

KARADENİZ TEKNİK ÜNİVERSİTESİ * SOSYAL BİLİMLER ENSTİTÜSÜ

BATI DİLLERİ VE EDEBİYATI ANABİLİM DALI
UYGULAMALI DİLBİLİMİ YÜKSEK LİSANS PROGRAMI

**IMPACT OF “TEXT MINING AND IMITATING STRATEGIES” ON LEXICAL
RICHNESS, LEXICAL DIVERSITY AND GENERAL SUCCESS IN SECOND
LANGUAGE WRITING**

YÜKSEK LİSANS TEZİ

Sevcan BAYRAKTAR ÇEPNİ

EKİM 2014

TRABZON

KARADENİZ TEKNİK ÜNİVERSİTESİ * SOSYAL BİLİMLER ENSTİTÜSÜ

BATI DİLLERİ VE EDEBİYATI ANABİLİM DALI
UYGULAMALI DİL BİLİMİ YÜKSEK LİSANS PROGRAMI

**IMPACT OF “TEXT MINING AND IMITATING STRATEGIES” ON LEXICAL
RICHNESS, LEXICAL DIVERSITY AND GENERAL SUCCESS IN SECOND
LANGUAGE WRITING**

YÜKSEK LİSANS TEZİ

Sevcan Bayraktar Çepni

Tez Danışmanı: Yrd. Doç. Dr. Elif TOKDEMİR DEMİREL

EKİM 2014

TRABZON

ONAY

Sevcan Bayraktar Çepni tarafından hazırlanan Impact of “Text Mining and Imitating Strategies on Lexical Richness, Lexical Diversity and General Success in Second Language Writing” adlı bu çalışma 14.10.2014 tarihinde yapılan savunma sınavı sonucunda oybirliği ile başarılı bulunarak jürimiz tarafından Uygulamalı Dilbilimi dalında **yüksek lisans tezi** olarak kabul edilmiştir.

[i m z a]

.....

[Yard. Doç. Dr Elif Tokdemir Demirel] (Başkan)

.....

[i m z a]

.....

[Yard. Doç Dr. Semin Kazazoğlu] (juri)

.....

[i m z a]

.....

[Yard. Doç. Dr. Tuncer Yılmaz] (juri)

.....

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduklarını onaylarım. ... / ... /

.....

Enstitü Müdürü

BİLDİRİM

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını, aksinin ortaya çıkması durumunda her tür yasal sonucu kabul ettiğimi beyan ediyorum.

İmza

Sevcan Bayraktar Çepni

14.10.2014

ACKNOWLEDGEMENTS

It is a real pleasure to thank all of those who supported and assisted me during this journey of discovery and learning. First, I would like to express my gratitude to my supervisor, Assist. Prof. Dr. ELİF TOKDEMİR DEMİREL for her endless support, constructive feedback, endless patience for me, and wonderful guidance even in hard and busy times. She encouraged and supported me throughout this hard period. I would also like to express my gratitude to Prof. Dr Douglas BIBER who works in Northern Arizona University for his help to tag all files with his tagging programme.

I would like to express my gratitude to Assist. Prof. Dr. Semin KAZAZOĞLU for being in the jury and for her valuable remarks and support.

I would like to express my gratitude to Assist. Prof. Dr. Tuncer YILMAZ for being in the jury and his valuable remarks and support.

I would like to express my gratitude to Assist. Prof. Dr. Mustafa Zeki Çıraklı for being in the jury and his valuable remarks and support.

I would like to express my gratitude to Prof. Dr. Erdoğan BADA for being in the jury and his valuable guidance and support.

Most importantly, none of this would have been possible without the love and patience of my family. My parents Edip and Güler Bayraktar, to whom this dissertation is dedicated to, have been a constant source of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my beloved husband Gökhan ÇEPNİ and my daughter Bilge ÇEPNİ.

Sevcan BAYRAKTAR ÇEPNİ

CONTENTS

ACKNOWLEDGEMENTS	V
CONTENTS	VI
ÖZET	IX
ABSTRACT	X
LIST OF TABLES	XI
LIST OF FIGURES	XIII
LIST OF ABBREVIATIONS	XIV

CHAPTER 1

1.INTRODUCTION	1-12
1.1 Introduction.....	1
1.2. Background to the Study.....	1
1.2.1. Written Productions and Corpus Linguistics.....	4
1.2.2. Reading and Writing Relationship	6
1.2.3. Approaches to Teaching Writing.....	7
1.3. Statement of the Problem.....	8
1.4 Purpose of the Research and Research Questions.....	9
1.5 Operational Definitions.....	9
1.5.1. Lexical Richness.....	9
1.5.2. Lexical Diversity.....	9
1.5.3. Learner Corpora.....	10
1.5.4. Concordances.....	10
1.6. Assumption of the Study.....	11
1.7. Limitation of the Study.....	11
1.8. Organization of the Thesis.....	12

CHAPTER 2

2. LITERATURE REVIEW	12-25
2.1. Introduction	13
2.2. Theories on the Relationship between Reading and Writing.....	13
2.2.1 Previous Research on Reading Writing Relationship.....	16
2.2.2 Strategies that Connect Reading and Writing.....	18
2.3. Incidental Vocabulary Learning Through Reading.....	20
2.4. Lexicographic Research and Lexical Diversity (LD).....	21
2.5. Corpus Linguistics.....	22
2.5.1. Learner Corpora.....	23

CHAPTER 3

3. METHODOLOGY	26-35
3.1. Introduction	26
3.2. Research Design	26
3.3. Participants	26
3.4. Instrumentation.	28
3.4.1. Parts of Speech Tagging.....	29
3.4.2. Corpus Description.....	30
3.5. Data Collection.....	31
3.6. Data Analysis.....	33

CHAPTER 4

4. RESULTS AND DISCUSSION	36-47
4.1. Introduction.....	36
4.2. Results.....	39
4.2.1. General Lexical Richness.....	39
4.2.2. The Use of Nouns in Frequency and Form: Singular or Plural.....	41

4.2.3. The Use of Adjectives in frequency and type: attributive comparative and superlative adjectives.....	43
4.2.4. The Use of Verbs in frequency and type:	44
4.2.5. General Success in Terms of Grading.....	45

CHAPTER 5

5.CONCLUSION.....	48-51
5.1. Introduction.....	48
5.2. Conclusions.....	49
5.3. Limitations and Suggestions.....	51
REFERENCES	52-60
APPENDIX 1.....	61
APPENDIX 2.....	63
APPENDIX 3.....	66
APPENDIX 4.....	67
APPENDIX 5.....	78
APPENDIX 6.....	80
APPENDIX 7.....	81
APPENDIX 8.....	82
APPENDIX 9.....	85
APPENDIX 10.....	86
APPENDIX 11.....	87
APPENDIX 12.....	89
CV	90

ÖZET

Bu çalışmanın amacı “metin madenciliği” ve bir kompozisyon modelini “taklit etme” stratejilerinin öğrenci kompozisyonlarındaki kelime zenginliği, çeşitliliği ve öğrenci başarısı üzerine olan etkisini incelemektir. Bu çalışmanın katılımcıları 2012/2013 Akademik yılında Karadeniz Teknik Üniversitesinin İngiliz Dili ve Edebiyatı 1.sınıf öğrencileridir (39 kontrol grubunda; 59 deney grubunda). Deney grubundaki öğrenciler örnek kompozisyonlar üzerinde çalışarak kelime kullanımlarını pekiştirmek için yazı inceleme ve taklit etme üzerine eğitildiler. Kontrol grubundaki öğrenciler ise sadece nasıl kompozisyon yazacakları üzerine eğitildiler ve bu süreçte örnek kompozisyonlar üzerine bir inceleme yapmadılar. Deney grubundaki eğitimden sonra her iki gruptaki öğrencilerden en az 250 kelimelik “Hayran olduğunuz biri” başlıklı beş paragraflı bir betimleyici kompozisyon, “Çalışanı Takip Etmek” başlıklı beş paragraflı bir avantaj dezavantaj kompozisyonları yazmaları istendi. Her iki gruptaki öğrencilerden (15826 kelimelik kontrol grubundan, 25027 kelimelik deney grubundan) oluşturulan iki derlem iki ayda toplandı ve incelemeyi kolaylaştırmak adına Biber (1993) etiketleriyle etiketlendi. İstatistiksel hesaplamalara göre iki grup arasındaki kelime zenginliğinde anlamlı bir farklılık olmasa da deney grubundaki öğrencilerin kompozisyonlarının daha zengin olduğu görüldü. Ayrıca genel isimler ve genel fiiller alt kategorisinde deney grubundaki öğrencilerin kompozisyonlarında istatistiksel anlamlı bir farklılık bulunurken, bu öğrencilerin genel başarısının da diğerlerine oranla yüksek olduğu saptandı.

Anahtar Kelimeler: Metin madenciliği, taklit etme, kelime zenginliği, kelime çeşitliliği, öğrenci derlemi.

ABSTRACT

This study aimed to find out the impact of “text mining and imitating” strategies on lexical richness, diversity and general success of students in their compositions in a writing class. The participants (39 in the control group; 59 in the experimental group) of this study were first year students of Karadeniz Technical University in English Language and Literature department in the academic year of 2012- 2013. Students in the experimental group working on model essays were instructed on how to mine and imitate their essays to be more competent in vocabulary while students in the control group were directly instructed on how to write their essays without using any reference book or working on model essays. After treatment to the experimental group, participants were asked to write one descriptive essay entitled “the person you admire” and one advantage disadvantage essay on “employee monitoring”. They were asked to write five paragraph essays with at least 250 words. Two corpora of 39 students in the control group and 59 students in the experimental group, (corpus of 15826 and a corpus of 25027 words respectively) were compiled in two months , and these essays were tagged (with Biber tagger’s tag descriptions) in order to ease the analysis. Though there is statistically slight difference in terms of lexical richness, the study revealed that compositions in the experimental group are lexically richer than those in the control group. Common nouns and general verbs subclasses were found to be used more in compositions of the experimental group. In addition to that, students in the experimental group got better grades than those in the control group.

