of the veil. Equally important are indications as to what learners get right, what they underuse and what they overuse" (Gilquin 2007, p. 288).

2.7.2. Computer-aided Error Analysis

According to (Dagneaux, Denness, & Granger, 1998), the early 90s saw the emergence of a new source of data for SLA research: the computer learner corpus (CLC),

which can be defined as a collection of machine-readable natural language data produced by L2 learners. “The emergence of computer learner corpora in the early 1990s enabled us to carry out computer-aided error analysis (CEA). It was developed to overcome most of the drawbacks of traditional EA, and it has an advantage in the storing and processing of enormous amounts of information about various aspects of learner language” (Abe, 2007). Computer-aided error analysis 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).

2.7.3. Interlanguage

Tarone (2006), stated that “the Interlanguage (IL) hypothesis claims that second-language speech rarely conforms to what one expects native speakers of the TL to produce, that it is not an exact translation of the NL, that it differs from the TL in systematic ways, and that the forms of the utterances produced in the second language by a learner are not random.”

“Interlanguage, a term coined by Selinker (1972), that seems to make somewhat stronger claims than implied by Corder (1971) with "idiosyncratic dialect" or Nemser (1971) in a more Skinnerian framework with "approximative system," makes even clearer the similarity to first language acquisition studies by focusing attention on the learner's knowledge of the target language as a whole, and provides room, through the notion of fossilization, for an ultimate connection with various sociolinguistic studies of language variation.”

(Spolsky, 2014)

Selinker (1972), suggested that the existence of a latent psychological structure within which interlingual identifications and the processes and strategies underlying second-language learning are located.

2.8. English Determiners

Determiners can be defined as the words that accompany nouns or as the noun modifiers which usually appear before noun phrases to specify or to serve as a reference

of a noun. Determiners form a set of closed- system items that are mutually exclusive with each other, i.e., there cannot be more than one occurring before the noun head. Both **a the boy* and **a some boy* are ungrammatical. The determiners are in a 'choice relation', i.e., they occur one instead of another. In this respect they are unlike *all*, *many* and *pretty* which are in a 'chain relation ', i.e., occurring one after another: *All the many pretty houses*. “Determiners are words which specify the range of reference of a noun in various ways, such as by making it definite (*the book*), indefinite (*a book*) or by indicating quantity (*many books*)” (Ho-Abdullah, Jaludin & Safinaz 2013). The articles are central to the class of determiners in that they have no function independent of the noun they precede (Zandvoort, 1974). As shown in the diagram below.

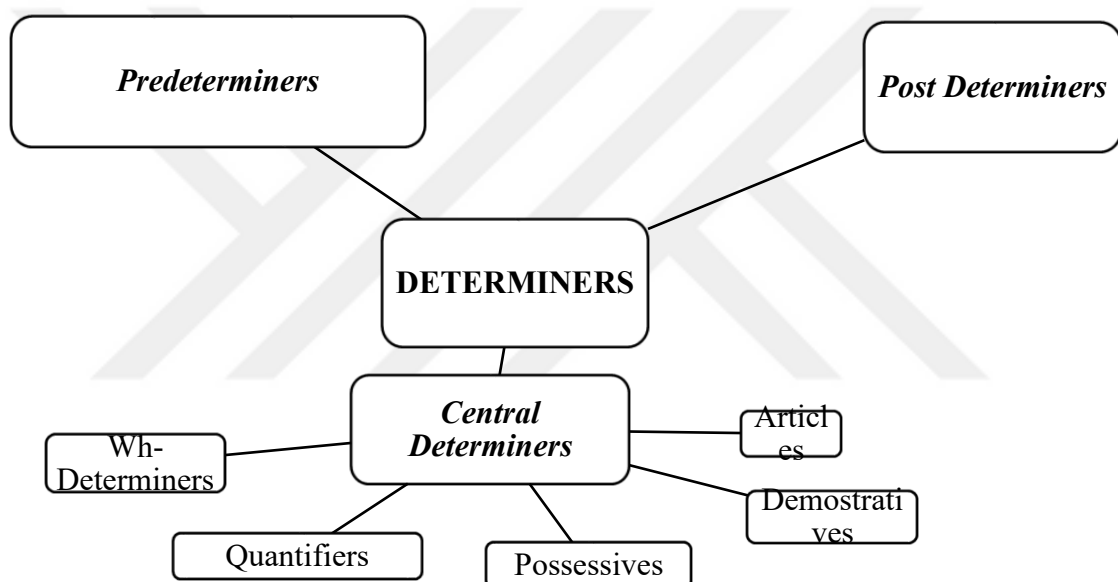


Figure 3. Categories of English Determiners

Source: Selvain, 2003, p. 3

English determiners are words which specify the range of reference of a noun, e.g., by making it definite (*the book*), indefinite (*a book*), or by indicating quantity (*many books*) (Ho-Abdullah, Jaludin & Safinaz 2013). Determiner is a key function in the structure of NP. When a determiner is added to a nominal, a construction at the NP level is formed. (“fuzziness of nominal determination”) In the case of nominals headed by singular nouns, addition of a determiner is generally obligatory:

- i. [new car] was stolen. [bare count singular nominal]
- ii. [The/One/Ally's new car] was stolen. [determiner + nominal]

In addition, the articles have no lexical meaning but solely contribute definite or indefinite status to the nouns they determine (Zandvoort, 1974).

2.8.1. English Articles

“The term article is used for the special subcategory of determinatives that provide the most basic expression of definiteness and indefiniteness. (“nouns-and-noun-phrases.pdf | Grammatical Number | Noun”).” According to the Longman grammar of spoken and written English, the definite and the indefinite article are the most common and most basic determiners. Both articles take a different spoken form when the word begins with a vowel:

<i>a house</i>	<i>an hour</i>
<i>a Member of the Parliament</i>	<i>an MP</i>
<i>the house</i>	<i>the hour</i>
<i>the Member of Parliament</i>	<i>the MP</i>

Source: McGregor, 1997, p. 285

English articles are classified into two groups definite and indefinite articles. A definite article shows the concreteness of an object, a definite article does not only tell us about the existence of the object, but also it serves as a reference for the object, e.g., John bought the car we saw yesterday. So that means we have some information about the car, the sentence is referring to the car we came across. Meanwhile, indefinite articles which are ‘a’ and/or ‘an’ serves to provide information about a noun that we have no idea on, like an abstract with which we did not come across yet. E.g., John bought a car. It is talking about an item which we do not know anything about. Here we heard that John bought a car, but we do not have any idea about the colour, or any other features of the car, in general.

2.8.1.1. The definite article (the)

According to McGregor (1997, p. 285) definite determiners realize their identifying function by ‘indexing’, that is, by locating ‘the indexed item in some real or abstract space in reference to the location of the person making the sign, or of the index itself’. Yet, Zandvoort, (1974) the dependence is not unilateral. A count noun like *boy*, for example, is, on its own, only a lexical item. It requires an ‘overt’ determiner of some kind to assume grammatical status. Furthermore, it is generally accepted that the meaning of indefinite determiners is ‘not sufficient to put the hearer in mental contact with a uniquely determined instance of the category’ (Langacker, 1991: p. 104).

The definite article [the] is the most basic indicator of definiteness. It is illustrated in [i] which shows that it is compatible with all types of common noun: count singular, count plural, and non-count. (“nouns-and-noun-phrases.pdf | Grammatical Number | Noun”)

i. *Bring me [the ladder/ladders/cement]!*

“Use of the definite article here indicates that I expect you to be able to identify the referent – the individual ladder, the set of ladders, the quantity of cement that I am referring to. (“nouns-and-noun-phrases.pdf | Grammatical Number | Noun”).”

Source: Langacker, 1991: p. 104

2.8.1.1.1. Definiteness and Specificity

Apart from the usual classification of articles, like definite (the) and indefinite articles (a, an), there is rather another classification of the articles as in definiteness and specificity. It has been suggested that articles cross-linguistically can be used to represent different semantic notions such as definiteness, specificity (e.g., Ionin et al., 2004).

Definiteness, as a semantic/pragmatic notion, is associated with the notion of identifiability, which depends on the existence and uniqueness of the referent in one of the pragmatically delimited domains mutually manifest to speaker and hearer in real time. In other words, a discourse referent is definite if the speaker intends to refer to it and expects the referent to be uniquely identifiable to the hearer, too (Trenkic 2008: p. 4).

Specificity is taken to mean the speaker's intent to refer to an individual or thing that exists in the world as conceptualized by the language user (Trenkic2008: p. 2). It is important to note that an intention to refer presupposes that the speaker has a particular referent in mind. Furthermore, the existence in the world without the speaker's intent to refer is insufficient to account for article choice in L2 English (Ionin et al. 2004; Trenkic 2008).

Ionin et al. (2004) briefly explained the matters of definiteness and specificity below.

“If a Determiner Phrase (DP) of the form [D NP] is...

1. [+definite], then the speaker and hearer presuppose the existence of a unique individual in the set denoted by the NP.
2. [+specific], then the speaker intends to refer to a unique individual in the set denoted by the NP and considers this individual to possess some noteworthy property.”

(Ionin et al. 2004, p. 5)

2.8.1.1.2. Uses of Definite Articles

Heim (1991) state that [+definite] is a semantic feature expressed morphologically as the. Furthermore, Hawkins (1978) provided the taxonomy of definite e.g., anaphoric, associative, larger situation.

I. Anaphoric use

- a. I saw a cat. I gave the cat some milk.

II. Associative use

- a. A man drove past our house in a car. The exhaust fumes were terrible.
 (“Final, non-type-set version of”)

III. Larger (cultural) situation uses

- a. (when invited to wedding)
 - i. Have you seen the bridesmaids?

1. Anaphoric Use
2. Associative Use
3. Immediate Situation Use
4. Larger Situation Use

Von Stechow and Kaiser (2003:44) define definiteness as “a discourse-pragmatic property which indicates that the discourse referent associated with a definite expression can be identified with an already introduced discourse item.”

Definiteness is a discourse-related semantic feature: it is related to the knowledge state of the speaker and hearer in the discourse (Ionin, 2003: 3).

This is a situation whereby the speaker and the audience share the same background knowledge about the item the speaker is talking about. As (Abbott, 2005:2) state, “the choice of the article depends on the knowledge of the speaker of his/her own knowledge and of others’ knowledge as well.”

Lyons (1999) argues that “definiteness is a grammaticalization category: that originally definite NPs were understood to denote identifiable entities, but as a consequence of the category becoming grammaticalized have acquired other use. In other words, definiteness is a category concerned with the grammaticalization of identifiability and non-identifiability of referents on the part of a speaker or addressee” (p. 278).

2.8.1.1.2.1. Anaphoric Use

Before getting acquainted with the usage of definite article in the anaphoric way, it would be more convenient to define the term anaphora. “The term anaphora implies that reference to a noun preceded by ‘the’ is known from the linguistic context” (Quirk et al., 1985:267). The above citation suggests that the information given by the speaker in the sentence serves to provide a comprehensive reference to the listener or the audience.

2.8.1.1.2.2. Associative Anaphoric Use

“Speaker and hearer may have (shared) knowledge of the relations between certain objects evoked by the discourse (the triggers) and their components or attributes (the associate): associative anaphoric uses of definite noun phrases exploit this knowledge (Lyons, 1999).

2.8.1.1.2.3. Immediate Situation Use

“The definite noun phrases whose referent is a constituent of the immediate situation in which the use of the definite noun phrase is located, without necessarily being visible (Lyons, 1999) Quirk et al. (1985:266) explain that the referent marked by the definite article ‘the’ is known from the immediate situation” (Al-saaidi, 2011).

2.8.1.1.2.4. Larger Situation Use

Hawkins (1978: 410) lists two classes of definite noun phrases that are used in situations in which the speaker appeals to the hearer ‘s knowledge of entities existing in the non-immediate or larger situation of utterance—knowledge they share by being members of the same community.

2.8.1.2. The indefinite article (a/an)

McGregor’s (1997) highlights, indefinite determiners do not have ‘indexical’ meaning. Indefinite nominals instruct the hearer to conceptualize instances as corresponding to the categorization specified by the speaker.

The indefinite article [*a*] is the most basic indicator of indefiniteness for singular count nouns. Its compatibility with plurals is due to its historical origin as an unstressed form of the numeral *one*. For example.

i. *a house*

ii. *a car*

The other case is specific to one grammaticized word, the indefinite article, which has *an* as a liaison form before vowels:

i.a. *a fool*

b. *an idiot*

ii.a. *a by no means ugly man*

b. *'an' in some ways handsome man*

With indefinite NPs the addressee is not being expected to be able to identify anything. NPs headed by singular count nouns permit a direct contrast between indefinite *a* and definite *the*:

- i. *Bring me a ladder!*
- ii. *Bring me the ladder!*

Source: McGregor, 1997, p. 286

2.9. Turkish Determiners

The Turkish language does not have a definite article, but it does have an indefinite article: *bir* 'a'. While this is the same word as the numeral meaning 'one', the position of the article is different from that of the numeral: the article follows any adjective in the noun phrase and immediately precedes the noun, while the numeral is phrase-initial:

- i.(415) **bir** giizel, olgun elma
(**one** nice, ripe apple)
["**one** nice ripe apple"]
- ii.(416) giizel, olgun **bir** elma
(nice ripe **an** apple)
["**a** nice, ripe apple"]

Source: Kornfilt, 1997, p. 106

2.9.1. Turkish Indefinite Article

Determiners fall into two classes, indefinite and definite, according to the referential status they impart to the noun phrase in which they occur.

Bir *a(an)*

The indefinite article *bir* usually follows the adjective, but can precede it:

- i. genç **bir** adam/**bir** genç adam
[**a** young man]

ii. açık **bir** penceere/**bir** açık pencere

[*an* open window]

There is a subtle difference in meaning between the two-word order with the indefinite article. Because modification within nominal phrases in Turkish operates left and right, placing the adjectival to the left of *bir* has the effect of making the sequence [adjectival + head] be perceived as a semantic unit. Consequently in (i) the form *bir genç adam* donates a member of the class ‘young man,’ whereas the form *genç bir adam* donates a young member of the class ‘man.’ This latter order, which is much the more common, gives more prominence to adjectives.

Source: Kornfilt, 1997, p. 106

2.10. English Articles in Language Acquisition Researches

According to Brown (1973) and Maratsos (1974), it is inferred that the native speakers, the proper use of English articles are acquired unconsciously and at an early stage. However, some studies claim that it is difficult to acquire English articles. Starting with Master (1987), stated that “unlike the acquisition of English article by L1 learners, the English article system is one of the most difficult of aspects of English grammar for ELLs (English Language Learning) and one of the last to be fully acquired.” Cook (1994) states that “the most persistent recurrent problems in terms of slips and minor errors amongst all groups of the most sophisticated users of English as a second language involve confusion with the words *the*, *a* (*an*) and their proper necessary non-use” (p. 27). Furthermore, Pica (2002; 2) also identifies certain sentence modality and modification of nouns with articles as grammatical structures that are very difficult for second language learners.

Moreover, Snape, García-Mayo, and Gurel (2013) studied L2 acquisition of English generic NPs by Spanish-speaking English Language Learners (+Art), Turkish-speaking ELLs whose L1 had an indefinite article but no definite article, and Japanese-speaking ELLs [ESL Learners] (–Art). They found that L2 article preference was mainly influenced by their L1.”

Furthermore, according to Bickerton (1981), the use of articles is governed by semantic binary features [\pm Specific Referent] and [\pm Hearer Knowledge] of noun phrases in discourse. Hawkins et al. (2006) put forward a feature-based account of the individual differences in the use of articles found among the Japanese speakers using the Distributed

Morphology model (based on Halle & Marantz, 1997), which predicts that the difficulty lies between matching the features of vocabulary items (the phonological exponents) and the terminal nodes.

Another study carried out by (Dagdeviren, 2010) with the aim of investigating the use of English articles by the Turkish speakers in the EFL setting. The participants of the study for two different proficiency groups (low and high proficiency), which consisted of 27 males and 33 females from different departments with the average age of 20.45 years. The result of the study came up with a significant difference compared to the previous studies like (Murphy, 1997; Master, 1997; Mizuno, 1999; Yoon, 1993; Robertson, 2000; and others). Moreover, it was found that “neither the low- nor the high- proficiency level participants’ overuse of *the* and *a* were found higher than the accuracy rates in [-definite, -partitive, +specific], [-definite, +partitive, -specific] sentences and [+definite, -partitive, -specific] sentences, respectively” (Dagdeviren, 2010).

A study on the Chinese speakers’ L2 article use was done by (Snape, 2009). After investigating 38 adult Mandarin Chinese speakers that are residing in Canada who are learning English for Academic Purpose between the ages of 18-40. The results of the study showed that the learners were consistent with the fluctuation hypothesis. “the intermediate L2 learners fluctuated between the two features [+definite] and [+specific], sometimes incorrectly selecting *the* as a marker for [+specific] and *a* as a marker of [-specific]” (Snape, 2009).

“L1 Chinese L2 English learners in the current study are able to acquire interpretable features but have problems in selecting the correct feature for English i.e., [+definite] for articles, not [+specific] because there are no L1 transfer effects i.e., no articles in Chinese so L2 learners have to work out whether the L2 encodes definiteness or specificity for articles. Because Mandarin Chinese is still a language which lacks articles i.e., equivalent semantic functions to article languages, there is no direct L1 transfer as there is from Spanish to English.”

(Snape, 2009: p. 45)

Likewise, the study of testing interface hypothesis (IH) [*If the Korean learners of English show higher accuracy rates with generic uses of English articles (internal interface) than with (in)definite uses of them (external interface), it would lend support to the IH*] on the Korean L2 learners’ acquisition of English articles was carried out by

(Park, 2013) by studying the total number of 77 Koreans L2 speaking learners of English and involved 21 English native speakers as control group. The results of the study suggested the two following points: “a) the prediction based on the distinction between Internal interface and external interface with IH (Interface Hypothesis) may not be sustainable regarding L2 acquisition of article uses. b) External interfaces are not entirely problematic in the L2 acquisition of articles” (Park, 2013).

Nevertheless, a study carried out by (Alimi, 2007) on some Batswana students’ use of English articles and modals in their writing tasks. The study was carried out with 514 graduate and ungraduated students from the Department of English at the University of Botswana. The results of the study revealed that the articles are regularly omitted, mixed up, or used redundantly; and that the students are more inclined to express different types of modality using only can/could. Moreover, the study added that “the complexities involved the articles and modals make them difficult to explain to the students, especially as the grammatical system of their L1 differs significantly from that of the target language” (Alimi, 2007, p. 220).

The above research suggested that lack of articles in L1 or the differences of article formation in L1 affect the students’ acquisition of English articles. As could be seen in the above studies, the problem of article acquisition could be resulted from lack of exposure and lack of transfer of article forms the mother tongue.

CHAPTER III

METHODOLOGY

3. INTRODUCTION

This chapter deals with the designs and methodologies used throughout the study. First of all, the data collection tools and the participants who took part in the study will be discussed. Then, the description of the programs used in the study, the use of corpus-driven techniques in language classroom will be stated.

3.1. Research Design

The study was conducted in order to determine the use of determiners (English articles), look into the errors made during the language usage and to look into the interlanguage influences on the use of determiners while using the target language by the Turkish EFL learners through analysing the written text of Turkish sub-corpora in the Cambridge Learner Corpus (CLC). Mixed method was used in the study, whereby qualitative method was used to analyse the effect of L1 on the target language and quantitative method was used to analyse and interpret the frequency of errors obtained from the data.

The present study focuses on the Turkish EFL learners across all (A1-C2) CEFR proficiency levels, the participants (learners) were between the age of ten to 50+. This study aims to use a quantitative research method to describe the frequency of English determiners (articles) used in the Turkish EFL learners' sub-corpus. Moreover, the study does not have to do with experimental study, however, in order to see the implications of corpus-driven techniques a short study was conducted, whereby the participants were divided into two different groups, a study (experimental) group and control group. The experimental group used corpus-driven techniques while the control group used traditional method of teaching English determiner (article).

Furthermore, the study seeks to find which of the English determiners (articles) has been underused and overused by the learners.

"Many studies of grammatical development in second language acquisition (SLA) could be classified into: a) studies that examine the formal features or b) studies that seek to

gauge overall progress by a development index" (Bardovi-Harlig and Bofman, 1989; p. 18). Considering the above statement of Bardovi-Harlig and Bofman, the present study falls into the second category, as the study seeks to find out the error made by the Turkish EFL learners while using the target language and to define the reasons behind the errors.

3.2. Turkish EFL Learners' Sub-corpora

The study bows to use the data provided by Cambridge Learner Corpus which consists of 29,266,800 words and is considered as the largest annotated learner corpus. The demographic information of the collected data (text) shows that the data was produced by different learners from 130 different native language backgrounds who took part in the Cambridge English for Speakers of Other Languages Examination. The texts provided the producers' demographic information, which consists of; their mother tongue, age, nationality, gender, CEFR level, test types, and text format.

From the aforementioned corpora, there extracted a Turkish EFL learner sub-corpus which comprises a total number of 275969 words collected from the learners from A1 - C2 CEFR level of proficiency.

Table 1.

Turkish EFL learners' Sub-corpora Breakdown

Mother Tongue	CEFR Levels	Corpus size	Exam	Age	Years	Exam format
Turkish	A1 – A2	31277	AC	10 –	1993-	- Advice
			BEC1	50+	2012	- Application/response
	B1 – B2	143191	BEC2			- Argumentative/opinion
			BEC3			- Business
			BECH			- Complaint/apology/resp
	C1 – C2	101501	BECP			onse
			BECV			- Composition/essay
	Total	275969	BULA			- Critical
			CAE			- Descriptive/creative/aut
			CPE			obiographical
			FCE			- Formal
			GT			- Informal
			ICFE			- Informative/ news
			IELTS			- Letter/reference
			PET			- Neutral/unmarked
TS			- Note/email/memo			
					- Story	

The above table insinuated the background of the corpus data of different genres produced by the EFL learners in the Cambridge examination overseen from 1993 – 2012.

The present study used the Cambridge Learner Corpus developed determiner typology to find the errors in the determiner usage. Learner errors in CLC (Cambridge Learner Corpus) are tagged through the use of the following convention (CUP, 2014, p. 14-15):

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

Incorrect Determiner Form (FD): this is a situation where indefinite article has been used in a sentence.

Examples

1. Miami is like a dream city . I <#TV> *was* *in* / *went* *to* </#TV><#FD> **a** /DT | **an** /DT </#FD> American Football match yesterday .
2. It's <#FD> **a** /DT | **an** /DT </#FD> old town with a harbour and a big
3. You can take photos with it and it has <#FD> **a** /DT | **an** /DT </#FD><#RP> *Mp* *3* / *MP 3* </#RP> player .

Missing Determiner (MD): this is the process by which a determiner needs to be added to complete the phrase or sentence.

Example,

1. But / *but* </#UP> suddenly it turned into <#MD> / | **a** /DT </#MD> rainy day .
2. I prefer watching <#MD> / | **a** /DT </#MD> comedy film which makes me feel happy .

Replace Determiner (RD): in this sense, a valid determiner has been used but it is not correct in the context.

Example,

1. On your last day you have <#RD> **an** /DT | **the** /DT </#RD> afternoon free .
2. You should take <#RD> **a** /DT | **the** /DT </#RD> guided tour which is available .

Unnecessary Determiner (UD): this is a process whereby a valid determiner has been used but its inclusion is incorrect.

Example,

1. I also want to go to <#UD> **the** /DT | / </#UD> " Universal Studios " .
2. You need to take a right turn on <#UD> **the** /DT | / </#UD> Main Street and walk towards the <#RP> *Public* / *public* </#RP> library .

3.3. Type-token Ratio

Type-token ratio analysis is used to find out the lexical variety of a corpus. Williamson (2013) states that Type-token ratio is a measure of vocabulary variation within a written text or a person speech and it is a helpful measure of lexical variety within a text. ("Type-Token Ratio (TTR)") Moreover, "T/t is obtained by dividing the type count, which is always < 1. A high T/t indicates a high degree of lexical variation while

a low T/t indicates a low degree of lexical variation. (“Trolljäger: Understanding Troll Writing as a Linguistic ...”) In addition, T/t can also be expressed as a percentage, multiplying the ratio by 100” (Thomas, 2015). In other words, Granger (2002) highlights that it is computed by the means of the formula below and the results obtained from the type/token calculation draw conclusion on lexical richness in learner texts. T/t is calculated as shown below.

$$\text{Type-token} = (\text{number of types} / \text{number of tokens}) \times 100$$

3.4. Log-likelihood

Log-likelihood is a test for statistical significance, similar to the chi-squared measure that is often used in corpus analysis such as for collocation or keyword analysis and it is sometimes called G-square or G score (Baker et al, 2006). In a statistical analysis of texts, to test the frequency distributions, LL (Log Likelihood) test is a reliable alternative to Pearsons’ Chi-square (Dunning, 1993). The LL test considers word frequencies weighted over two different corpora. It measures higher or lower frequencies than expected. G2 score or LL is Log-likelihood value is as p-value in Pearsons’ Chi-square (McEnery, Xiao& Tono, 2006).

Log-likelihood and effect size calculator

To use this wizard, type in frequencies for one word and the corpus sizes and press the calculate button.

	Corpus 1	Corpus 2
Frequency of word	<input type="text"/>	<input type="text"/>
Corpus size	<input type="text"/>	<input type="text"/>

Notes:

1. Please enter plain numbers without commas (or other non-numeric characters) as they will confuse the calculator!
2. The LL wizard shows a plus or minus symbol before the log-likelihood value to indicate overuse or underuse respectively in corpus 1 relative to corpus 2.
3. The log-likelihood value itself is always a positive number. However, my script compares relative frequencies between the two corpora in order to insert an indicator for '+' overuse and '-' underuse of corpus 1 relative to corpus 2.