Key words: Text mining, imitating, lexical richness, lexical diversity, learner corpora.

LIST OF TABLES

<u>Table Nr.</u>	<u>Table Name</u>	<u>Page Nr.</u>
1	Receptive and productive vocabulary knowledge of Nation (2001) considering “form”.....	2
2	Receptive and productive vocabulary knowledge of Nation (2001) considering “meaning”	2
3	Receptive and Productive Vocabulary Knowledge of Nation (2001) considering “Use”	2
4	Selected concordance lines for globe in The New York Times, 1990, sorted on the first word to the left and the second word to the left.....	12
5	Common Ground Between Reading and Writing.....	17
6	Independent Samples t test Results: Group statistics for Passing Grades of Students 28.....	25
7	Independent Samples t test results: Passing grades of Students.....	26
8	Independent Samples t-test results: Group Statistics for Narrative Essay Grading.....	27
9	Independent Samples t-test results: Narrative Essay Grading.....	28
10	Sentences from Tagged Texts.....	31
11	Treatments in the control and the experimental groups.....	33
12	Measures of Lexical Richness.....	34
13	Tags for Adjective Category (adapted from Biber 1993).....	37
14	Tags for Noun category.....	38

15	Tags for Verb category.....	38
16	Example of Norming.....	39
17	Type Token Ratio of Essays: Group Statistics.....	40
18	Independent Samples T-test Results : Type Token Ratio of Essays.....	41
19	Log-likelihood ratio (LL) of overall frequency of Nouns among the experimental and the control group corpora.....	42
20	Log-likelihood ratio (LL) of overall frequency of adjectives among the experimental and the control group corpora.....	44
21	Log-likelihood ratio (LL) of overall frequency of verbs among the experimental and the control group corpora.....	45
22	Reliability Statistics for two raters	46
23	Independent Sample T-test result: Group Statistics for General Success Rates of Students.....	46
24	Independent Sample T-test result General Success Rates of Students.....	47

LIST OF FIGURES

<u>Figure Nr.</u>	<u>Figure Name</u>	<u>Figure Nr.</u>
1	Selected concordance lines for globe in The New York Times.....	4
2	Example of a graphic organizer for the control group.....	29
3	Noun distribution according to its subclasses.....	38
4	Adjective distribution according to its subclasses.....	40
5	General verb distribution according to its subclasses.....	42

LIST OF ABBREVIATIONS

ICLE	: International Corpus of Learner English
LD	: Lexical Diversity
L1	: First Language
L2	: Second Language
POS	: Parts of Speech
SLA	: Second Language Acquisition
TTR	: Type-Token Ration

CHAPTER ONE

1. INTRODUCTION

1.1. Introduction

This chapter includes background to the study, statement of the problem, the purpose of the study, research questions, assumptions and limitations of the study, and lastly, organization of the thesis.

1.2. Background to the Study

Development of vocabulary is a fundamental part of learning a second or a foreign language. Wilkins (1992: 111) declared that “little can be conveyed without grammar but nothing can be conveyed without vocabulary.” In spite of this fact, this development has been neglected in foreign language teaching. This neglect, according to Carter (1987) stems from two main factors. (1) More emphasis was traditionally placed on grammar because it was considered that if grammar instruction was given well and learners were exposed to target language sufficiently, they would easily pick up words; (2) there was a perception that syntactic relations, as they are finite, can be efficiently learned while lexis was hard to be learned as they are infinite. (as cited in Priz, 2008)

In order to fill this gap, many studies on vocabulary learning have been conducted so far (see Nassaji, H. 2006; Lee, S. H. & Muncie, J. 2006; Nation, I. S. P. 2008). First and second language research attempted to clarify what is meant by “vocabulary knowledge” or “lexical knowledge” (Nation, 1990; 2001; Qian, 1998). Knowledge of words has two forms: Receptive and productive. Receptive means understanding or recognizing a word in a context while productive means using these words in written and spoken context. Nation (2001) has described lexical knowledge as taxonomy of components. He came up with a set of 18 questions classified into three categories, each containing receptive and productive aspects, to know a word:

Table 1: Receptive and productive vocabulary knowledge considering “form”

Aspects	Receptive	Productive
Spoken Language	What does the word sound like?	How is the word pronounced?
Written Language	What does the word look like?	How is the word spelled?
Word Parts	What parts are recognizable in this word?	What word parts are needed to Express meaning?

Source: Nation, 2001

Table 2: Receptive and productive vocabulary knowledge considering “meaning”

Aspects	Receptive	Productive
Form and meaning	What does this word form signal?	What word form can be used to express this meaning?
Concept and referents	What is included in the concept?	What items can the concept refer to?
Associations	What other words does this make us think of it?	What other words could we use instead of this one?

Source: Nation, 2001

Table 3: Receptive and Productive Vocabulary Knowledge of Nation (2001) considering “Use”

Skills	Receptive	Productive
Grammatical functions	In what patterns does the word occur?	In what patterns must we use this word?
Collocations	What words or types of words occur with this one?	What words or types of word must we use instead of this one?
Constraints on use (register, frequency...)	Where, when and how often would we expect to meet this word?	Where, when and how often can we use this word?

Source: Nation, 2001

This study focuses on the productive aspect of Nation's taxonomy through text mining and imitating strategies. Depending on Nation's (2001) taxonomy, it can be concluded that knowing vocabulary requires both having an idea about the usage of the vocabularies and the relation among one another.

Armbruster, Lehr, and Osborne (2003) mentioned two ways of learning vocabulary: direct and indirect vocabulary instruction. Direct vocabulary teaching means teaching specific words with direct explicit instruction. It can be done by teaching students listing and memorizing unknown words before and activity or teaching students how to analyze word roots (affixes). On the other hand, indirect vocabulary teaching is related with exposure to reading and spoken materials such as conversation with others, reading stories or magazines etc... Learners learn vocabulary implicitly during this process.

Readers with varying proficiency level commonly come across with unknown words while reading, and should take advantage of these incidental encounters in building their L2 lexicon. Frequency of such encounters is the most popular predictor of learning (Ellis 2002). Depending on this fact, it has been presumed that foreign language learners pick up most of the words incidentally while reading a context, which led to research on the relationship between vocabulary size and reading comprehension in the L2 (Laufer, 1990; Qian,2002). These studies show a correlation between vocabulary size and learners' reading abilities. In other words, students' vocabulary knowledge is considered to be an important predictor of learners reading fluency and reading comprehension.

Albrechtsen, Haastrup and Henriksen (2008: 22) wrote that:

L1 lexicon items are not stored randomly but in a well-organized web with connections or pathways between the words, which makes retrieval of these words rapidly and easily because of the varied and well- established access routes in the lexical store. When it comes to the mental lexicon of foreign learners, it is not unusual to think that they have different stages of interlanguage development. That is, L2 learners expose less to the target language, which results in difference in lexicon size and structural properties of the word store.

However, Albrechtsen et al. (2008) stated that writing in a foreign language is not different in terms of appealing to components of knowledge and fluency for NS and NNS. Both groups are expected to have similar difficulty in both linguistic EFL knowledge (vocabulary, grammar, and orthography) and fluency and accessibility of this linguistic knowledge.

Schellekens (2007: 103) claimed that “Many students struggle with writing and they often find it hard to write at all, even about topics that they know well, such as their home life, their children, or their job”. This difficulty may stem from the fact that students’ lexical knowledge is not enough to help them produce detailed expressions while writing in a second or foreign language. Therefore, improving lexical knowledge, with direct or indirect instruction, may mean raising proficiency levels of students at the same time. This study is concerned with improvement in lexical richness, diversity and general success of students, through text mining and imitating strategies applied on model essays. In order to analyze student written production corpus linguistics methods which entail “the analysis of very large collection of electronically stored texts, aided by computer software” and “firmly rooted in empirical, inductive forms of analysis” (Baker, 2010: 93-94) were used.

1.2.1. Written Productions and Corpus Linguistics

Improvements in computer technologies have become a great use for linguists and teachers of foreign languages who seek for better insights on language learners’ learning process. Instead of storing a large number of written or transcribed texts with old storage methods such as piling them up in libraries or in files, corpus tools whose storable characteristic make them a great source, have been widely utilized in the analysis of learners’ written and spoken productions.

As O’Kafee, McCarthy and Carter (2007: 1) put it, “A corpus is a collection of texts, written or spoken, which is stored on a computer”. Definition of corpus by Biber, Conrad and Reppen (1998: 4) suggested that “corpus is a principled collection of texts available for qualitative and quantitative analysis”. O’ Kafee et al. (2007) found this definition useful as it captures some important issues as follows. In addition to the definition that “corpus is a principled collection of texts”, they say that any collection of text may not be a corpus as a corpus must represent something, and it must be judged according to the level of its representatives, so great care should be given for the designing process of corpus to ensure its representatives. O’ Kafee et al. (2007: 1) gives example for the preparation stage of collecting data as follows:

If we wished to build a corpus to represent classroom discourse in ELT, we would have to create a design matrix that would ideally capture all the essential variables of age, gender, location, type of school (e.g. state or private sector), level, teacher (e.g. gender, qualifications, years of experience, whether native or non-native speaker), class size (large groups, small groups or one-to-one), location, nationalities and so on.