How to calculate log likelihood

Log likelihood is calculated by constructing a contingency table as follows:

	Corpus 1	Corpus 2	Total
Frequency of word	a	b	a+b
Frequency of other words	c-a	d-b	c+d-a-b
Total	c	d	c+d

Note that the value 'c' corresponds to the number of words in corpus one, and 'd' corresponds to the number of words in corpus two (N values). The values 'a' and 'b' are called the observed values (O), whereas we need to calculate the expected values (E) according to the following formula:

$$E_i = \frac{N_i \sum_j O_j}{\sum_i N_i}$$

In our case N1 = c, and N2 = d. So, for this word, E1 = c*(a+b) / (c+d) and E2 = d*(a+b) / (c+d). The calculation for the expected values takes account of the size of the two corpora, so we do not need to normalize the figures before applying the formula. We can then calculate the log-likelihood value according to this formula:

$$-2 \ln \lambda = 2 \sum_i O_i \ln \left(\frac{O_i}{E_i} \right)$$

Figure 4. Screenshot of Log-likelihood Calculator

Table 2.

Contingency Table for Log-likelihood Calculation

	CORPUS ONE	CORPUS TWO	TOTAL
Freq. of words	a	b	a+b
Freq. of other words	c-a	d-b	c+d-a-b
TOTAL			

Adapted from: Rayson and Garside, 2000, p.3

Source: McEnery, Xiao & Tono, 2006

Note that the value ‘c’ corresponds to the number of words in corpus one, and ‘d’ corresponds to the number of words in corpus two (N values). The values ‘a’ and ‘b’ are called the observed values (O). We need to calculate the expected values (E) according to the following formula:

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

In our case $N_1 = c$, and $N_2 = d$. So, for this word, $E_1 = c*(a+b) / (c+d)$ and $E_2 = d*(a+b) / (c+d)$. The calculation for the expected values takes account of the size of the two corpora, so we do not need to normalize the figures before applying the formula. We can then calculate the log-likelihood value according to this formula:

$$-2 \ln \lambda = 2 \sum_i O_i \ln \left(\frac{O_i}{E_i} \right)$$

“This equates to calculating LL as follows: $LL = 2*((a*\log(a/E_1)) + (b*\log(b/E_2))$ ”

Source: McEnery, Xiao & Tono, 2006

The word frequency list is then sorted by the resulting LL values. This gives the effect of placing the largest LL value at the top of the list representing the word which has the most significant relative frequency difference between the two corpora (McEnery, Xiao & Tono, 2006).

3.5. Participants of the Experimental and Control Group

This part of the study is experimental study as Turkish EFL learners were divided into two groups of the same proficiency levels. There are 46 participants in total throughout the study, 30 of them were in the experimental group and 16 were in the control group. The participants are seventh grade students with the age range of 12 and 13. They are not equally divided because of some inconveniences. All the participants were Turkish native speakers, A1-A2 CEFR proficiency levels. They are all learning English as a foreign language for almost the same year and for the same purpose. The groups are heterogeneous in terms of gender, amongst the participants 22 were female and 24 were male seventh grade students.

3.6. Procedure for Using Corpus

A1 - A2 Turkish EFL learners were divided into control and study groups, in order to compare the traditional and corpus-driven ways of teaching English articles.

First of all, each of the groups was given a pre-test which consists of 31 fill-in-the-blanks questions in order to figure out what the learners know about English articles. After the tests results a plan was made for teaching the learners use in corpus driven and traditional method.

In the traditional method, the teacher explained what determiners with examples. The teacher differentiated the types of determiner in order to lay a background about the general determiners and then specify and put more emphasis on the English articles. The teacher explained the differences and usages of definite and indefinite articles with concrete examples.

In a corpus-driven technique, the teacher explained the topic of matter, which was the determiner and its types. To bring authentic language into the classroom the teacher provided the participants with some example data from the learner corpora obtained. In this process, there were some corpus driven activities implemented by the teacher to enhance learning and to make English articles more comprehensive for the learners. The corpus-driven activities are as follows.

i. Fill in the Blanks

In the fill in the blanks activity there were given some words starting with vowels and some starting with consonants, participants had to figure out where to

use indefinite articles (a & an). However, there were some complicating vocabulary which they had to learn their pronunciations before they could answer them, like; honest, honour, hour, united, union, etc. On the other side of the paper there were given activity that dealt with the use of definite articles (the).

ii. Error Correction

In the error correction activity, the participants are presented with some data obtained from the analysed learner corpora. The participants were supposed to correct any determiner (article) errors made by the essay writers (Turkish EFL learners). The participants corrected the errors and constructed their own sentences or paraphrased the given sentence when necessary.

iii. Picture Description

The picture description activity was more like a speaking skills-based activity. Each of the participants was given a picture and they interpreted the picture using and emphasizing the learned determiners (articles). The picture contained items that could be described using articles. For instance, one of the pictures what of a classroom where desks, chairs, markers, board, maps, bookshelf, etc. could be seen.

Excerpt:

- a. The book is on the table.
- b. There is a black marker between the book and the eraser.
- c. I can see an eraser next to the paper clip.

iv. Listening Activity

The listening activity was to make the participants acquainted with how the native speakers use articles in their spoken language. The participants had to listen and write any article they heard from the audio text. At the beginning of the activity the participants were divided into British and American groups. The British listened to a short drama on BBC Radio, while the American group listened to

news highlights on CNN. At the end of the listening activities both groups compared and contrasted what they obtained from the listening texts.

3.7. Duration of the Study

The control and experimental study lasted for four consecutive weeks, each week participants were presented with different approaches and activities to enhance learning of English determiners (article) effectively.



CHAPTER IV

FINDINGS

4. INTRODUCTION

This chapter deals with the results of the English determiners which are analysed through the use of Loglikelihood to describe the statistics. First of all, there shown the descriptive results of English articles used in the various corpus. Secondly, the functions of the English articles used in the corpora are presented and there shown the comparison of the usage. Finally, the overall comparison of determiners across the three proficiency levels is presented.

4.1. Overall Findings of the Study

At the commencement of the analysis, the determiners used in the whole Turkish EFL learners' sub-corpora were gathered and calculated without considering to what determiner category they belong. The English Articles were extracted from the used determiners through sketch engine. Then, the number of English articles (a, an, and the) were spontaneously cast.

In the end, the entirety and the frequency of each determiner were obtained. In order to ascertain the dimension of English articles used in the entire corpora frequency count was conducted. Therefore, the frequency results of each learner proficiency level could be compared with one another. The overall summary of the results of the determiners used in the corpora is shown in the table below.

Table 3.

Corpus Size and English Determiner Frequency

Proficiency Level	Corpus size	Determiners	Percentage
A1 – A2	31277	1321	4.22%
B1 – B2	143191	7632	5.33%
C1 – C2	101501	7025	6.92%
TOTAL	275969	15978	16.47%

The above table shows the description of the total data used and the number of determiners found in the given corpora. As could be seen in the above statistics, the results show that with 6.92% it obvious that C1 – C2 proficiency level use more determiners than the other proficiency levels, then followed by B1 – B2 and at last with 4.22% A1 – A2 level's use of determiner is the least among the proficiency levels.

4.2. Type-token Ratio Analysis Results

Table 4 shows the type-token ratio results of the three proficiency levels, which are A1-A2, B1-B2 and C1-C2. As stated in the definition, Type-token (T/t henceforth) serves to show the lexical variety within a text. T/t ratio of English determiner represents the percentage of each English determiner within all the words in the text.

Table 4.

Comparative Frequency of English Determiners (articles) in the Corpora

Category	Token	Type	T/t Ratio (<=/%)
A1 – A2	1321	312	0.24 (24) %
B1 – B2	7632	541	0.709 (7.09) %
C1 – C2	7025	753	0.107 (11) %

n= raw frequency of epistemic bundles

T/t ratio= Type/token ratio; percentage of number of stance lexical bundles (types) in total of words (tokens) in each corpus

As Table 4 displays, the lowest number of English articles in total belongs to C1-C2 (753). Yet, when the T/t ratio of English articles is compared, it can be seen that the T/t ratio of English articles used in C1-C2 is 0,11 (11%); while it is 0,709 (7.09%) in B1-B2 and 0,24 (24 %) in A1-A2. This means that A1-A2 learners used more varied English articles; 24 of their every 100 words were English articles. C1-C2 learners used 11 of their every 100 words as an English article and B1-B2 learners used 7.09. The percentage of T/t ratios is nearly similar in C1-C2 and B1-B2 corpora, however, it is lower than in the A1-A2 proficiency level.

Frequency differences presented in Table above are all proven to be statistically significant. From an overall view, the table above shows that there are striking differences

between A1-A2, B1-B2 and C1-C2 Turkish learners of English proficiency levels in terms of overall determiner usage.

4.3. Log-likelihood Analysis Findings

Log-likelihood is used to analyse the test of the frequency distribution of the corpora. Log-likelihood is considered as a test which weighs the two different corpora word frequencies. The of English articles obtained from the various data produced by the Turkish EFL learners of different CEFR (Common European Framework) proficiency levels were compared as follows:

1. A1 - A2 vs. B1 - B2
2. A1 - A2 vs. C1 - C2
3. B1 - B2 vs. C1 - C2

Moreover, in order to see the condition among the proficiency levels, overall determiner frequency of A1 - C2 CEFR levels of proficiency were also compared.

The first Log-likelihood analysis was done between A1 - A2 and B1 - B2 corpora to spot the overall significant frequency difference that transpired between the given proficiency levels. The table below shows the Log-likelihood ratio of the determiner (English articles) usage in comparison with A1 - A2 and B1 - B2 CEFR levels of proficiency.

Table 5.

Log-likelihood Ratio of A1-A2 and B1-B2

Word	Freq. in A1 – A2	Freq. in B1 – B2	Log-likelihood	Sig.		
MD	227	313	169,19	0,000	***	+
a	93	106	86,57	0,000	***	+
an	9	6	13,12	0,000	***	+
the	125	201	75,08	0,000	***	+
UD	70	108	44,73	0,000	***	+
a	36	41	33,54	0,000	***	+
an	3	1	6,21	0,013	*	+
the	31	66	11,10	0,001	***	+
FD	4	5	3,36	0,067		+
a		1	#NUM!	#NUM!	##	-
an	4	4	4,24	0,039	*	+
RD	11	25	3,38	0,066		+
a	8	15	3,71	0,054		+
an		1	#NUM!	#NUM!	##	-
the	3	9	0,37	0,541		+

+ indicates overuse in A1-A2 relative to B1-B2,

- indicates underuse in A1-A2 relative to B1-B2.

As illustrated in the table above, the determiner frequency in B1 – B2 corpora is far more than those of A1 – A2 levels of proficiency. As can be seen from the table, almost all the determiners are been overused across the error forms, apart from FD ‘a’ and RD ‘an.’ Moreover, when attention is paid to the above table, it can be seen that MD (Missing Determiner) has the highest frequency in both corpora, followed by UD (Unnecessary Determiner), and at last, FD followed afterwards as the item with lowest frequency of determiner in the used corpora.

Table 6.

Log-likelihood Ratio of A1-A2 and C1-C2

Word	Freq. in A1 – A2	Freq. in C1 – C2	Log-likelihood	Sig.	
MD	227	445	35,88	0,000	*** +
a	93	140	30,67	0,000	*** +
an	9	13	3,24	0,072	+ +
the	125	292	9,03	0,003	** +
UD	70	217	0,11	0,740	+ +
a	36	59	9,72	0,002	** +
an	3	9	0,01	0,907	+ +
the	31	149	4,30	0,038	* -
FD	4	13	0,00	0,998	- -
a		6	#NUM!	#NUM!	## -
an	4	7	0,91	0,341	+ +
RD	11	78	7,13	0,008	** -
a	8	34	0,50	0,481	- -
an		10	#NUM!	#NUM!	## -
the	3	34	6,12	0,013	* -

+ indicates overuse in A1-A2 relative to C1-C2,
- indicates underuse in A1-A2 relative to C1-C2.

As anticipated or predicted from the frequency differences between A1- A2 and C1 - C2, Loglikelihood values between the corpora of the two proficiency levels revealed the overuse and underuse of the determiners. A1-A2 learners used 312 determiners (Articles), meanwhile, C1-C2 used the total number of 753 determiners (articles) in the whole analysed corpora.

The loglikelihood ratio between the two corpora is of two groups, one-half of the values are found to be overused, and the other half are found to be underused. In the log-likelihood analysis, the two error forms with higher frequencies (MD and UD) tend to be overused by the learners, while the remaining two (FD and RD) are found underused in the A1-A2 and C1-C2 CEFR proficiency levels.

Table 7.

Log-likelihood Ratio of B1-B2 and C1-C2

Word	Freq. in B1 – B2	Freq. in C1 – C2	Log-likelihood	Sig.		
MD	313	445	90,86	0,000	***	-
a	106	140	23,66	0,000	***	-
an	6	13	5,61	0,018	*	-
the	201	292	62,73	0,000	***	-
UD	108	217	84,36	0,000	***	-
a	41	59	12,40	0,000	***	-
an	1	9	10,41	0,001	**	-
the	66	149	67,78	0,000	***	-
FD	5	13	6,97	0,008	**	-
a	1	6	5,89	0,015	*	-
an	4	7	2,19	0,139		-
RD	25	78	49,90	0,000	***	-
a	15	34	15,55	0,000	***	-
an	1	10	11,97	0,001	***	-
the	9	34	25,36	0,000	***	-

+ indicates overuse in B1-B2 relative to C1-C2,

- indicates underuse in B1-B2 relative to C1-C2.

As illustrated in the table above, compared to the previous tables, the frequencies of both groups are relatively close to each other. In the table, all the values included the relative frequency of determiners in the analysed corpora. The loglikelihood ratio of B1-B2 and C1-C2 shows that the Turkish EFL learners of these two proficiency levels underused the determiners. The results show obtained through the LL (Log Likelihood) ratio show that all the articles are underused despite the high frequency in B1-B2 and C1-C2 corpora.

4.4. Traditional and Corpus-driven Techniques Findings

Before implementation of corpus-driven and traditional method in the teaching of English determiners (articles) to Turkish EFL learners, a pre-test was conducted to see what the participants know. After the who study, a post-test was done to see the

participants' improvement and to see if using corpus-driven activities brings change or rather helps the participants learn better and in a more comprehensive way. The tables below show the results of the pre-test and post-test in the control group.

Table 8.

Experimental Group Participants' Test Scores

Participants	Pre-test	Post-test
Study 1	24	30
Study 2	24	31
Study 3	23	31
Study 4	22	29
Study 5	22	28
Study 6	19	27
Study 7	18	29
Study 8	17	28
Study 9	17	27
Study 10	17	27
Study 11	17	24
Study 12	16	20
Study 13	14	26
Study 14	13	25
Study 15	13	26
Study 16	12	27
Study 17	25	30
Study 18	23	31
Study 19	22	29
Study 20	23	30
Study 21	21	29
Study 22	21	30
Study 23	19	28
Study 24	19	26
Study 25	19	27
Study 26	18	29
Study 27	17	28
Study 28	17	27
Study 29	16	26
Study 30	15	29

The above table shows the scores obtained by the participants in pre-test and post-test. Mere looking, it could be seen that there is massive improvement in their post-test

scores in the study group (corpus-driven). There is some significant increase in the post-test, some of the participants rose from 12 points to 27, which is very impressive, meanwhile there are some participants with a slight improvement like 16 – 20 and 22 – 28.

Table 9.

Control Group Participants' Test Scores

Participants	Pre-test	Post-test
Control 1	23	26
Control 2	20	24
Control 3	19	22
Control 4	17	24
Control 5	17	19
Control 6	16	21
Control 7	16	20
Control 8	16	21
Control 9	14	19
Control 10	13	22
Control 11	17	21
Control 12	11	19
Control 13	22	25
Control 14	22	26
Control 15	16	20
Control 16	15	18

The above table shows the results of the control who were taught using the traditional method of teaching English articles. The table infers that there is some improvement, however, the improvements were not quite applaudable as the ones in the study group. The highest difference is 9, in which the participants scored 13 in the pre-test and 22 in the post-test. Moreover, the lowest difference was 2 whereby the participants scored 17 in the pre-test and 19 in the post-test.

Table 10.

Descriptive Results of Pre-test and Post-test

		Description of Pre-test and Post-test							
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Pre-Test	Experimental Group	30	18,77	3,588	0,655	17,43	20,11	12	25
	Control Group	16	17,13	3,344	0,836	15,34	18,91	11	23
	Total	46	18,20	3,557	0,524	17,14	19,25	11	25
Post-Test	Experimental Group	30	27,80	2,340	0,427	26,93	28,67	20	31
	Control Group	16	21,69	2,600	0,650	20,30	23,07	18	26
	Total	46	25,67	3,801	0,560	24,55	26,80	18	31

The above table shows the description of the results shown on SPSS using independent variable analysis. The result shows that the minimum score of the participants in the experimental group was 12 in the pre-test and 20 in the post-test. However, the result shows different in the control group, where 11 was the minimum score in the pre-test before teaching the English determiner (articles) and the result changed in the post-test by which the minimum score increased to 18.

Moreover, to talk about the maximum scores, the experimental group had 25 in the pre-test and emerged with 31 in the post-test. Meanwhile, the same improvement could not be said in the case of the control group, where the score changed from 23 in the pre-test to 26 in the post-test. There is an improvement but not as significant as the experimental group.

Table 11.

ANOVA Results

		ANOVA				
		Sum of Squares	df	Mean Square	F	Sig.
Pre-Test	Between Groups	28,122	1	28,122	2,287	0,138
	Within Groups	541,117	44	12,298		
	Total	569,239	45			
Post-Test	Between Groups	389,871	1	389,871	65,918	0,000
	Within Groups	260,238	44	5,914		
	Total	650,109	45			

When the ANOVA was ran on SPSS, the above results were obtained. The table above shows the noteworthy results of the tests compared between the groups. In the pre-test of the experimental and control group there is a significant difference between the groups as the result shows '0.137.' On the other hand, in the results of the post-test, the experimental and control group suggests that there is a significant difference in the test scores. The significant score of '0.000' shows that there is an influence between the groups, which implies that the experimental group post-test results present the role of using corpus-driven technique (activities) in learning grammatical item better in a foreign language classroom.

CHAPTER V

DISCUSSIONS

5. INTRODUCTION

This chapter summarizes the findings of the study in brief. After the summary there follows the review of the research questions, the limitation of the study. Finally, some recommendations for further studies have been given.

5.1. Summary of the Study

English determiners play a significant role in language use and language teaching and learning. They are unavoidable elements of language use. The linguistic and grammatical fluency and accuracy also depend on the use of the language in a correct way, and that cannot be done without the acquisition/learning the English determiners. They show success and progress toward target language learning. In other words, English determiners are considered to be the most difficult grammatical structure to be acquired/learned by the Turkish EFL learners due to the absence of the indefinite article in their L1.

In spite of the role English Determiners play in the spoken and written, academic prose and scientific discourse, etc., there are very few studies conducted on the English articles, there are, in fact, very few studies done comparing the use of determiner (English articles) by native and native, or native and non-native speaker or learner. However, the present study aimed at finding answers to the following research questions:

5.1.1. Research Question 1:

What are the most common errors pertaining to the use of English Determiners along with the proficiency range A1 – C2 according to CEFR?

The main aim of the study was to determine which English determiners are used by the Turkish EFL learners and to determine which of the determiners were used more frequently. The results obtained from the analysis revealed that 15970 determiners were used and among them, by considering the whole proficiency level 1516 English articles (a, an, and the) were used by the learners. Moreover, when the frequency was compared

it was seen that C1-C2 proficiency levels used higher with the frequency of 753 in the 101501 word-count. Furthermore, the proficiency level with lowest determiner usage was A1-A2 with 312 in 1321 word-count. The results obtained from the present study show that Turkish EFL learners' 275969 corpus size has 16,47% used of determiners when it is considered overall. In the 16,47% determiner use, C1-C2 has 6,92%, B1-B2 5,33% and A1-A2 has 4,22% use of English determiners in general.

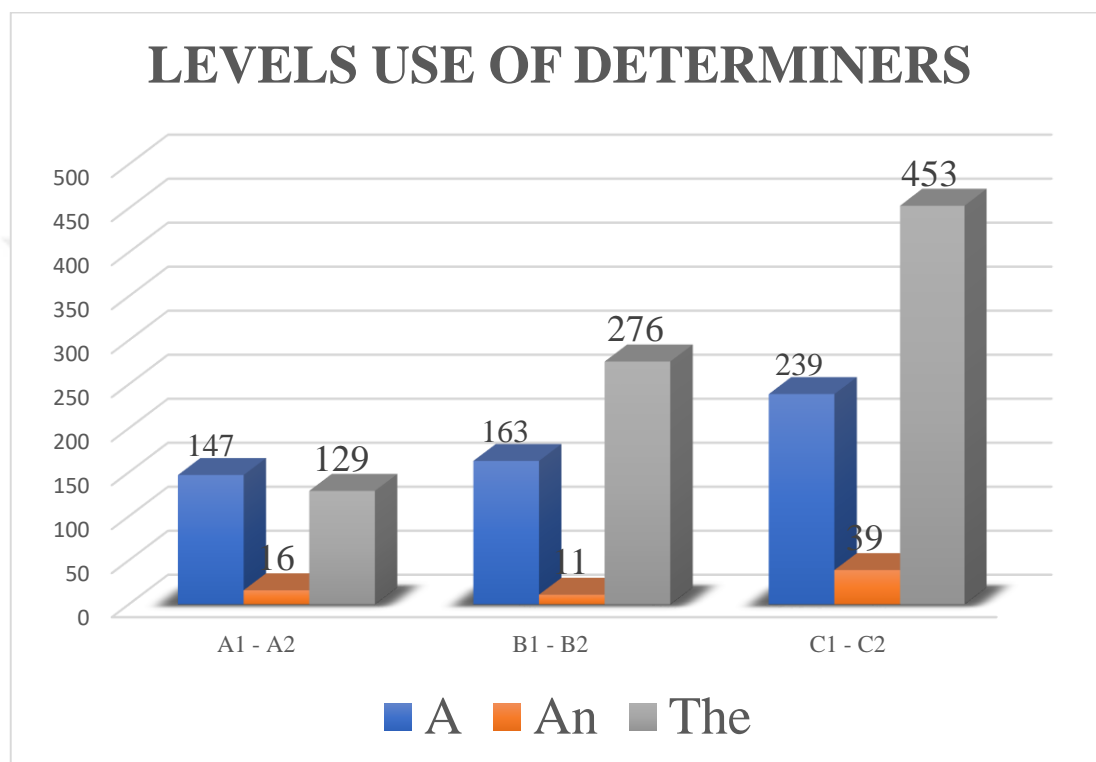


Figure 5. English Determiner (articles) Frequency among the Levels

Hence, it can be concluded that the number of determiners used in the C1-C2 proficiency levels is higher and compared to all proficiency levels. A1-A2 the use of English determiners is lower, nevertheless, notwithstanding the higher and lower frequencies in both levels, it is found that B1-B2 and C1-C2 used "the" more than the other English determiners (Articles). Meanwhile, in the A1-A2 proficiency level, the use of "a" is higher than "an" and "the". Moreover, the use of "an" is common on all levels.

5.1.2. Research Question 2:

Which determiners are over-/under used by the Turkish EFL learners in various levels of proficiency range?

When the log-likelihood ratio was used to analyse the overuse and underuse of the English determiners by Turkish EFL learners, the analysis was divided into three distinct groups to compare the use of determiners. The comparison was as follows.

- i. A1-A2 vs B1-B2
- ii. A1-A2 vs C1-C2
- iii. B1-B2 vs C1-C2

The first comparison between A1-A2 and B1-B2, it is found that almost all the determiners were overused by the learners except FD "a" and RD "an" in the Missing Determiner (MD) all of the determiners were overused and in use of Unnecessary Determiners (UD), all of them were overused including "an", only in the incorrect determiner form (FD) "an" and the replace determiner (a) are underused.

In the second comparison between A1-A2 and C1-C2, it is found that in the Missing Determiners (MD) all of the determiners (a, an, and the) were overused as well as the ones in the unnecessary determiners (UD) except the definite article "the" is being underused by the learners. Meanwhile, in the incorrect determiner form (FD) "a" is underused and "an" is overused by the Turkish EFL learners. Moreover, in the replace determiner (RD) all the English determiners are being underused.