As for the other definition which suggests that “a corpus is available for qualitative and quantitative analysis” O’ Kafee et al. (2007) state that there are many different ways of looking at the language in a corpus. For example, by using a corpus of newspapers, the number of times the words “fire” and “blaze” occur can be examined, which will give quantitative results that can be compared with frequencies in other corpora such as casual conversation or general written English. For qualitative use, it can be seen how a word or phrase is used across a corpus.

Flowerdew (2002) states that the compiling sources, size, and design of the corpus is determined according to purpose of the researchers. In terms of size and design, corpora can be divided into two kinds: general corpora and specialized corpora. For example, BNC = British National Corpus, or Bank of English are general corpora that aim at representing a language or variety as a whole (contain both spoken and written language, different text types etc.) One type of specialized corpora is learner corpora such as ICLE which stands for International Corpus of Learner English. This corpora aims at representing the language as produced by learners of this language.

Corpus linguistics can be regarded as methodology based on electronically collected texts or as a method of carrying out linguistic analyses through collected corpus. Bennet (2010) states that corpus linguistics serves to answer two fundamental research questions:

What particular patterns are associated with lexical or grammatical features?
How do these patterns differ within varieties and registers?
In an attempt to answer these questions, some researchers commented on the uses of such analysis.

Sahillioglu Sahinkayasi, and Sahinkayasi (2011: 231) state that “corpus linguistics has opened a new horizon in language research”. Shin Chul (2010: 58) reported that, “Corpus based descriptions (e.g., collocation, concordance, and semantic prosody) of native speakers’ language use could guarantee that learners would use native-like language”. The Corpus Approach (Biber, Conrad and Reppen 1998: 4) is comprised of four major characteristics:

1. It is empirical, analyzing the actual patterns of language use in natural contexts.
2. It utilizes a large and principled collection of natural texts as the basis for analysis.
3. It makes extensive use of computers for analysis
4. It depends on both quantitative and qualitative analytical techniques.

Given an abbreviated form by Lindquist (2009: 8-10), advantages of corpus linguistics that are presented in 1992 by Jan Svartvik are as follows.

- Corpus data are more objective than data based on introspection
- Corpus based data can easily be verified by other researchers, and researchers can share the same data instead of always compiling their own
- Corpus data are needed for studies of variation between dialects registers and styles
- Corpus data provide frequency of occurrence of linguistic items
- Corpus based data do not only provide illustrative examples, but are a theoretical resource.
- Corpus based data give essential information for a number of applied areas, like language teaching and language technology
- Corpora provide the possibility of total accountability of linguistic features- the analyst should account for everything in the data, not just selected features.
- Computerized corpora give researchers all over the world access to the data.
- Corpus data are ideal for non-native speakers of language.

In this study, rather than focusing on particular linguistic features, either a word or grammatical construction, impact of text mining and imitating strategies through using model essays on written productions of learners was analyzed. These strategies can be counted as reading strategies which are used as pre-writing activities in this study.

1.2.2. Reading and Writing Relationship

Since the emergence of writing as a distinct area of scholarship in the 1980s, factors that affect writing performance of learners have been investigated, and many theories have been developed by linguists such as Bazerman (1988), Bhatia (2004), Biber (2006).

The question of whether L2 reading skills are linked with L2 writing skills has been the topic of many research studies (Krashen 1993; Escribano 1999; Ito 2011). Krashen (1993) stated that writing style, the special language of writing, is acquired by reading. He supported his claim by pinpointing that the language acquisition process is similar to learning writing. He (1982) added that, “ Most of language acquisition takes place subconsciously, not through deliberate study, and it is a result of input (comprehension), not output (production) Thus, if you write a page a day, your writing style or your command of mechanics will not improve” (Krashen, 1993:27).

As learners are confronted with various characteristics of text in expository writing classes where the purpose is to inform, describe, explain, or define the author's subject to the reader, learners need to be provided with different writing styles and instructional implications of different texts (Hyland 2009). Grabe (2002: 14) stated that:

Once students (and teachers) begin to see how text information is organized in consistent ways, they will also begin to recognize all types of discourse signals as contributing to a small set of recurring patterns. These signals will no longer be seen as hundreds of individual and arbitrary cues (that can seem very confusing to students). Rather, they will be

seen as working together to help convey the larger text structure. Students will be able to connect information in ways that will assist their overall comprehension of texts. With improved overall comprehension of a text, it is also easier to focus on key grammatical complexities and vocabulary learning. Students will not be so lost with the text that they cannot focus on some difficult details. The text will then become a real context for guessing only partially known words .

Kennedy & Bolitho (1985: 85) claimed that there is a link between reading and writing in terms of signals and indicators. They are quoted as saying:

The actual content of the written text may be the same but the difference is that reading is concerned with the recognition of aspects of that structure, whereas writing has to do with the production of the text. In this respect reading may be regarded as a necessary precondition for any writing task, since the writer must be aware of the structure of a particular type of writing before he can produce it.

To a relation between reading and writing, Raimes (1983: 50) stated that :

A short story, a newspaper column, an advertisement, a letter, a magazine article, a poem, or a piece of student writing can work as a picture to provide shared content in the classroom. That is, readings can be used to create an information gap that leads communicative activities: if the students work with variety of readings at the same time, then they will be dealing with different content, and anything they write to each other will thus be authentic communication, conveying new and real information .

Reading and writing relationship may reveal factors that affect improvement in both skills. Besides this relationship, approaches to teaching writing should be mentioned to find out which approach includes reading in itself.

1.2.3. Approaches to Teaching Writing

This study investigated the possible impact of one of three major types of ESL /EFL writing approaches: product approach (Hyland, 2002; Silva, 1990). Other approaches are process approach (Brown, 2001; Silva, 1993;) and genre-based approach (Casanave, 2004; Hyland, 2003a). In the product approach, students are expected to produce texts that conform to the model texts provided by the teacher. Final products of students are important as students are taught to “develop competence in particular modes of written communication by deconstructing and reconstructing model texts” (Seifoori ;Christmas, 2011: 1).

The product approach comprises of four stages. At stage one, students analyze a model text and then they highlight the features of the genre. Stage two consists of controlled practice of the highlighted features, usually in isolation. At stage three, students organize their ideas. The organization of the ideas is more important than the ideas themselves and as important as the control of the language in this approach. Stage four is

the end of the learning process. At this stage, students choose one from many comparable writing tasks. Students individually use the skills, structures and the vocabulary they have been taught to produce the product in order to show they can be fluent and competent users of the target language (Steele, 2004). Hyland (2002: 5) categorizes general approaches to researching and teaching writing into three categories. Closely related with product approach, his first approach” focuses on the products of writing by examining texts in various ways, either through their formal surface elements or their discourse structure”. Teachers with this approach focus on the final piece of writing and measure it according to the criteria of “vocabulary use, grammatical use, and mechanical use as well as content and organization” (Brown, 1994: 320). With this approach, students are encouraged to mimic a model text, which is usually presented and analyzed at an early stage. Closely related with product approach but not encouraging students to produce “mindless copies of a particular organizational plan or style” (Eschholz, 1980: 24), few reading strategies are used to encourage students to write essays.

1.3. Statement of the Problem

Weigle (2002: 1) wrote that “The ability to write effectively is becoming increasingly important in our global community, and instruction in writing is thus assuming an increasing role in both second- and foreign- language education”. To write effectively, students need to use their vocabulary knowledge properly. However, as vocabulary is generally learnt through lists and memorization, students come up with short term memorizations and cannot use these words properly and actively in the long term use of the language learning process. In other words, learners have difficulties in adapting their vocabulary knowledge into writing or speaking skills.

Gorell (1987: 53) stated” unskilled writers do not have a clear sense of form. The sense for them is that sense is all mixed up with jumbled, half-remembered rules, unsuccessful trials at writing, heavily marked papers, and insufficient and ineffectual reading”. The question of what increases learner competence in the use of vocabulary for effective writing has been the focus of much research so far. As writing in a second language is different from writing in the first language, most of the research conducted on impact of L1 reading on L1 writing production fails to answer questions in EFL research. It has been put that there is a positive correlation between reading and writing in the target language; however, the question which reading strategies for writing should be used to

facilitate vocabulary use in written productions has not yet been put clearly. In order to understand how reading informs writing in terms of vocabulary and what happens when readers are also writers, students should be asked to use certain strategies called “text mining” and “imitating” to reach a conclusion about the usefulness of these strategies on vocabulary knowledge. These questions are important as they can lead to a shift in emphasis from teaching of reading and writing to the nature of learning how students use and adopt these strategies.

1.4. Purpose of the Research and Research Questions

The main aim of the current study is to analyze learner corpora to find out the possible impact of reading strategies called “text mining” and “imitating” on writing production in terms of the lexical richness, diversity and general success. Bearing this in mind, the research addresses the following questions:

1. How does “text mining” and “imitating”, used as prewriting activities, affect students’ writings in terms of:
 - The general lexical richness of the written output in terms of type/token ratio?
 - The use of nouns in frequency and form: singular or plural?
 - The use of adjectives in frequency and type: attributive, predicative and comparative + superlative adjectives?
 - The use of verbs in frequency of base form of a verb, past tense verb and progressive verb?
 -
2. General success in writing in terms of essay grading?

1.5. Operational Definitions

1.5.1. Lexical Richness

Lexical richness shows second language use in terms of the sophistication and range of an L2 learner’s productive vocabulary (Wolfe-Quintero; Inagaki; Kim, 1998). Intimate link between vocabulary and text length bring out the need to analyze indices of lexical richness. Measure of lexical richness is done depending on type token ratios (TTR), which are calculations with which number of different words (types) by total number of words (tokens) are divided in a given text. The idea behind this measure is that lexically richer texts are indicative of proficiency and larger lexicons. (Corssley et. al., 2010)

Therefore, TTR has been considered to be an important construct in L2 teaching and research. Lexical richness is directly related to the learner's ability to communicate effectively in both spoken and written form.