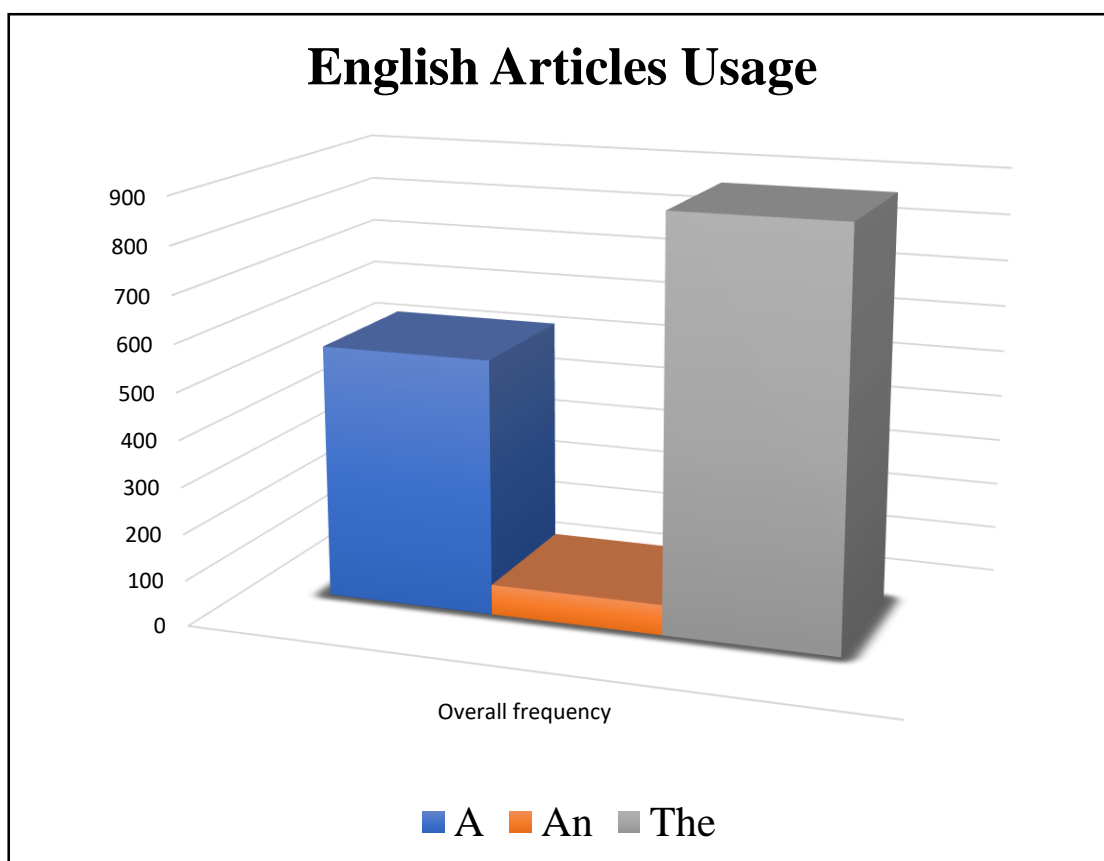


Figure 6. Overall Frequency of English Determiners (articles)

Moreover, the LL ratio of B1-B2 and C1-C2 revealed that both the two proficiency levels underused all the English determiners (Articles) despite the fact that they belong to the different error categories.

Finally, mere looking at the above statements, it can be seen that in some cases English determiners were overused e.g., in the comparison between A1-A2 and B1-B2, and in the comparison between A1-A2 and C1-C2, there are overuse and underuse in different error forms. And all the English determiners were underused in the comparison between B1-B2 and C1-C2 proficiency levels.

5.1.3. Research Question 3:

What are the persistent errors of Turkish EFL learners in their use of English determiners along with their proficiency levels?

In the case of the persistent errors, the Missing determiners form (MD) "a" and "the" have the higher frequency then followed by unnecessary determiners (UD) which likewise has the highest frequency in "a" and "the". Although, when "a" and "the" are compared, "the" has the highest frequency, which means to tell that "the" is persistent among the English determiners (Articles) used in the whole corpora.

Figure 7. Percentages of Determiner Error Form

The chart above suggests that the percentage of missing determiner is apparently higher according to the data obtained, thus, this means that MD is persistent among the error from due to its high percentage of usage, followed by UD which has 26.06 percentage among the articles used in the Turkish EFL learners' sub-corpora.

Table 12.

Frequency in Missing Determiner Error Typology

Error typology		A1-A2	B1-B2	C1-C2	TOTAL
Missing	Determiner (MD)	227	313	445	987

Table 8 displays the overall missing determiner error typology found among the groups. When we refer to the table, it can be seen that the C1-C2 proficiency level has the highest frequency compared to the other proficiency level. Examples of missing determiners extracted from the corpus.

A1-A2

1. Yesterday I went to <#MD> / | a /DT </#MD> big Shopping <#SA> Center | Centre </#SA>

2. We must bring our crayons , <#MD> | *a* </#MD> pencil
, <#MD> | | **an** /DT </#MD> eraser and <#MQ> | *some* </#MQ> paper
with us.
3. I went to <#MD> | | **the** /DT </#MD> maritime museum, it's very interesting.

B1-B2

1. As you know , a hotel <#UN> *whose name* |
</#UN><#RV> *is* | *called* </#RV><#MD> | | **the** /DT </#MD> Palace Hotel
has been booked for your group .
2. But if it is night
or <#MD> | | **a** /DT </#MD><#RP> *sunday* | *Sunday* </#RP><#MP> | , </#M
P> unfortunately the nearest shop is always closed and not only
3. I 've bought <#MD> | | **an** /DT </#MD> " X-Box- 360 " for my bedroom.

C1-C2

1. Especially when we asked him if he could turn the air-
conditioning <#W><#MD> | | **a** /DT *little bit down* | *down a little bit* .
2. As result of being <#MD> | | **an** /DT agricultural country , we
are <#RJ><#DJ> *luck* | *lucky* | *able* to buy fresh vegetable from markets .
3. We would like to start with <#MD> | *a* warm welcome to all <#MD> | | **the**
/DT new students from all over the world .
- 4.

Table 13.

Frequency in Unnecessary Determiner (UD) Error Typology

Error typology		A1-A2	B1-B2	C1-C2	TOTAL
Unnecessary (UD)	Determiner	70	108	217	395

The table above displays the overall unnecessary determiner usage found among the groups. When we look at the table, it can be seen that the C1-C2 proficiency level has the highest frequency compared to the other proficiency level. Examples of unnecessary determiners extracted from the corpus;

A1-A2

1. It </#RP><#RV> *has* | *plays* </#RV><#UD> **a** /DT | / </#UD> fantastic music .
2. I think it will take one and <#MD>
| *a* </#MD> half <#UD> **an** /DT | / </#UD><#AGN> *hour* | *hours* </#AGN> .
3. <#MD> **Dinner** /NN | **The** /DT **dinner** /NN </#MD><#TV> *is* | *will*
be </#TV> on Saturday at 6.30 .

B1-B2

1. As you know , it is not suitable to wear <#UD> **a** /DT | / </#UD><#IN> *jean* | *jeans* </#IN> or something <#MJ> | *similar* </#MJ>
2. Then I decided to read a book <#RT> *under* | *by* </#RT><#UD> **the** /DT | / </#UD><#RP> *candle*
light | *candlelight* </#RP> .
3. Being able to obtain <#UD> **an** /DT | / </#UD> information is quicker <#MP> | , </#MP> which speeds up the <#RN> *process* | *pace* </#RN> of work .

C1-C2

1. One winter day , as Lucy desperately tried to forget what she was enduring , the sound of <#UD> **a** /DT | / soft music attracted her to the window .
2. I feel that you should take <#UD> **an** /DT | / immediate action <#RT> *to* /
on this matter and I think you should compensate <#MA> / *us* and refund the money <#UY> *back* / as soon as possible .
3. <#UD> **The** /DT **accommodation** /NN | **Accommodation** /NN in the hall <#MV> | *is* much cheaper than anywhere else .

Table 14.

Frequency in Incorrect Determiner Form (FD) Error Typology

Error typology	A1-A2	B1-B2	C1-C2	TOTAL
Incorrect Determiner Form (FD)	4	5	13	22

Table 12 shows the overall incorrect determiner form error typology found among the groups. When we put the table above into consideration, it can be seen that the frequency is low across proficiency levels compared to the other error typologies. Examples of incorrect determiner form extracted from the corpus.

A1-A2

1. I <#TV> *was in* / *went to* </#TV><#FD> *a* /DT | *an* /DT </#FD> American Football match yesterday .
2. It's <#FD> *a* /DT | *an* /DT </#FD> old town with a harbour and a big <#DN> *shopping mile* | </#DN> .
3. You can take photos with it and it has <#FD> *a* /DT | *an* /DT </#FD><#RP> *Mp 3* | *MP 3* </#RP> player . #FD

B1-B2

1. I am really sorry <#MP> | , </#MP> I will not be able to come for a meal in your new house this time , because I just had <#FD> *a* /DT | *an* /DT </#FD> interview <#RT> *about* | *for* </#RT><#RD> *my* | *a* </#RD> job today .
2. ...20 who would like to improve their English and <#MA> | *it* </#MA><#RV> *is* | *provides* </#RV><#FD> *an* /DT | *a* /DT </#FD> unique opportunity for us to...
3. When </#RP> you changed your distributor all <#MD> | *the* </#MD> packages <#RV> *were* | *started to get* </#RV> damaged . Please send us <#FD> *a* /DT | *an* /DT </#FD> up-to- date catalogue .

C1-C2

1. Also the group <#MV> | *called* RADIANT did not play for only 35 minutes <#RP> , | ; they played for <#FD> **an /DT | a /DT** <#SX> *hole | whole* hour and it was really great .
2. *we are* going to have <#FD> **an /DT | a /DT** <#DJ> *sport | sports* event in July .
3. <#MD> *Lack | A lack* of love , <#FV> *to raise | raising* children in <#FD> **an /DT | a /DT** hostile environment , being oblivious to their emotional and psychological needs ...

Table 15.

Frequency in Replace Determiner (RD) Error Typology

Error typology	A1-A2	B1-B2	C1-C2	TOTAL
Replace Determiner (RD)	11	25	78	114

Table 13 displays the overall replace determiner error typology found among the groups. When we refer to the table, it can be seen that the C1-C2 proficiency level has the highest frequency compared to the other proficiency level. Examples of replace determiners extracted from the corpus.

A1-A2

1. <#RD> **My /PP\$ | The /DT** <#RD> window is broken in the bathroom .
2. Yesterday I went to <#RD> **some /DT | a /DT** <#RD> museum <#MP> *it* | .
3. Hello Sue I
have <#IV> *wroten | written* <#IV> <#RD> **that /IN/that | a /DT** <#RD> <#RP> *post card | postcard* <#RP> from Turkey .

B1-B2

1. She dashed <#RT> *into | onto* <#RT> the train and took a seat next to <#RD> **the /DT | an /DT** <#RD> old lady .
2. <#UN> **Mr / Mrs** | <#UN> Dear friends <#MP> | , <#MP> thank you for <#MD> *the* <#MD> invitation <#MV> | *to*

come </#MV><#RT> *about* / *for* </#RT><#RD> **the** /DT | **a** /DT </#RD> meal .

3. They are all <#MV> / *linked* </#MV> together , we cannot <#RV> *cut* / *break* </#RV><#RD> **their** /PP\$ | **the** /DT </#RD> relationship <#S> *betwen* / *between* </#S><#RA> *each other* / *them* </#RA> .

C1-C2

1. Please , read this sheet carefully because it will give you some <#UN> *piece of* / advice you might find <#UA> *it* / useful <#RT> *for* / *in* <#RD> **your** /PP\$ | **the** /DT near future , in our country.
2. were only <#RD> **the** /DT | **a** /DT few <#M> / *of the things* <#R> *to start with* / .
3. <#RD> **The** /DT | **An** /DT old lady was sitting in her rocking chair and looking out of the window .

Finally, the above data suggest that among the determiner error typologies Missing Determiner is more persistent in all the proficiency levels mere looking at the frequency and the usage.

5.1.4. Research Question 4:

Do the errors relate with the use of determiners stem from Turkish L1 or imply an interlanguage property?

The above research question dwelled on the implication of interlanguage properties. In order to answer the given research question, the frequency of English determiners was identified in the corpora, then compared with each proficiency level. The results obtained from the corpora were evaluated by looking into the frequency of the English determiners (articles) used by the learners from each group. Moreover, the data and the number of the English determiners used of all the levels were compared through the use of log-likelihood analysis.

In the analysis, it was found that C1-C2 proficiency levels used more English determiners compared to the other levels. According to the frequency obtained and loglikelihood analysis, there are certain overuse and underuse of English determiners in

the Turkish EFL learners corpora when compared with the various CEFR proficiency levels (A1-C2).

It can be seen that the tendency of overuse and underuse of English determiners by the Turkish EFL learners from different CEFR proficiency levels could be related or could be resulted due to the interlanguage properties. English determiner (definite article) is not present in the Turkish language: therefore, it could be presumed that the overuse and the underuse of the definite article have resulted from the learners' mother tongue (L1 transfer).

It was also found that the students' native language has interfered in their use of English articles. This may be due to the fact that the students first acquired the rules of article sequences which are acceptable in their native language when they are used in the modification of a noun. In the process of repositioning or reordering the articles, the students face difficulties in knowing the correct order of English articles, especially when it comes to the definite articles. Thus, the learners' mother tongue might be the reason of the errors in the learner language production, the mother tongue might cause overgeneralization of the articles, in which "the learner extends the use of a grammatical rule beyond its accepted uses" (Richards et al., 1989).

These findings go in hand with Bhela's (1999), who affirmed that the way learners used their L1 properties to help them form their L2 texts designates a direct interference of L1 in L2. Prins (2006) highlighted that Students fall back on the rules of their first language when they do not know the rules of the second language. These errors might reflect poor academic performance amongst the students.

According to the result obtained in the present study, the L2 article preference was mainly influenced by L1 due to the absence of definite articles in the mother tongue. Referring to the results of the study carried out by Ionin et al. (n.d.) on the use of articles among the adult L1 Russian and L1 Spanish English learners, they found out that there are three factors which are related to English article acquisition. These are L1-transfer, input, and Universal Grammar. Speakers of Spanish, coming from an L1 with articles, are able to transfer article semantics from their L1 to their L2. Speakers of Russian, which lack articles, are not able to do this, so their learning proceeds through a combination of UG-access and input processing. UG tells learners what possible patterns of article choice in a natural language may look like. Learners fluctuate between these patterns – definiteness and specificity – until the input tells them that only one of these patterns is appropriate for English (p. 21).

To add more information about research carried out to find out about article usage, Ekiert & Han, (2017) carried out a study of the case of L2 acquisition of English articles by Slavic speakers, in which data from sixty-five adult ESL learners were collected, among the participant were 42 Polish, 12 Ukrainian and 11 Russian. The study focused on the acquisition of English articles among the L2 learners with Slavic background, and the multifunctional use of ‘*a*’ and ‘*the*’ were examined. Therefore, the study concluded that “the three types of *the* and two types of ‘*a*’ in questions posed unequal challenges to the participants, with referential use of *a* and the situational use of *the* being the most, and textual use of *the* least, difficult” (p. 165), according to the three different collected from the students’ tasks. The researchers added that “the acquisition of English article by L2 learners whose L1s do not utilize article-like forms requires not only the learning of a new form, but also of a new grammatical meaning.

5.1.5. Research Question 5:

Do corpus-driven techniques assist the learners in their acquisition of determiners?

To answer the above position, experimental and control groups were formed, experimental group participants were taught English determiners (articles) through the use of corpus driven techniques. Meanwhile, the control group participants were taught English determiners (articles) by the use of traditional methods of teaching articles.

Results obtained from the pre-test and post-test implemented before and after the study processes show that there is a significant difference between the groups, which suggests that using corpus-driven techniques in the language classroom assist the Turkish learners to learn English determiners (articles) better.

As predicted, the use of corpus driven techniques in teaching English determiners (articles), the result proved that the traditional methods have lower influence in the acquisition/learning of English determiner (articles) than the corpus driven techniques (activities) to be precise.

The traditional teaching method does not give the language learners a room to practice what is learned in the classroom conveniently, it does not give a language learner a platform to practice the target language beyond the classroom language context. While corpus-driven techniques give a chance to bring the outside language into classroom. The

corpus-driven techniques open the gate to the language learners to practice more with the given sets of rules in real-life without any boundary.



CHAPTER VI

CONCLUSION

6. INTRODUCTION

This chapter dwells on the conclusion of the results found in the study and the presents the implications of the study in English Language Teaching and last it ends with the limitations of the study and recommendations for further studies.

6.1. Conclusion

The aim of the study was to analyse, comprehend and describe how Turkish EFL learners acquire/learn and use English determiners, more specifically English articles (a, an, and the). As determiners are defined as the words that accompany nouns or as the noun modifiers which usually appear before noun phrases to specify or to serve as a reference of a noun. "The term article is used for the special category of determinatives that provide the most basic expression of definiteness and indefiniteness." (McGregory, 1997).

Learning/acquiring English articles is proven to be one of the most difficult grammatical items and they are among the ones to be fully acquired or learner last, for instance, researchers like Master (1987) highlighted that "unlike the acquisition of English article by L1 learners, the English article system is one of the most difficult of aspects of English grammar for ELLs [English Language Learners] and one of the last to be fully acquired." and in terms of error usage of the acquired/learned language, English articles are found more likely to be overused by the second language learners of English, for instance, Cook (1997) laid a background that "the most persistent recurrent problems in terms of slips and minor errors amongst all groups of the most sophisticated users of English as a second language involve confusion with the words the, a, (an) and their proper necessary non-use."

Moreover, in the present study, the sub-corpus of the Turkish EFL learners from the A1-C2 CEFR proficiency levels was analysed in order to find out about the English determiners (articles) used in the text collected from 1993-2012. The study has indicated that one of the most common errors found in the Turkish EFL learners' corpora has been proven to be the use of the definite article (the). The English article (the) has been

overused in all the proficiency levels compared to other English articles (a, an), and the article (an) has been underused within all the proficiency levels. Additionally, the results have revealed that C1-C2 proficiency levels have overused more English determiners (articles) than the other levels and there was higher underuse in B1 - B2 proficiency levels.

Furthermore, among the determiner typologies [Missing Determiner (MD), Unnecessary Determiner (UD), Incorrect Determiner Form (FD), and Replace Determiner (RD)] used in the Turkish EFL learners' sub-corpora, the study has indicated that the learners used MD more persistent, followed by UD and RD, and the least and underused error typology found in the study was FD, the results signified that the learners tended to avoid the used of English determiners (articles) most of the times.

Notwithstanding, the application of corpus-driven techniques was a positive achievement, it was realized that the corpus-driven activities have extremely helped the students comprehend and use English determiners (articles) better and more conveniently. The huge differences found when traditional method is compared and contrasted to corpus-driven techniques in teaching English determiners (articles) to Turkish EFL learners suggest that the implementing corpus-driven activities in the foreign language classroom could be a tremendous success. Indeed, providing real-life data in a language classroom could help foreign language learners get acquainted with how the target language is used in the outside world (outside classroom), especially in a country where there is not enough input of the target language.

Finally, the overuse and underuse of the English determiners (articles) by the Turkish EFL learners across the CEFR proficiency levels could be resulted from the L1 influence due to the lack of definite article in the language. The absence of definite articles leads the learners to overgeneralize and avoid the use of articles during language production (written or spoken). Thus, mother tongue might be the reason for the errors in the learner language production and it leads to the overuse and underuse of a grammatical item.

6.2. Implications of the Study in English Language Teaching

Master (1987), states that “unlike the acquisition of English articles by L1 learners, the English article system is one of the most difficult of aspects of English grammar for ELLs and one of the last to be fully acquired.” Moreover, the role of transfer

in the use of English articles by Moroccan learner is been investigated by (Trimasse, 2016), in which 80 university students' essays were studied, among them 40 from the first year and 40 from the third year. The study aimed to investigate these three items: 1) the effect of proficiency level on transfer errors by Moroccan learners in the use of articles, 2) different between transfer errors made in comprehension and in production in using articles and 3) the difference in transfer errors according to the three English articles. In the meantime, the result of the study concluded that “proficiency level has an effect on students’ transfer errors in comprehension but not in production. There is an effect of task type on transfer errors in that students make more transfer errors in the grammaticality judgment test than in the essay. There is a difference in students’ performance according to the three categories of articles. Students make most errors in using ‘the’ and then zero article” (Trimasse, 2016).

Thus, this study has revealed that Turkish EFL learners encounter some difficulties using the English determiners (articles), the reason for those difficulties could be due to the absence of the definite articles in their mother language. With a corpus-based application, learners could be provided with authentic language directly and they can individually examine the use of English determiner (articles) in context. With the use of corpora, learners could also learn position choices of native speakers with the appropriate place to use English determiners (articles). Moreover, with corpus research, the most appropriate way of teaching and acquiring/learning how to use English determiners.

Corpus linguistics is applied in many fields, such as analysing spoken data, grammar, applied linguistics, culture, sociolinguistics, semantics, psycholinguistics, pragmatics, and lexical studies. Moreover, corpus linguistics provides a room for the use of data-driven approach in language teaching and learning, whereby corpus data is vastly in use to form a new way of language teaching, where learning is driven through the use of the provided corpus data.

6.3. Recommendation for Further Study

The current study is narrowed to the size of the A1 - C2 CEFR proficiency levels Turkish EFL learners' corpora, therefore the results of the study are bounded to the analysis of them. In addition, the corpus-based study has been conducted based on the Turkish EFL learners’ use of English determiners (articles), and the study has been carried

out by restricting its range to overuse and underuse of English determiners (articles) by Turkish EFL learners by looking into four different error forms, these are Missing Determiner (MD), Unnecessary Determiner (UD), Replace Determiner (RD), and Incorrect Determiner Form (FD). The corpus-driven techniques were implemented on secondary school A1 – A2 CEFR proficiency level Turkish EFL learners, to improve their use of English determiners (articles). In the further study, corpus-driven techniques could be implemented in a language classroom with adult Turkish EFL learners.



REFERENCE

- Abe, M. (2007). "A Corpus-Based Investigation of Errors across Proficiency Levels in L2 Spoken Production", *JACET Journal*, volume 44, 2007, pp. 1-14.
- Aijmer, K. (2009). *Corpus and Context. Investigating Pragmatic Functions in Spoken Discourse*. *International Journal of Corpus Linguistics*, 14(3), 419–425. <https://doi.org/10.1075/ijcl.14.3.06ajj>
- Akbana, Y. E. (2018). *A Contrastive Interlanguage Analysis of the Highest-Frequency Vocabulary in Advanced and Native English*, (January).
- Al-khresheh, M. (2011). *An investigation of interlingual interference in the use of 'and' as a syntactic coordinating structure by Jordanian EFL*
- Al-khresheh, M. (2013). *The Misuse of Word Order in the Writing of Jordanian EFL Learners*. Unpublished doctoral dissertation, University of Malaya, Kuala Lumpur
- Al-khresheh, M. (2015). *A Review Study of Interlanguage Theory*. *International Journal of Applied Linguistics & English Literature*, 4 (3), 124 - 131.
- Alimi, M. M. (2007). *English Articles and Modals in the Writing of Some Batswana Students*. *Language, Culture and Curriculum*, 209–222. <https://doi.org/10.2167/lcc336.0>
- Al-saaidi, Sawsan. (2011). *English Definite Article "the " in Term of Definiteness With Reference to Literary Works: A Linguistic Study*.
- Altenberg, B., & Granger, S. (2002a). *The grammatical and lexical patterning of make in native and non-native student writing*. *Applied Linguistics*, 22, 173–194.
- Altenberg, B., & Granger, S. (2002b). *Recent trends in cross-linguistic lexical studies*. In B. Altenberg & S. Granger (Eds.), *Lexis in contrast* (pp. 1–48). Amsterdam: John Benjamins.
- Baker, P., Hardie A. and McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh University Press.
- Bardovi-Harlig, K., & Bofman, T. (1989). *Attainment of Syntactic and Morphological Accuracy by Advanced Language Learners*. *Studies in Second Language Acquisition*, 11, 17-34. <https://doi.org/10.1017/S0272263100007816>
- Barlow, Michael (1996), *Corpora for Theory and Practice*. In: *International Journal of Corpus Linguistics* 1(1), 1–37.
- Barnbrook, Geoff (1996), *Language and Computers. A Practical Introduction to the*

- Computer Analysis of Language. Edinburgh: Edinburgh University Press.
- Belz, Julie and Vyatkina, Nina (2005), "Learner corpus research and the development of L2 pragmatic competence in networked intercultural language study: The case of German modal particles", *Canadian Modern Language Review*, 62(1), 17-48.
- Bernard, S., (2014). *Contrastive Analysis*, 63(5), 250–257.
- Bhela, B. (1999). Native language interference in learning a second language: Exploratory case studies of native language interference with target language usage. *International Education Journal*, 1(1), pp 22-23.
- Biber, D. (2009). 'A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing', *International Journal of Corpus Linguistics* 14 (3): 275–311.
- Bickerton, D. 1981. *Roots of Language*. Ann Arbor: Karoma.
- Braun, S. (2005) 'From Pedagogically Relevant Corpora to Authentic Language Learning Contents', *ReCALL* 17(1): 47–64.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Can, C. (2017). A Learner Corpus-Based Study on Verb Errors of Turkish EFL Learners. *Journal of Education and Training Studies*, 5(9), 167. <https://doi.org/10.11114/jets.v5i9.2612>
- Can, C. (2018). Agreement Errors in Learner Corpora across CEFR: A Computer-Aided Error Analysis of Greek and Turkish EFL Learners Written Productions, 6(5), 77–84. <https://doi.org/10.11114/jets.v6i5.3064>
- Cheng, W. (2012). *Exploring corpus linguistics: language in action*.
- Claudio Fantinuoli and Federico Zanettin (ed.). 2015. *New directions in corpus-based translation studies (Translation and Multilingual Natural Language Processing 1)*. Berlin: Language Science Press
- concerning error analysis. *TESOL Quarterly*, 11(4), 440-451. <http://dx.doi.org/10.2307/3585740>
- Cook, D. (1994) Determining the determiners. In W. Adegbite and C. Onukaogu (eds) *Language in Education in Nigeria: Some Critical Perspectives* (pp. 25–45). Ile Ife, Nigeria: Centre for Language Education and Development
- Corder, S. P. (1967). The significance of learners' errors. In J.C. Richards (Eds), *Error analysis: Perspectives on second language acquisition* (pp. 19-27). London: Longman.