1.5.2. Lexical Diversity

Frequency and type of each word used in spoken or written corpora give information on lexical richness and diversity of a text. Lexical diversity (LD, also known as lexical variation) is defined as “the range and variety of vocabulary deployed in a text by either a speaker or a writer” (McCarthy and Jarvis, 2007: 459). High lexical diversity shows that student is not repetitive in his choice of a word. The richer lexicons used in the discourse, the higher degree of variations and sophistications perceived. (Zhang, 2013)

1.5.3. Learner Corpora

Electronic collections of spoken or written texts that are produced by foreign or second language learners are defined as computer learner corpora. Corpus-based study of learner language is a relatively recent development in corpus linguistics. Two types of corpora have been focus of SLA research so far: corpora by learners and corpora for learners. Sylviane Granger and her team developed the International Corpus of Learner English (ICLE) in 1998, which resulted in growing interest in producing corpora that can be used to study on interlanguage of learners. The general idea behind such work is that if learners' language is identified or analyzed, it may be possible to focus on teaching methods and contents in order to make teaching context more fruitful. Learner corpora can be compiled from both writing and spoken products of learners.

1.5.4. Concordances

Concordancing means finding every occurrence of a particular word or phrase. In other words, Lindquist (2009: 5, 7) defined a concordance as “a list of all the context in which a word occurs in a particular text”. He indicated that getting information in this way is easier and quicker than any other method. Besides being quicker than any other method, he added that “reading and analyzing concordances skillfully to extract the semantic and grammatical information hidden in them is an important aspect of corpus linguistics”.

O' Kafee et al (2007:8) state that:

The search word or phrase is often referred to as the 'node' and concordance lines are usually presented with the node word/phrase in the center of the line with seven or eight words presented at either side. These are known as Key-Word-In-Context displays (or KWIC concordances). Concordance lines are usually scanned vertically at first glance, that is, looked at up or down the central pattern, along the line of the node word or phrase.

Table 4: Selected concordance lines for globe in The New York Times, 1990, sorted on the first word to the left and the second word to the left.

321 ur months' time required to circle the	globe	and of the minimum \$30,000 to pay for it.
322 of stratospheric winds that circle the	globe	at the equator. These winds reverse direction
323 500 pounds each, that would circle the -	globe	several hundred miles up in so-called low
324 h hundreds of sailors have circled the	globe	alone and a dozen have made such a trip
325 last year Mr. Malamud has circled the	globe	seven times himself, logging 250,000 air
326 1976 Montreal Olympics. He circled the	globe	to hone his tactical skills and fitness.

Source: Lindquist (2009)

1.6. Assumption for the Study

It is assumed that students participating in the study are good at reading and comprehending what they read. It is also assumed that students took the advantage of strategies training on "text-mining" and "imitating" and applied what they learnt on their writings.

1.7. Limitations of the Study

Since the study was not conducted with a large number of students, the result of the investigation may deliver the reliable data about the topic but cannot be generalized. This research only involves first year students of the English Language and Literature

Department of Karadeniz Technical University. The students are both female and male within the age range of 18-24.

1.8. Organization of the Thesis

This thesis is introduced in five chapters. The first chapter presents the background to the study, purpose of the study and research questions, assumptions and limitations of the study, and operational definitions. Chapter two reviews the studies on theories on reading writing relationship, reading strategies text-mining and imitating, corpus linguistics and learner corpora. Chapter three defines the methodology of the research. Chapter four introduces the findings of the study, and chapter five delivers the discussion, conclusion and implications of the study.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Introduction

This chapter reviews studies in the field of learner corpora analysis and reading writing relations, strategies that connect reading with writing, lexical diversity and corpus linguistics.

2.2. Theories on the Relationship between Reading and Writing

Krashen's (1985) "Input Hypothesis" sheds light on reading and writing relationship. With this hypothesis, Krashen claims that language is acquired by understanding messages or obtaining comprehensible input. When it comes to second language learning, it holds that students acquire a second language through comprehensible input in the second language. According to this hypothesis, acquisition comes with exposure to large amount of meaningful, interesting or relevant L2 input material. Krashen's (1989) study on power of reading on language acquisition revealed that reading becomes a comprehensible input for language learners as long as texts are both interesting and understandable. The study indicated that reading exposure helps students increase both reading comprehension and vocabulary acquisition, and grammatical development and writing style. Krashen (1989: 109) states that "reading exposure is the primary means of developing language skills". That is, he (2004: 37-132) claimed that "reading is a powerful means of developing reading comprehension ability, writing style, vocabulary, grammar, and spelling" ; we acquire writing style, the special language of writing, by reading".

Focusing on the similarities between reading and writing, Cooper (1993) suggested that reading and writing should be taught together as they are both constructive processes, share similar processes and kinds of knowledge, improve achievement, foster communication, and lead to outcomes not attributable to either process alone.

Similar to Krashen's and Cooper's ideas on impact of reading on writing, some L2 writing researchers such as Grabe (2002), Paradi (2006), Coons (2008) who highlight the

relationship between reading and writing claim that L2 learners' writing ability can be facilitated by encouraging students to use model essays as they present rhetorical elements, principles and patterns of written discourse. According to such views, reading and writing are connected as they depend on similar knowledge representations, cognitive processes, contexts and contextual constraints. Therefore, it is not unusual to think that reading and writing development has close interaction, which may lead some pedagogical combinations that may be useful in making learning more efficient in. On reading and writing relation, Shanahan (1984: 466) stated that:

If reading and writing involve analogous cognitive structures and processes, it is possible that instruction in one would lead to increased ability in the other. Learning that occurs as a result of reading instruction could be useful in engendering writing achievement, and learning obtained through writing could be beneficial to reading development.

Biber (1986) state that different writing styles have different linguistic characteristics, but these styles overlap to a considerable extent. For example, so-called narrative style has some, but not all of the characteristics of formal, expository prose. Thus, Krashen, Lee and Ying indicated that reading anything will help writing, to at least some extent. In addition to that, Smith (1988:560) is quoted as saying: "To learn to write for newspapers, you must read newspapers; textbooks about them will not suffice. For magazines, browse through magazines rather than through correspondence courses on magazine writing ...To write poetry, read it".

Depending on the assumption that cognitive knowledge is shared by domains of reading and writing, (Fitzgerald and Shanahan, 2000), researchers have been investigating the relationship between reading and writing since 1930s (Nelson and Calfee, 1998). Research has focused on four basic types of knowledge. According to Fitzgerald & Shanahan, (2000: 40) the first type of superordinate knowledge is meta-knowledge which refers to several subcategories of knowledge, including being aware of the functions and purposes of reading and writing. The second one is domain knowledge about the substance and content, which refers to "all knowledge and awareness that readers and writers have, no matter what the source is". This knowledge type includes both "world knowledge" or "prior knowledge" and the knowledge that results from reading and writing interaction.

Knowledge of universal text attributes constitutes the third kind of superordinate knowledge which includes three subcategories as graphonics (letter and word identification), syntax, and another form of syntax. Graphonics include phonological

awareness, grapheme awareness, and morphology. While reading, readers must know how to deal with letters and phonemes and how they are combined. Likewise, writers must know how to deal with letters and sounds if they are supposed to spell correctly. To achieve this, readers and writers should increase their phonological awareness; develop the ability to perceive and discriminate letter shapes.

A second “subcategory of text attributes” is syntax which is defined in Fitzgerald & Shanahan, (2000: 40) by Kellogg (1994) as the rules or grammar for constructing sentence and for using punctuation. Readers and writers must know how to produce meaningful syntactic orderings of words and how to use punctuation. Even though oral language also has syntactic knowledge, some syntactic knowledge structures can only be found in text, which must only be learnt from reading texts. The third subcategory of text attributes is what we call as text format, which includes syntax of larger chunks of text, such as story grammars, and more general forms of text organization such as graphics (Shanahan, 1984 as cited in Fitzgerald and Shanahan, 2000).

The fourth kind of superordinate knowledge is procedural knowledge and skill to negotiate reading and writing (Langer, 1986 as cited in Fitzgerald and Shanahan, 2000). This type of knowledge refers not only to knowing how to access, use, and generate knowledge in any of the areas previously motioned, but also to the ability to integrate various processes such as recalling relevant information from memory, using intentional strategies such as prediction, questioning, or trying to find analogies. (Kellogg, as cited in Fitzgerald and Shanahan, 2000). These four subordinate categories of shared knowledge between reading and writing has attracted the attention of researchers who aimed to measure correlations among these knowledge types. The following table provides common ground for reading and writing.

Table 5: Common Ground between Reading and Writing

Strategic Processing	Writer	Reader
Searching for meaning	Generates ideas with an audience in mind	Uses print to construct meaning
Monitoring for meaning	Checks that the message makes sense	Checks that the message makes sense
Searching for structure	Anticipates the order of words based on how book language and oral language sound	Groups words together in phrases to represent the intended message
Monitoring for structure	Checks the order of words supporting the intended message	Rereads (out loud or holding the message) to check that the word order communicated the intended message
Searching for graphophonic information	Uses knowledge of how letters, words, and print work to record the message	Seeks out graphophonic input from print in relation to meaning and structure
Monitoring for graphophonic information	Checks and detects any discrepancies between anticipated message and graphophonic input	Checks and detects that the print represents the message
Self correcting	Detects and corrects	Detects and corrects

2.2.1 Previous Research on Reading and Writing Relation

Following Krashen's theory that holds writing skill is affected positively by extensive reading, a number of studies on reading and writing relationships in L2 have been conducted. Ito (2011) investigated relationship between L2 reading skills and persuasive essay quality in EFL Japanese high schools students. It was found that L2 reading has an effect on L2 writing quality. Özçelik (1996) sought an answer to the question whether the writing process of low level EFL students improves if it is taught

through reading with the help of reading texts. Participants of the study were 20 low level prep school students. He divided these participants into two and made one group an the experimental group and one group of a the control group. Participants in the the experimental group were exposed to pre-writing activities through reading with the help of reading texts but the participants in the control group was not exposed to pre-writing activities during the study. He found that reading had an effect on writing. That is teaching writing through reading text produced a significant increase in the composition profile total score of learners. Shanahan (1984) tried to find out the relationship between reading and writing by examining second and fifth graders phonic skills, reading comprehension, reading vocabulary, spelling, and prose writing. She analyzed writing samples for syntactic complexity, diversity of vocabulary, and organizational structure. She found that reading and writing measures were positively correlated, but only to a small or moderate extent. Shanahan concluded that the association between diversity of vocabulary in writing, and reading ability had an increase from second to fifth grade.

Based on Krashen's input hypothesis, Wai-King Tsang (1996) conducted research on effectiveness of an extensive reading programme and a frequent writing programme on the acquisition of descriptive writing skills in English. His participants were a group of Hong Kong secondary school students. His study highlighted the importance of linguistic input in the acquisition of writing abilities.

Parodi (2006) investigated reading and writing connection to identify psycholinguistic variables that account for the microstructural, macrostructural and superstructural levels of comprehension/production processing. He analyzed argumentative essays of 439 eight graders. Correlation results of the research showed significant coefficients between reading and students essay in all the psycholinguistic levels. Depending on these results, he concluded that the processes involved in both activities share some common knowledge-based strategies.

Perin (2006) conducted a research on reading-writing relationships in adult students who were health care workers attending an adult literacy programme. Participants reading different levels are measured with 5 points standardized adult literacy comprehension measure to compare productivity, syntactic complexity, single-word error type (base and function words, verbs, morphological suffixes, spelling),and use of apostrophes. Poor

readers compositions showed significantly lower syntactic complexity as measured by mean T-unit length. She found that poorer readers wrote in a less sophisticated manner.

2.2.2. Strategies that Connect Reading and Writing

Learners need to apply certain strategies to improve their writing skills. Taking this into account, many strategies that connect reading and writing have been theorized and put into practice. Carrell & Carson (1997: 55-56) stated that “Reading strategies must be taught that will enable learners to comprehend the text in a way that will allow them to produce an appropriate essay”.

Greene (1991) proposed a set of strategies that connect reading and writing and explored some ways students employ strategies that are involved in what he termed “mining” a text- reconstructing context, inferring or imposing structure, and seeing choices in language. Greene (1991) mentioned about two reading strategies for writing before defining the role of “text mining”. As to the question of how reading can inform writing, “imitation” and “immersion” strategies have been proposed to define the role of models of a well-wrought prose.

It is believed that students can learn about writing by imitating good models of written discourse. In such an approach, students are expected to internalize the style, grace, and correctness that make these works exemplary. Gorell (1987: 54) claimed that unskilled writers learn from imitation by focusing on form and structure while generating and finding the expression for their own ideas. By imitating, they learn to shape their sentences, develop their paragraphs, express their own voices, and perform many of the complicated tasks that writing process involves. In addition to this, he asserted that “when writers read, they pick up not only meaning but also the way in which that meaning is expressed.”. He also (1987) touched upon functions of imitation explaining one as having a problem solving capacity with which students make use of experience- one’s own and that of others to find solutions. Applied to writing, imitation means students do not need to invent a new form every time they want to express an idea. However, role of imitation has also brought some speculations as Greene (1991: 152) said:

One might wonder if students can articulate or apply the discourse knowledge they tacitly learn through imitation to their writing in different situations and across a number of varying tasks. Will imitation serve our students when they must transform their knowledge in order to contribute something new to an ongoing conversation in a given field?

Another expectation is that students will be able to write well by being immersed in what Atwell (1985; 1987) called “literature environment” (as cited in Greene 1991). With the immersion, students can see social purposes of writing and contribute to the development of community. Greene (1991) supported the view that employing these strategies students can only learn “the forms and genres and the ways of speaking that writing is a discipline demands” (Jolliffe and Brier, as cited in Greene 1991: 55). The question as to what extent these approaches teach students— individual writers— to negotiate the complex demands that a rhetorical situation places upon remains unanswered. Accordingly, Greene (1991) claimed that if students are expected to make reasonable choices and decisions in widely different rhetorical situations, abstract instruction will not suffice to them.

Greene’s (1991: 155) metaphor of “mining” holds that in order to achieve goals in composing, writers should read purposefully and intently. Tsui (2006) defines mining as a strategic approach that is used to dig out valuable language sources such as grammar and vocabulary. By using this strategy, writing and vocabulary skills of students are expected to improve as students pay attention to grammatical and lexical features of the texts, organization of the articles, and expressions which are unfamiliar to them. In addition to that this strategy is thought to help students improve their reading skills, and, at the same time, build the foundation of future writing. This kind of pragmatic reading, according to Greene, is fueled by three key strategies. Three keys strategies that can inform reading are “reconstructing context”, “inferring or imposing structure”, and “seeing choices in language”. As language is a lens through which people can understand something in a particular way, and even subtle changes in language can change the ways how meaning is located, a process which involves a plan, selective evaluation and organization of information in order to get a sense of the topography is required. So people can reflect upon one’s choices and decisions about the use of this accumulated knowledge to the best effect. Mining process is like an excavation during which miner uses certain tools that are convenient for the situation to help uncover what is most desired. This means that readers who are also writers use these strategies to reconstruct context, infer or impose structure, and see choices in language. This leads readers to make informed guesses about the use of the ideas or discourse features of a given text in light of his or her goals as a writer.

(Greene, 1991). Mining suggests a strategic process that can be considered to be the key factor in raising student awareness on how discourse patterns organize subset information.

Raimes (1983) stated that while reading a text, learners of a language not only interact with texts that somebody has written but also they can learn a great deal about writing. Depending on this fact, she suggested some reading strategies that help learners write through reading. Some of her techniques are (a)“examining cohesive links” with which students examine a piece of writing closely and make discoveries about the devices writers used to connect ideas and sentences to make the text cohesive; (b)“examining punctuation and grammar” with which students can make discoveries about the rules of grammar and punctuation; (c) examining sentence arrangement, with which students examine a text carefully to find out if the sentences go well according to basic principles that old information should be followed by a new one. These strategies of Raimes are related with what Greene called “mining a text” strategy with which students dig out each valuable information on a text to write.

2.3. Incidental Vocabulary Learning Through Reading

Importance of learning lexis in second language learning context for better and fluent productions in the target language gave rise to questions on how learners learn new words. Lewis (1993) suggested followings for language teachers and linguists.

- Lexis is the basis of language.
- Grammatical mastery is not a requirement for effective communication.
- Any meaning-centered syllabus should be organized around lexis rather than grammar.

In the light of what Lewis suggested, lexis constitutes important part of learning a second language. The relationship between reading and learning lexis is mentioned to foster vocabulary learning. Krashen’s Comprehension Hypothesis (Input Hypothesis) holds that reading helps learners acquire vocabulary, syntax, and spelling subconsciously. Pulido (2004: 20) asserts:

Reading is a complex cognitive process, one in which the reader, using previous knowledge, interacts with the information in the text to construct and integrate meaning (...). During reading there is simultaneous cognitive processing involving pattern recognition, letter identification, lexical access, concept activation, syntactic analysis, propositional encoding, sentence comprehension, intersentence integration, activation of prior knowledge, and comprehension monitoring.

Depending upon these explanations, many research have been conducted on the impact of reading on vocabulary acquisition. Positive effects of extensive reading were found by Book Flood studies which were conducted by Elley in 1991. These studies immersed learners in a great number of interesting texts. Lasting eight months, these studies showed dramatic improvements in many language skills including reading comprehension, knowledge of grammatical structures, word recognition, oral repetition, and writing (as cited in Waring and Nation, 2004).

Elgord and Warren (2014) investigated acquisition of second language (L2) vocabulary from reading a connected authentic text. The study revealed than number of encounters with new words in reading helped learners gain explicit word knowledge. However, advanced learners and lower proficiencies differ in gaining such knowledge. While extensive reading may be sufficient to sustain vocabulary development for advanced learners, lower proficiency readings needs to be supplemented with deliberate word learning and vocabulary learning strategy training.

Ponniah (2001) conducted a study on incidental vocabulary learning. He compared performance of the students who devoted their time to reading, and the students who learned meaning of words consciously to develop their vocabulary knowledge. Results of the study showed that the group who tried to learn subconsciously from their readings could use words that they have learnt in sentences while learners who spend their time learning meanings of words from dictionary could not use the previously unknown words in sentences. This study shows that words that are leant incidentally are retrieved better during writing process and learners can put their knowledge of new words into practice better when they read and learn a new word.

Erçetin (2012) investigated contribution of text comprehension, with different level of familiarity, on incidental vs. intentional vocabulary learning and found that text comprehension from less familiar texts resulted in only intake of unknown vocabulary while more familiar texts helped learners gain the unknown words. Retention of words from more familiar reading comprehension text was higher than that of from less familiar texts.