- Corder, S. P. (1975). Error Analysis, Interlanguage and Second Language Acquisition. *Language Teaching*, 8(04), 201. <https://doi.org/10.1017/S0261444800002822>
- Corder, S. P. 1970. The significance of learners' errors. *International Review of Applied Linguistics* 5.161-170.
- Corder, S. P. (1973). *Introducing applied linguistics*. Harmondsworth: Penguin Education.
- Dagdeviren, G. (2010). Use of English Articles by Speakers of Turkish. *Novitas-ROYAL (Research on Youth and Language)*, 4(2), 242–250.
- Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System*, 26(2), 163–174. [https://doi.org/10.1016/S0346-251X\(98\)00001-3](https://doi.org/10.1016/S0346-251X(98)00001-3)
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language Two*. New York: Oxford University Press.
- Dunning, Ted (1993) 'Accurate Methods for the Statistics of Surprise and Coincidence', *Association for Computational Linguistics*, 19(1).
- Ekiert, M., & Han, Z. (2017). L1-fraught difficulty: The case of L2 acquisition of English Articles by Slavic Speakers. *Crosslinguistic Influence in Second Language Acquisition*, (September 2017), 147–172. <https://doi.org/10.1007/978-3-319-49959-8>
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University.
- Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.
- Ellis, R. (1997). *SLA Research and Language Teaching*. Oxford: Oxford University Press.
- Erdogan, V. (2005). Contribution of error analysis to foreign language teaching. *Mersin University Journal of the Faculty of Education*, 1(2), 261-270.
- Evans, D. (2004). Corpus building and investigation for the Humanities: An on-line information pack about corpus investigation techniques for the Humanities. *Linguistics*, 15–16. <https://doi.org/10.1109/SocialCom.2010.106>
- Federation, N., Language, M., Associations, T., Federation, N., Language, M., Associations, T., ... *Journal, L.* (2014). *Contrastive Analysis*, 63(5), 250–257.
- Flowerdew, J., & Mahlberg, M. (2009). Lexical cohesion and corpus linguistics. *Benjamins Current Topics*, (v 17), 124 p. <https://doi.org/10.1075/ijcl.11.3.03mor>
- Flowerdew, J., & Mahlberg, M. (2009). Lexical cohesion and corpus linguistics. *Benjamins Current Topics*, (v 17), 124 p. <https://doi.org/10.1075/ijcl.11.3.03mor>
- Fries, C. C. (1951), *The structure of English*. London: Longman.

- Gass, S. & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd Ed.). London: Lawrence Erlbaum Associates.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319–335. <https://doi.org/10.1016/j.jeap.2007.09.007>
- Goksel, A., & Kerslake, C., (2005). *Turkish; A Comprehensive Grammar*
- Granger, S. (1996), 'From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora', in Aijmer, K., Altenberg, B. and Johansson, M. (eds), *Languages in Contrast. Text-Based cross-linguistic studies. Lund Studies in English* 88. Lund: Lund University Press, pp. 37–51.
- Granger, S. (1998), 'The computerized learner corpus: A versatile new source of data for SLA research', in Granger, S. (ed.), *Learner English on Computer*. London and New York: Addison Wesley Longman, pp. 3–18.
- Granger, S. (2002), 'A Bird's-eye view of computer learner corpus research', in Granger, S., Hung, J. and Petch-Tyson, S. (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Language Learning and Language Teaching* 6. Amsterdam and Philadelphia: Benjamins, pp. 3–33.
- Granger, S. (2003). 'Error-tagged learner corpora and CALL: a promising synergy', *CALICO (special issue on Error Analysis and Error Correction in Computer-Assisted Language Learning)* 20 (3): 465–80
- Granger, S. (2004), 'Computer learner corpus research: Current status and future prospects', in Connor, U. and Upton, T. (eds), *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam and Atlanta: Rodopi, pp. 123–145.
- Granger, S. (2004). *Computer Learner Corpus Research: Current Status and Future Prospects. Applied Corpus Linguistics: A Multidimensional Perspective (Language and Computers)*, 52, 123–145. Retrieved from <http://www.ingentaconnect.com/content/rodopi/lang/2004/00000052/00000001/art00008>
- Granger, S. (2009), 'The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation', in Aijmer, K. (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins. 13–32.
- Granger, S., Hung, J., & Petch-Tyson, S. (2002). *Computer learner Corpora, Second Language Acquisition and Foreign Language Learning. Language Learning & Language Teaching*, 258.

- Gries, S. (2009). *Quantitative Corpus Linguistics with R*. New York: Routledge
- Halle, M. (1997). Distributed morphology: Impoverishment and fission. In B. Bruening, Y. Kang, & M. McGinnis (Eds.), *MIT working papers of linguistics 30: Papers at the interface* (pp. 125- 149). Cambridge: MITWPL
- Hamad Al-Khresheh, M. (2016a). A Review Study of Error Analysis Theory. *International Journal of Humanities and Social Science Research*, 2, 49–59. Retrieved from <http://lifescienceglobal.com/pms/index.php/ijhssr/article/viewFile/3722/2184>
- Hamad Al-Khresheh, M. (2016b). A Review Study of Error Analysis Theory. *International Journal of Humanities and Social Science Research*, 2(February), 49–59. Retrieved from <http://lifescienceglobal.com/pms/index.php/ijhssr/article/viewFile/3722/2184>
- Hawkins, R., S. Al-Eid, I. Almahboob, P. Athanasopoulos, R. Chaengchenkit, J. Hu, M. Rezai, C. Jaensch, Y. Jeon, A. Jiang, Y.-K. Leung, K. Matsunaga, M. Ortega, G. Sarko, N. Snape, & K. Velasco-Zárate. 2006. 'Accounting for English article interpretation by L2 speakers', *EUROSLA Yearbook*
- Heim, I. 1991. Articles and definiteness. In *Handbook of Semantics. an international handbook of contemporary research*, A. Stechow & D. Wunderlich (eds.), 487–535. Berlin: Walter de Gruyter
- Ho-Abdullah, I., Jaludin, A., & Zainudin, I. S. (2013). Translating definiteness and indefiniteness: A corpus-based study of Malay translational equivalences of English articles. *Pertanika Journal of Social Science and Humanities*, 21(SUPPL), 27–40.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hymes, Dell (1972), On Communicative Competence. In: Pride, John B./Holmes, Janet (eds.), *Sociolinguistics*. Harmondsworth: Penguin, 269–293.
- Hymes, Dell (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.
- Ionin, T., Ko, H. & Wexler, K. (2003). Specificity as a grammatical notion: evidence from L2-English article use. In G. Garding and M. Tsujimura (eds.), *WCCFL 22 Proceedings*, (pp. 245-258). Somerville, MA: Cascadilla Press.

- Ionin, T., Ko, H. and Wexler, K. 2004. "Article Semantics in L2 acquisition: The role of specificity." *Language Acquisition* 12: 3–69.
- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. London: New York: Longman.
- James, C. (1998). *Errors in Language Learning and Use. Exploring Error Analysis*. London & New York: Longman.
- Johansson, M. (1996). Fronting in English and Swedish: A text-based contrastive analysis. In C. Percy, C. F. Meyer & I. Lancashire (Eds.), *Synchronic corpus linguistics* (pp. 29–39). Amsterdam/Atlanta: Rodopi.
- Johansson, S. 2007. *Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies* [Studies in Corpus Linguistics 26]. Amsterdam: John Benjamins.
- Johansson, Stig (2007), *Using Corpora: From Learning to Research*. In: Hidalgo/Quereda/Santana 2007, 17–30.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials (pp. 1-13).
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Harlow: Longman
- Khansir, A. A. (2012). Error Analysis and Second Language Acquisition. *Theory and Practice in Language Studies*, 2(5), 1027–1032. <https://doi.org/10.4304/tpls.2.5.1027-1032>
- Kornfilt, J. (1997). *Turkish*. London: Routledge.
- Lado, Robert. 1968 [1957]. *Linguistics across cultures: applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.
- Langacker, R. W. 1991. *Foundations of Cognitive Grammar, Vol. II: Descriptive Application*. Stanford CA: Stanford University Press.
- learners. *European Journal of Social Sciences*, 18(3), 426-433
- Leech, G. (1997). Teaching and language corpora: A convergence. In *Teaching and language corpora* (pp. 1-23)
- Leech, G., & Svartvik, (2002). *A Communicative Grammar of English*. 3rd Edition ISBN:0582506336
- Leuc, A., Laylsileaac, I., Already, L., Handbook, T., Vdlred, C. L., Handbook, E. T., ... Handbook, T. (n.d.). *Blackwell Handbooks in Linguistics*.
- Ludeling, A., & Kyto, M. (2008). *Corpus Linguistics: An International Handbook*. Mouton de. Retrieved from

<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Corpus+linguistics.+An+international+handbook#5>

- Mahmoodzadeh, M. (2012, April). A cross-linguistic study of
- Mahmoud, A. (2005). Collection errors made by Arab learners of English. *Asian EFL Journal*, 6 (2).
- Maratsos, M. (1974). Preschool children's use of definite and indefinite articles. *Child Development*, 45, 446-455.
- Master, P. (1987). A cross-linguistic interlanguage analysis of the acquisition of the English article system (Unpublished doctoral dissertation). University of California at Los Angeles.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: method, theory, and practice*. Cambridge Textbooks in Linguistics, xv, 294 p.
<https://doi.org/10.1017/CBO9781107415324.004>
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics* Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advance resource book*. New York: Routledge.
- McGregor, W. (1997). *Semiotic grammar*. Oxford: Clarendon.
- Meunier, F. (2011). Corpus linguistics and second/foreign language learning: exploring multiple paths. *Revista Brasileira de Linguística Aplicada*, 11(2), 459–477.
<https://doi.org/10.1590/S1984-63982011000200008>
- Meyer, C. 2002. *English Corpus Linguistics: An Introduction*. Cambridge University Press.
- Meyer, C. F. (2004). *English Corpus Linguistics: An introduction*. Cambridge: Cambridge University Press
- Mukherjee, J. (2006). 'Corpus linguistics and English reference grammars', in A. Renouf and A. Kehoe (eds.) *The Changing Face of Corpus Linguistics*, pp. 337–54. Amsterdam: Rodopi.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 125–157). Amsterdam: John Benjamins Publishing Company.
- O'Keeffe, A., & McCarthy, M. (2010). *The Routledge handbook of corpus linguistics*. New York, NY: Routledge.
- Park, S. (2013). Testing the Interface Hypothesis: Acquisition of English Articles by Korean L2 Learners. *Proceedings of the 12th Generative Approaches to Second*


- Language Acquisition Conference (GASLA 2013), (1), 155–161.
- Pica, T. (2002). Subject-Matter Content: How does it Assist the Interactional and Linguistic Needs of Classroom Language Learners? *Modern Language Journal*, 86(1), 1–19. <https://doi.org/10.1111/1540-4781.00133>
- prepositions in Persian and English. *The Effect of Transfer Theory and Practice in Language Studies*, 2(4), 734-740.
- Prins, H. (2006). "Conquering Chinese English in the ESL Classroom: The Internet TESL Journal, 12 (11): 78-90.
- Pullum, K., G., Huddleston, R., (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press
- Quirk, R., Greenbaum, S. L., & Leech, G. G. and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Harlow: Longman.
- Römer, U. (2005). Corpora and language teaching. *Corpus Linguistics Volume 1*, 112–131. <https://doi.org/10.1016/j.esp.2003.12.001>
- Schachter, J. & Celce-Murcia, M. (1977). Some reservations
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-231.
- Sinclair, J. (2004). *How to use corpora for language teaching*. Amsterdam & Philadelphia: John Benjamins.
- Snape, N. (2009). Exploring Mandarin Chinese speakers' L2 article use. *Representational Deficits in SLA: Studies in Honor of Roger Hawkins*, (January), 27–52. <https://doi.org/10.13140/2.1.2866.5604>
- Sridhar, S.N. (1980). Contrastive analysis, error analysis and interlanguage: three phases of one goal. In K. Croft (Ed) *Readings on English as a Second Language*, 120-31. Cambridge Mass: Winthrop Publishers:
- Tarone, E. (2006). The Interlanguage Hypothesis. *Science Direct*, 4, 747–752. Retrieved from <http://www.sciencedirect.com/science/article/pii/B0080448542006180>
- Teubert, W. (1996), 'Comparable or parallel corpora?', *International journal of lexicography*, 9: 238-264.
- Tognini-Bonelli, E. (2002). Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. *Lexis, in Contrast*, 73–95.
- Tognini-Bonelli, E. (2002). Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. *Lexis in Contrast*, 73–95.
- Trenkic, D. (2008). The representation of English articles in second language grammars:


- Determiners or adjectives? *Bilingualism: Language and Cognition*, 11(1), 1–18.
<https://doi.org/10.1017/S1366728907003185>
- Trimasse, N. (2016). The Role of Transfer in the Use of English Articles by Moroccan Learners. *International Journal of Linguistics*, 8(5), 58.
<https://doi.org/10.5296/ijl.v8i5.9855>
- Widdowson, H. G. (2004) *Text, Context, Pretext: Critical Issues in Discourse Analysis*. Oxford: Blackwell. ———(2007) *Discourse Analysis*. Oxford: Oxford University Press.
- Willis, Dave (1990), *The Lexical Syllabus. A New Approach to Language Teaching*. London: Harper- Collins.
- Wizner, S. (2011). Corpus Linguistics and the Description of English. *Corpora*, 6(1), 111–114.
- Zandvoort, R. W. (1974). A grammar of contemporary English. *Lingua*, 33, 185–190.
[https://doi.org/10.1016/0024-3841\(74\)90042-4](https://doi.org/10.1016/0024-3841(74)90042-4)

APPENDICES

APPENDIX I: List of English Determiners (articles) Used in A1-A2


	A1 – A2		
	A	AN	THE
Missing Determiner MD	93	9	125
Unnecessary Determiner UD	36	3	31
TOTAL	129	12	156


Incorrect Determiner Form - FD			
			Frequency
1.	a	an	4
TOTAL			4

Replace Determiner - RD			
			Frequency
1.	the	a	6
2.	a	my	1
3.	a	the	1
4.	a	these	1
5.	some	a	1
6.	that	a	1
7.	the	this	1
8.	these	the	1
9.	my	the	1
TOTAL			14

APPENDIX II: List of English Determiners (articles) Used in B1-B2


B1 – B2			
	A	AN	THE
Missing Determiner MD	106	6	201
Unnecessary Determiner UD	41	1	66
TOTAL	147	7	267

Incorrect Determiner Form - FD			
			Frequency
1.	a	an	4
2.	an	a	1
TOTAL			5

Replace Determiner - RD			
			Frequency
1.	the	a	9
2.	a	the	4
3.	one	a	4
4.	the	their	2
5.	their	the	2
6.	my	a	1
7.	a	his	1
8.	a	her	1
9.	any	a	1
10.	the	an	1
11.	an	the	1
12.	the	your	1
13.	your	the	1
14.	my	the	1
15.	the	my	1
TOTAL			31

APPENDIX III: List of English Determiners (articles) Used in C1-C2


	C1 – C2		
	A	AN	THE
Missing Determiner MD	140	13	292
Unnecessary Determiner UD	59	9	149
TOTAL	199	22	441

Incorrect Determiner Form - FD			
			Frequency
1.	a	an	7
2.	an	a	6
	TOTAL		13

Replace Determiner - RD			Frequency
	→		
1.	the	a	26
2.	a	the	18
3.	the	an	7
4.	your	the	3
5.	the	your	3
6.	our	the	3
7.	the	its	2
8.	this	a	2
9.	the	our	2
10.	the	this	2
11.	the	their	2
12.	that	the	2
13.	this	the	2
14.	an	the	2
15.	my	the	1
16.	my	a	1
17.	that	an	1
18.	any	a	1
19.	our	a	1
20.	some	an	1
21.	a	their	1
22.	a	his	1
23.	some	a	1
24.	a	that	1
25.	one	the	1
26.	one	a	1
27.	another	a	1
28.	the	my	1
29.	their	the	1
30.	our	the	1
31.	any	an	1
TOTAL			93

APPENDIX IV: List of English Determiners (articles) Used Overall

OVERALL			
	A	AN	THE
Missing Determiner MD	393	28	618
Unnecessary Determiner UD	136	13	246
TOTAL	529	41	864

Incorrect Determiner Form - FD			
			Frequency
1.	a	an	15
2.	an	a	7
TOTAL			22

Replace Determiner - RD			
	→		Frequency
1.	the	a	37
2.	a	the	23
3.	the	an	7
4.	your	the	4
5.	the	your	4
6.	our	the	3
7.	the	its	2
8.	this	a	2
9.	the	our	2
10.	the	this	3
11.	the	their	4
12.	that	the	2
13.	this	the	2
14.	an	the	3
15.	my	the	3
16.	my	a	2
17.	that	an	1
18.	any	a	2
19.	our	a	1
20.	some	an	1
21.	a	their	1
22.	a	his	2
23.	some	a	2
24.	a	that	1
25.	one	the	1
26.	one	a	5
27.	another	a	1
28.	the	my	2
29.	their	the	3
30.	our	the	1
31.	any	an	1
32.	a	my	1
33.	a	these	1
34.	that	a	1
35.	these	the	1
36.	a	her	1
TOTAL			138

APPENDIX V: Pre-test and Post-test Questions

Determiners in Use

Fill in the blank with the appropriate article [a, an, the]

1. I am in Turkey now. I decided to send postcard to you I saw in Istanbul museum.
2. I am having great time.
3. You know that's historic place and very interesting.
4. I have got wonderful alarm clock and big stamp collection and new shoes.
5. My telephone number is 3421982. It is 33:10 Nokia mobile phone. It's been used for about six months.
6. We must bring our crayons, pencil, eraser and some papers with us.
7. I think it will take one and half hours.
8. The new art class starts on Monday and it's 2 -month class.
9. You should bring your paints and drawing sheet with you.
10. It has got radio, camera, calculator, etc. I don't know how much it was.
11. I love living room, I can eat in front of the TV.
12. You should bring your history book for history exam.
13. I am landlord of the flat.
14. She has very beautiful blue eyes.
15. Miami is like dream city. I went to American Football match yesterday.
16. You can take photos with it and it has MP 3 player.
17. I think you don't have to bring anything. Best wishes, **art** class is on Monday, 19th of June and it's between 5 p.m. and 8 p.m., for 3 hours.
18. It will always be on my mind. Dear Maria, we are at hotel in Istanbul with my husband. We can see Bogaz bridge from window.
19. We started trip in Nottingham. First, we visited Nottingham Castle, then we went to train station and after short journey we arrived in London.
20. But I am sure this will come to end in the next 4 days because on radio this morning politician informed the public that they could find solution with workers who are on strike.

APPENDIX VI: Lesson Plan 1**Traditional Method**

Proficiency Level: A1-A2

Native language: Turkish

Grade: 7th

Duration: 40 Minutes (Each week)

Topic: English Determiners (Articles)

Materials: Marker, Board, Papers

First Week:

- i. Introduction to English Determiners
- ii. Extracting articles from the main determiner stem
- iii. Guiding with clear example sentences

Second Week

- i. Introducing determiner-based activities
- ii. Playing games using articles in sentences
- iii. Comparing Turkish and English articles

Third Week

- i. Presenting articles in a different context e.g., a house/a beautiful house
- ii. Article usage in plural and in singular form
- iii. Use of articles with consonants and vowels (pronunciation related)

Fourth Week

- i. English determiner-based reading activity
- ii. Making sentences with Articles
- iii. Revision of the general English articles
- iv. Post-test

APPENDIX VII: Lesson Plan 2**Corpus-driven Activity**

Proficiency Level: A1-A2

Native language: Turkish

Grade: 7th

Duration: 40 Minutes (Each week)

Topic: English Determiners (Articles)

Material: Marker, Board, Extracted Corpus Materials, Printed Activities

First Week

- i. Introduction to English determiners
- ii. Extracting articles from the central determiners
- iii. Giving examples using simple sentences

Second Week

- i. Giving more examples to make the participants comprehend the core of the English articles
- ii. Presenting ***fill in the blanks*** activities
- iii. Following up the activities

Third Week

- i. Presenting articles in a different context e.g., a house/a beautiful house
- ii. Differentiating between using articles with nouns and adjectives
- iii. Guiding the students to do the ***error correcting*** activities.

Fourth Week

- i. Presenting ***picture description*** activity
- ii. Guiding the students to the ***listening*** activity
- iii. Checking up the activities
- iv. Post-test

APPENDIX VIII: Fill in Blanks Activity**Name:****Group:****Determiners in Use**

Let us have fun with articles. Please, insert the appropriate article (a/an/the) where necessary.

a/an

1.mistake
2.uniform
3.hour or two
4.abbreviation
5.union
6.hole in the ground
7.dream
8.untrue story
9.hill
10.interesting dream
11.urgent message
12.handsome man
13.empty box
14.universal problem
15.honest man

a/an/the

1. My mother asked me to close window
2. If I pass the exam my father will buy me Walkman
3. Ali has got unique personality.
4. The government is not willing to close schools
5. Ball is round object.
6. The instruction says we should answer all questions
7. I want my father to buy me puzzle as my birthday present.
8. They are eager to open..... organization
9. oranges that my father bought are fresh
10. He is looking for opportunity to move to another country.
11. I saw the advertisement on newspaper
12. Upon honest man's fortune
13. I saw ugly picture.
14. She needs internet connection to watch..... Olympic games
15. Simon Bolivar is hero to many people.
16. I called Jim by the wrong name. It washonest mistake.
17. I hadunusual experience yesterday.
18. Ann hadunique experience yesterday.
19. I often ask my parents foradvice.
20. Jack has wallet in his back pocket.

APPENDIX IX: Error Correction**Determiner in Use****Error Correction***A. Rewrite the following sentences*

1. I ate very big pizza.

.....

2. I have seen museum and lots of interesting *things* here.

.....

3. I hope they'll win **an** another match.

.....

4. I *like* this gift much *more* than others.

.....

5. I want to *have* a *big new* flat, *in* area *that* is *good quality*.

.....

6. If you use coach you can get there *in* 1 hour.

.....

7. If you want to use train, you can get there *in* 2 and a half hours.

.....

8. It is **a** *friendly*.

.....

9. It is Ericson T 10.

.....

10. My new flat *is a* small.

.....

11. She is from *a Somalia*, she is tall and very slim.

.....

12. She likes wearing black skirt and *T-shirt*.



.....

13. Take care ! I want to sell old rug.

.....

14. They did n't cost a lot, not even *a* 100 YTLs.

.....

15. We are *from* same country .

.....

B. Circle the appropriate article

1. We 've seen **the** | **a** very nice hotel.
2. But my favourite present is **the** | **a** CD of Euro 2004. My uncle gave it to me.
3. My favourite present is **the** | **a** jigsaw puzzle.
4. We are **a** | **the** same age and she looks Turkish.
5. We went to **the** / **a** disco last night with my husband.



APPENDIX X: Picture Description

Source: <https://tr.pinterest.com/pin/394557617349466803/>

CURRICULUM VITAE

Name : ABDULKADIR
Surname : ABDULRAHIM
Father Name : ABDULRAHIM ARJALI
Nationality : NIGERIAN
Date of Birth : 25.07.1993

CONTACT

ababdulrahim1@gmail.com

ababdulrahim@live.com

Website: www.arjalisticpen.com

FOCUS

Teaching English as a Foreign Language

Corpus Linguistics

Applied Linguistics

EDUCATION

- Çukurova University, Adana
Master of Arts (MA), English Language Teaching
(2016 - 2020)

- Çukurova University, Adana
Bachelor of Arts (B.A.), English Language Teaching
(2012 - 2016)

LANGUAGES

- i. Arabic (Beginner)
- ii. English (Native or Bilingual)
- iii. French (Beginner)
- iv. Hausa (Native or Bilingual)

- v. Ngizim (Native or Bilingual)
- vi. Turkish (Very Good User)

WORK EXPERIENCE

- Bahcesehir College, Mezitli
English Language Teacher
August 2019 – Present Mersin, Turkey

- Appen
Linguist
June 2017 - Present

- Ortadogu okullari, Adana
English Second Language Teacher
March 2019 - July 2019, Adana, Turkey

- Beren Children Academy
English Second Language Teacher
July 2018 - March 2019, Adana, Turkey

- Deulcom International, Adana
English Second Language Teacher
July 2016 - October 2018 Adana, Turkey

- Kampus Okullari
English Second Language Teacher
September 2017 - June 2018 Adana, Turkey

- Okyanus Kolejleri
ESL Teacher
March 2017 - June 2017 (Adana, Turkey)

CERTIFICATIONS/TRAININGS

1. Creative Writing Course
2. Implementation of the European Language Portfolio System
3. Train the Trainer
4. Montessori Eğitim Felsefesi & Günlük Yaşam Becerileri (Philosophy of Montessori Education and Daily Living Skills)
5. Oyunlarla İngilizce & Yaratıcı Drama (English with Games and Creative Drama)
6. 2019 Language Documentation Training Workshop

RESEARCH EXPERIENCE

Published Paper

1. Problems and Challenges Faced by the African Students in Adana and Some Solutions to the Problems
2. Perceptions of African Students About Turkish Language and Culture
3. Türkçenin Yabancı Dil Olarak Öğretiminde Kullanılan Ders Kitaplarında Yer Alan Konuşma Becerileri Üzerine Bir Çalışma (*A Study on Speaking Skills in the Textbooks Teaching Turkish as a Foreign Language*)

Unpublished

1. The Role of Authentic Reading Materials to Improve Speaking Skills of Turkish Adult EFL Learners
2. Yabancı Öğrencilerin Türkçe Konuşma Becerileriyle Başa Çıkma Yolları Üzerinde Bir Çalışma