2.4. Lexicographic Research and Lexical Diversity (LD)

Research on meaning and use of words is related with lexicography. Biber (1998) mentioned about several advantages of corpus based studies that are eased with advances in computer technology. Apart from other uses of corpora, he focused on lexicographic

research on corpus as he claimed that computers have paved the way for storing and collecting larger corpora from variety of sources. Another advantage, he said, is being able to store complete texts or large chunks of text with aid of computers, which make analysis of these corpora not limited to sentence-length excerpts. In addition to that having chance to store more texts, corpora that are more representative of the overall language can be compiled. Reliability of the analyses is another factor that makes such analyses advantageous as unlike human readers, computers can analyze the pattern of word associations with more complex scale than a human does. Finally, together with all of these features Biber (1998: 22-24) stated that:

The greater size of corpora, their more representative nature, the more thorough and more complex analyses allow us to investigate a range of lexicographic research questions that were not feasible before. Today, corpus-based lexicographic investigations address six major types of research questions, each of which is illustrated among these questions are the following:

- What are the meanings associated with a particular word?
- What is the frequency of a word relative to other related words?
- What non-linguistic association patterns does a particular word have?
- What words commonly co-occur with a particular word, and what is the distribution of these “collocational” sequences across registers?

This kind of study seeks answers to “what is the frequency of a word relative to other related words?” Frequency and type of each word used in spoken or written corpora give information on lexical richness and diversity of a text. Lexical diversity (LD, also known as lexical variation) is defined as “the range and variety of vocabulary deployed in a text by either a speaker or a writer” (McCarthy and Jarvis, 2007: 459). Johansson (2008) states a highly lexically diverse text means that speakers or writers have to use many different words, with little repetition of the words that have already been used. Philip and Scott (2010) wrote that “ LD has been used by researchers in fields as varied as stylistics, neuropathology, language acquisition, data mining, and forensics; and LD indices have been found to be indicative of writing quality, vocabulary knowledge, speaker competence, Alzheimer's onset, hearing variation, and even speaker socioeconomic status (see Malvern, Richards, Chipere, and Durán, 2004; McCarthy and Jarvis, 2007: 1).” Johansson (2008) states a highly lexically diverse text means that speakers or writers have to use many different words, with little repetition of the words that have already been used.

Measuring lexical richness is generally concerned with how many different words are used in a text. There are different aspects of measuring lexical richness. Read (2000) defines them as lexical diversity which is the proportion of individual words in a

text; lexical variation that is the same as lexical diversity but focuses only on lexical words; lexical sophistication, the proportion of advanced words in a text; lexical density which is the proportion of lexical words in the whole text, and lexical individuality that is the proportion of words used by only one person in a group.

2.5. Corpus Linguistics

The word "corpus" refers to a large collection of texts presenting sample of a particular variety or use of languages that are presented in machine readable form. Corpus linguistics focuses on discovery of patterns of authentic language use through analysis of actual usage. (Krieger, 2003) Since 1960s computer corpora has had an impact on all language related research from lexicography to literary criticism through artificial intelligence and language teaching. This growing number of interest on computer linguistics brought the concept of "corpus linguistics" which Leech (1992: 106) referred as a 'new research enterprise', a different angle of thinking about the language.

Lindquist (2009: 1) states that corpus linguistics is not a "branch of linguistics on par with other branches, since "corpus" does not tell you what is studied, but rather that a particular methodology is used". Therefore, he defined corpus linguistics as a methodology that comprises a large number of related methods which can be used by scholars from different fields.

English is the language which has been continuously analyzed with corpus, which resulted in growth and diversification in English corpora. This growth in corpus size resulted in appearance of corpus types to cover wide range of varieties: diachronic, stylistic (spoken vs. written; general vs. technical) and regional (BritishAmerican, Australian, Indian, etc.) (Granger 1998) No attempt has been made to collect English learner data till recently. Only in 1990s, theoretical and practical potential of computer learner corpora has been recognized by academicians, EFL specialists and publishing houses alike and they began to initiate several projects (Granger 1998). As SLA research is concerned with 'the mental representations and developmental processes which shape and constrain second language (L2) productions' (Myles 2005: 374), and language production happens in the minds of learners; it is impossible to keep track of L2 interlanguage development of learners. It is for this reason that learner corpora, which help teachers analyze the interlanguage of learners with the help of corpora, have attracted such attention.

2.5.1. Learner Corpora

Developments in computer text-processing capabilities made it possible to investigate large samples of learner writing through corpus based research. In accordance with the development in corpus studies, learner corpus studies flourished, more and more has started to be discovered about the linguistic problems learners have.

Philip (2010) states that researchers have taken the advantage of pre-compiled corpora and have realized that they can compile their own corpora from the texts that are produced by their own learners. This realization brought the concept of analysis of learner corpora which is a key to determine the items of the language code that are mostly used by students. Granger (2003) defines computer learner corpora as “systematic electronic collections of spoken or written texts produced by foreign or second language learners” (2004: 124).

According to Granger (1998), the main purpose in compiling a learner corpus is to gather objective data for the description of learner language, which is considered imperative for valid theory and research in language learning and teaching. Such corpus is often analyzed by comparing learner data with that of native speakers. For example, researchers such as Granger 2002; Nesselhauf 2004; Flowerdew 1998; 2000; Aston 2000) tried to focus on quantitatively feasible trends in learner language in general at a given stage in the learning process and compare these trends with native corpora. For instance, Granger and Tyson (1996) compared adverbial connector usage of native and nonnative speakers of English and found that French learners of English focus more on the usage of connectors which add to, exemplify or emphasize a point; however, they use less connectors which are used for contrast or to take an argument logically forward. Aart & Granger (1998) studied on tag sequences in learner corpora. Their aim was to present the technique of automatic tag extraction and assess its potential for SLA research. Their analysis showed that the technique is able to highlight distinctive inter-language patterns, marking off those that are common to all learners from those that are specific to one learner.

Corpus studies on learner texts reveal frequency level and misuse of some word combinations which organizes a text, signal rhetorical relations, such as introducing a topic, contrasting, explaining, summarizing, concluding; ,or combine clauses within a text. Therefore, it is important to identify the aspects that might not be accepted or misunderstood by native speakers. (Campoy et al. 2010)

Campoy et al. (2010) mention about these research studies as not being sufficient to implement better instructions on language learning and teaching context. It is claimed that such analysis is not enough to design EAP teaching materials and must be complemented with studies analyzing students' interlanguage. Aston as cited in Campoy et al. (2010) gives two reasons why analysis of such corpora is not sufficiently qualified for syllabus design. First, the findings from such analysis 'provide no information as to the relative difficulty and learnability of particular features to be taught'; second, these findings are not of use in identifying the productivity of particular features from the learner's perspective. It is for this reason that he proposes using learner corpora to get relevant information on interlanguage development of learners, which can be used to code and classify recurrent errors, along with over-uses and under-uses, with a view to identifying features which teaching should perhaps emphasize and to evaluating their difficulty'. Therefore, Mukherjee and Rohrbach (nd) highlight the importance of individual learner's language and progression over general and supraindividual description of learner language. In learner corpus analysis, both theoretical foundations that are taken from corpus linguistics and second language acquisition investigations are used to analyze aspects of learner lexis, grammar, and discourse in order to investigate their links with L2 learning. Therefore, better insights into learner language will be gained for the benefit of language teachers. Leech (1997: 20) mentioned about the value of this kind of research for SLA studies, and emphasized its value in providing 'authoritative answers' for common types of language errors. Seidlhofer (2002: 125) is quoted as saying "[...] the most constructive way forward is to recognize and act upon the need for empirical classroom-based action research conducted by teachers who are aware of the potential as well as the limitations of corpus linguistic."

Studies using learner corpora can be listed under two categories: (i) *hypothesis-driven/corpus-based* studies and (ii) *hypothesis-finding/corpus-driven* studies (Granger 1998 (ed). The former is interested in testing specific hypothesis or research questions based on the literature on SLA research. Granger (1998: 15) states "The advantage of this approach is that the researcher knows where he is going, which greatly facilitates interpretation of the results. The disadvantage is that the scope of the research is limited by the scope of the research question". Hypothesis- finding research involves gathering data to see interlanguage of learners; following the data collection, researchers reach possible conclusions and hypothesis on the learner data. (as cited in Granger 1998: 132). Depending

upon analysis of papers collected in recent edited volumes within the field (e.g Granger et al. 2002 and Aston et al. 2004), it can be said that majority of studies within the area of learner corpus research fall within second (ii) category.

CHAPTER THREE

3. METHODOLOGY

3.1. Introduction

This chapter presents the research design, description of the participants, instrumentation, data collection procedure, and data analysis.

3.2. Research Design

This is an experimental study aiming to find out possible impact of certain reading strategies called “text mining” and “imitating” on lexical richness, diversity, and general success in learner compositions.

3.3. Participants

The participants of this study, 98 students ; 39 in the control group, 59 in the experimental group, were first year students of Karadeniz Technical University in English Language and Literature department in academic year 2012- 2013. The students, who were native speakers of Turkish were chosen with convenience sampling technique. Classrooms which could be easily accessed to were chosen for this study. Even though the participants chosen via convenience sampling might not represent the whole population (Paton, 2002), students who were eager to take part in the study were chosen. After having a year of preparatory class of intensive English courses: writing, speaking, listening, reading and grammar, students advanced to first year in their departments. Table 6 shows passing grades of students in the experimental and the control group.

Table 6: Independent Samples T test Results: Group Statistics for Passing Grades of Student

Group Statistics				
Group	N	Mean	Std. Deviation	Std. Error Mean
Grades experimental group	59	72,56	2,996	,390
control group	39	73,10	4,453	,713

Table 7: Independent Samples T test Results: Passing Grades of Student

Group Statistics				
group	N	Mean	Std. Deviation	Std. Error Mean
Rater1 experimental	20	54,25	9,187	2,054
control	20	52,10	9,492	2,122
Rater2 experimental	20	53,00	8,046	1,799
control	20	50,75	9,358	2,092

As it is seen from the table 6, there is no significant difference between means of these two groups. Regardless of their passing grades, they were divided to different classes randomly. In order to show equality of proficiency level of these students in writing skill, all students were assigned to write a narrative essay before the treatment. First drafts of their essays were compiled, and randomly selected 20 essays from each group were submitted two independent lecturers giving writing courses to grade these essays according to essay scoring rubric developed by Oshima and Hogue in 2006 (see appendix 3). Grades of students were computed with independent samples t test.

Table : 8 Independent Samples t-test results: Group Statistics for Narrative Essay Grading

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
grades	Equal variances assumed	,896	,346	-,105	96	,916	-,073	,698	-1,459	1,313
	Equal variances not assumed			-,110	91,786	,913	-,073	,670	-1,403	1,257

Table : 9 Independent Samples t-test results: Narrative Essay Grading

		Independent Samples Test									
		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
Rater1	Equal variances assumed	,004	,948	,728	38	,471	2,150	2,954	-3,830	8,130	
	Equal variances not assumed			,728	37,960	,471	2,150	2,954	-3,830	8,130	
Rater2	Equal variances assumed	,461	,501	,815	38	,420	2,250	2,760	-3,336	7,836	
	Equal variances not assumed			,815	37,165	,420	2,250	2,760	-3,341	7,841	

Independent samples t-test indicated there is no statistically significant difference between the experimental and the control group ($p > 0,005$). The analysis showed that students in both groups had nearly same writing performances. Another evidence that indicates students in both groups have similar proficiency level is their grades at the end of the preparatory class (see table 8).

In order to make sure that students understood how to apply text mining and imitating strategies, they were checked one by one while analyzing their model essays. If needed, researcher provided them with one by one training on using these strategies.

3.4. Instrumentation

For this study, participants were assigned to write one descriptive essay entitled “the person you admire” and one advantage disadvantage essay on “employee monitoring”. They were asked to write five paragraph essays with at least 250 words. Besides being instructed on how to mine and imitate their model essays, students in the experimental group used “Reason to Write Advanced Student Book” written by Miller, Cohen, Colonna and Gilbert as a reference book. This book contains short, thought-provoking reading passage for students to analyze and practice these strategies which help students proceed along a path from reading to writing. Students in the control group are instructed with power point presentation on how to write a descriptive and advantage and disadvantage essays. They are provided with charts and graphic organizers and instructions on rhetorical style and organization (see appendices 5-12 for sample student essays).

Their essays were compiled and tagged with Biber (1993) tags (see appendix 1). Bennet (2002: 14) stated that:

When a corpus is tagged, it means that each word included in the corpus has a marker added to it that gives additional information. Often, tags are part of speech markers, enabling users of corpora to search not only for specific words, but also for specific words used as a particular part of speech.

3.4.1. Parts of Speech Tagging (POS)

Using tags simplifies the corpora analysis process. Lindquist (2009: 43) gives example on how a tag can be of great use while analyzing part of speech in a corpora. She states that:

Let us say that you are interested in the lexical variation between the nouns *can* and *tin* referring to a metallic container in different varieties of English, and want to check for frequencies of these lexical items in a few different corpora. a. A search for the word forms *tin* and *can* in The New York Times, 1990, gives the frequencies 540 and 15,993, respectively. But if you look at the actual concordance lines, you will find that only about 0.2 per cent of the *can* examples refer to containers – all the others are instances of the modal verb *can*!

This is a problem that is encountered with in an untagged corpus. As searching for all frequencies for “can” will give you all instances of “can” whether a modal verb or a word that means metallic container. In order to keep analysis under control and to reach more reliable data, Parts of Speech Tagging was used to ease such analysis. In parts of speech tagging each word has a tag stating its parts of speech (word class). (Lindquist 2009)

One more problem comes with the size of the corpus in tagging process. A small corpora can be hand-tagged by an individual or a group of researchers. However, when it comes to larger corpora such as BNC, it is impossible to tag all word types by hand. It is for reason that automatic taggers have been utilized by researchers. “Such taggers can work with a more “fine-grained” classification of the words than just calling them nouns, verbs, adjectives or adverbs (Lindquist , 2009: 45) These automatic taggers often have 97/97 % success rate. Hence, Biber (1993) tags (see Appendix 4) were used in this study to find concordances of nouns and sub categories of nouns, adjectives and subcategories of adjectives, and verbs and sub categories of verbs were found with a software program called AntConc 3.2.4 which is used for carrying out corpus linguistics research and data-driven learning.

3.4.2. Corpus Description

Two parallel corpora of learners studying in the first year of English Language and Literature Departments were compiled in expository writing lesson. 39 essays acquired from the control group that constituted a corpus of 15826 tokens while 59 essays acquired from the experimental group which constituted a corpus of 25027 tokens. All texts in student corpus were grammatically annotated using an automatic grammatical “tagger” (a computer program developed and revised over ten year period by Biber) (see table 2). A large number of linguistic features in spoken (transcribed) and written forms are identified by this tagger. Tagging this corpus made it possible to conduct a series of more sophisticated analyses than would have been possible with an untagged corpus. Preparing student texts for the program was labor intensive and extremely time consuming. Spelling of each word is edited as not only English characters are required but also spelling of words should be correct. Each part of speech is tagged according to its classes. If a word has two functions, the one that is more commonly used is chosen to be tagged. For example, “name” has both verb function and noun function. As “name” as a noun is more used in dictionaries and corpus of native speakers, the programme tags it as a name. Following table shows sentences from tagged texts.

Table 10: Sentences from Tagged Texts

Sample Tagged Extract of an Essay: experimental Group	Sample Tagged Extract of an Essay: control group
Communication ^nn+nom+++=Communication technologies ^vbz+++??+=technologies which ^whp+who+++=which develop ^vb++++=develop with ^in++++=with science ^nn+nom+++=science and ^cc++++=and technologies ^nns+++??+=technologies advances ^nns+nom+++=advances make ^vb++++=make employers ^nns+nom+++=employers and ^cc++++=and employees ^nns+nom+++=employees	Some ^dti++++=Some thinks ^vbz+vprv+++=thinks that ^tht+vcmp+++=that there ^ex+pex+++=there are ^vb+ber+vrb++=are a ^at++++=a lot ^nn++++=lot of ^in++++=of advantages ^nns+nom+++=advantages. . ^zz++++=EXTRAWORD Some ^dti++++=Some thinks ^vbz+vprv+++=thinks that ^tht+vcmp+++=that

Table 10 (cont.)

Sample Tagged Extract of an Essay: The experimental Group	Sample Tagged Extract of an Essay: The control group
work ^nn++++=work easier ^jjr++++=easier. . ^zz++++=EXTRAWORD But ^cc+cls++++=But, , ^zz++++=EXTRAWORD	there ^ex+pex++++=there are ^vb+ber+vrb++++=are too ^ql++++=too much ^ap++++=much disadvantages ^nns+nom++++=disadvantages.

3.5. Data Collection

At the beginning of the term, students in the the experimental group were taught certain reading strategies called “text mining” and “imitating”. They were trained to work on model essays to gain insights into how and where writers use words where needed. For each class and topic, they were provided with model essays. By reading and mining these essays, they tried to produce their own compositions (see Appendix 5-12 for learner compositions)

The Best Pizza in Town and Maybe the World

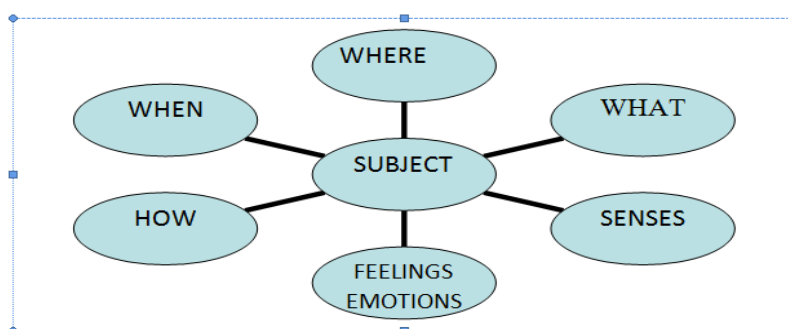
I have suffered a great deal because of a terrible addiction to pizza. Actually, I enjoy pizza too much. In fact, I enjoy it so much that I won't share it, not even with my mother. People often think of pizza as a junk food, because you can put almost anything on a pizza, but you can make it quite healthy if you want. People in my hometown of Cabimas, Venezuela, love it and call me the Pizza King of Cabima. Actually, it is a name that I am proud of. I have eaten pizza in many places, and none is as good as the pizza of Camibas. (Introduction paragraph of an essay; taken from Effective Academic Writing by Oxford University Press 2012: 4)

This is an example of mining an essay. Students can also imitate the rhetorical style and adopt sentences according to their own topic. For example, for a descriptive essay entitled “the person you admire” instead of underlined sentence, a student may write “I

have known many people in my life, but none is as devoted as my mother” for thesis statement.

On the other hand, students in the control group learnt writing essays with charts and organizers; they were not exposed to reading texts or analyzing texts. They were only taught how to write an essay, and were asked to write their own essays with these graphics. (See Appendix 1 and 2 for lesson plans).

Figure 2: Example of a graphic organizer for the control group



Students used these organizers to write their essays and learnt how to structure their descriptive essays (see table 11 for treatments)

Table 11: Treatments in the control and the experimental groups.

The experimental Group	The control group
<p>1. Training session:</p> <p>Students receive training on how to dig out valuable information from model essays. They are trained by power point presentation and hands on activity to be capable of using text mining and imitating strategies.</p> <p>2. Pre-writing stage:</p> <ul style="list-style-type: none"> • Students work on model essays by using text mining and imitating strategies. • Brainstorming • Planning <p>3Writing stage</p> <p>Writing the 1st draft Only first drafts of essays were compiled to avoid from teacher interference.</p>	<p>1. Pre-writing stage</p> <p>Brainstorming</p> <p>Planning</p> <p>2. Writing stage</p> <p>Writing the 1st draft</p>

The data for the writing were collected via writing tasks covering the topics that were assigned to them. Two parallel corpora of students in the control group and students in the experimental group were compiled and (39 essays in the control group (a corpus of 15826 words); 59 essays (a corpus of 25027 words) in the experimental group) these essays were tagged (with Biber tagger’s tag descriptions) in order to ease the analysis.

3.6. Data Analysis

Corpus-based studies often measure vocabulary richness in terms of Type-Token Ratio (TTR) in which the number of different words (types) a learner writes in a text is divided by the total number of words (tokens) in order to determine the degree of variation. However, this traditional model brings out some problems with it. Johansson (2008:63) mentions about this problem as follows:

A problem with the TTR measure is that text samples containing large numbers of tokens give lower values for TTR and vice versa. The reason for this is that the number of word tokens can increase infinitely, and although the same is true for word types, it is often necessary for the writer or speaker to re-use several function words in order to produce one new (lexical) word. This implies that a longer text in general has a lower TTR value than a shorter text, which makes it especially complicated to use TTR in developmental comparisons, e.g., between age-groups, where the number of word tokens often increase with age.

Table 12: Measures of Lexical Richness

Measure	Label	Formula
1. total number of words	tokens	N
2. number of different words	types	V
3. number of different dictionary entries	lemmas	
4. number of types occurring only once	hapaxes	
5. type/token ratio	TTR	V/N
6. corrected TTR	TTR(c)	$V/2\sqrt{N}$
7. 'Indice de Richesse'	Guiraud	V/\sqrt{N}
8. Index of Herdan	logTTR	$\log V / \log N$
9. Uber index	Uber	$(\log N)^2 / (\log N - \log V)$
10. 'theoretical vocabulary'	Menard	

As an alternative to TTR model, some models have been proposed to measure lexical richness (see table 10) “*Theoretical vocabulary*” is one of these models which has been proposed to (Broeder, Extra & van Hout 1986) measure word types in the samples. The principle behind this measure, as Johansson (2008) says, is to pick a number of words

(e.g 100 words) from a text randomly and calculate the number of word types in the sample. Therefore, theoretical vocabulary takes into account all possible ways of choosing 100 words from the text. By doing this, one can easily compare texts of different lengths with limiting number of words by random selection.

In order to find out lexical richness, this theoretical vocabulary model was used. 200 words are chosen randomly from all essays. These essays are analyzed in terms of Type-Token Ratio. To measure lexical density, tagged files were analyzed by corpus analysis tools named AntConc 3.2.4 which is used for carrying out corpus linguistics research and data-driven learning. It helps researchers with a comprehensive set of tools including a powerful concordancer, word and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot.

Written productions of learners are compiled and tagged. AntConc 3.2.4 was performed to find frequencies of parts of speech. Obtained frequencies were assessed and compared with Log likelihood ratio to find statistical differences, if any, between the frequency of nouns, verbs, and adjectives. In corpus studies, when comparing different sized datasets, chi-square value has often been performed to compare word frequencies across corpora; however, Rayson and Garside (2000) stated that log-likelihood tests are considered to have higher reliability than other statistical methods. Significance difference is tested by log-likelihood ratio which computes overuse and underuse of words. If the log-likelihood ratio is ± 3.84 or more a significant difference exists between the two datasets at a 5% significance level. Rayson and Garside (2000: 40) described log-likelihood ratio as follows:

“Log Likelihood has] 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 Words which appear with roughly similar relative frequencies in two corpora appear lower down the list.

General success of students is evaluated by two raters. These raters independently read the essays according an essay scoring rubric (developed by Oshima and Hogue 2006). Peat (2006) suggested that, because of their explicitly defined criteria, rubrics lead to increased objectivity in the assessment of writing. Results of this grading were analyzed in terms of reliability first. This analysis is done with SPSS 16. Program and Cronbach Alpha's of these grades are found.

CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Introduction

This chapter is going to introduce results and the discussion of the study. The data for this study was collected from two groups of participants: participants in the the experimental group (n=59) and participants (n=39) in the the control group. Their essays were compiled into a written learner corpus in the spring semester of academic year 2012-2013. Participants in the the experimental group were given instructions on writing essays with reading, focusing on strategies such as text-mining and imitating, as a pre-writing activity; while participants in the the control group were only instructed on how to write essays through power point presentations which included important points to consider while writing different types of essays. The students in the the control group did not carry out any text-mining strategy exercises or imitating exercises as prewriting. Participants in the the experimental group were provided with model essays. They tried to dig out valuable information on these essays in order to improve their writing. Their reference book was “Reason To Write” by (Colonna and Gilbert, 2006) in which they read reading passages and did many exercises related to reading while participants in the control group were only instructed on how to write essays without having model essays and exercises. Only first drafts of their essays were compiled, as this study is concerned with the impact of reading strategies on lexical richness and diversity in essays of participants, so that teacher interference on use of vocabulary was kept under control. In order to analyze their essays, each essay was controlled for Turkish characters to aid the computerized tagging process. Occurrences of Turkish characters should be avoided or changed to English as the process of tagging only requires English characters. Final versions of their essays were sent to USA to be tagged through the Biber (1993) tagger program at the Corpus Lab of Northern Arizona University. For example, regarding the word categories focused on in this study; for adjectives he used tag “jj”, for nouns “nn”, and for verb “vb”. Each general category is divided in its subcategories and each different subcategory is tagged with a special symbol. For the present study, three subcategories of nouns, verbs and adjectives were analyzed and compared.

Table 13: Tags for adjective category

jj+atrb+++	adjective + attributive function
jj+atrb++xvbg+	adjective + attributive function + + -ing form
jj+atrb++xvbn+	adjective + attributive function + + past participle form
jj+pred+++	adjective + predicative function
jj++++	adjective + indeterminate function
jjb+atrb+++	attributive-only adjective + attributive (chief, entire)
jjr+atrb+++	comparative adjective + attributive function
jjr+pred+++	comparative adjective + predicative function
jjt+atrb+++	superlative adjective + attributive function

Source: Adapted from Biber (1993)

All adjective concordances were found by analyzing the tagged version of the student essay corpus with AntConc 3.2.4 with the tag “jj”. Three types of adjectives in students’ essays were analyzed in specific: adjective + attributive function, comparative adjective + attributive function, superlative adjective + attributive function. The same procedure was applied for nouns as well.

Table 14: Tags for Noun category

nn++++	singular common noun
nn+nom+++	singular noun + nominalization
nvbg+++xvbg+	singular noun + + + -ing form
nn+++xvbn+	singular noun + + + past participle form
nns++++	plural common noun
nns+nom+++	plural noun + nominalization
nnu++++	unit of measurement (lb, kg, ...)
np++++	singular proper noun
nps++++	plural proper noun
npl++++	locative noun
npt++++	singular titular noun
npts++++	plural titular noun
nr++++	singular adverbial noun (east, west, today, home, ...)

nrs++++	plural adverbial noun
---------	-----------------------

Among all these noun types, only three of them were analyzed in learner essays. These types are singular common noun, singular noun + nominalization and plural noun + nominalization.

Table 15: Tags for Verb category

vb++++	base form of verb, excluding verbs in infinitive clauses (uninflected present tense, imperative)
vb+++xvbn+	base form of verb + + + past participle form (e.g., cut, hit, hurt, ...)
vb+be+aux++	base form of verb + be + auxiliary verb
vb+be+vrb++	base form of verb + be + main verb
vb+bem+aux++	verb + am + auxiliary verb
vb+bem+aux++0	verb + am + auxiliary verb + + contracted ('m)
vb+bem+vrb++	verb + am + main verb
vb+bem+vrb++0	verb + am + main verb + + contracted ('m)
vb+ber+aux++	verb + are + auxiliary verb
vb+ber+aux++0	verb + are + auxiliary verb + + contracted ('re)
vb+ber+vrb++	verb + are + main verb
vb+ber+vrb++0	verb + are + main verb + + contracted ('re)
vb+do+aux++	verb + do + auxiliary verb
vb+do+vrb++	verb + do + main verb
vb+hv+aux++	verb + have + auxiliary verb
vb+hv+aux++0	verb + have + auxiliary verb + + contracted ('ve)
vb+hv+vrb++	verb + have + main verb
vb+hv+vrb++0	verb + have + main verb + + contracted ('ve)
vb+seem+++	base form of verb + seem / appear
vb+vprv+++	base form of verb + private verb (believe, feel, think, ...)
vb+vprv+tht0++	base form of verb + private verb + that deletion **
vb+vpub+++	base form of verb + public verb (assert, complain, say, ...)
vb+vpub+tht0++	base form of verb + public verb + that deletion **
vb+vsua+++	base form of verb + suasive verb (ask, command, insist, ...)

Source: Adapted from Biber (1993)

There were 59 essays with 25027 tokens in the experimental group, while there were 39 essays with 15826 tokens in the control group. As each essay has different word count, the percentage of concordances of all nouns, adjectives, and verbs were taken to norm the word counts. So, in order to make the word counts comparable across the two corpora: the experimental group and the control group, the word counts were normed to out of 1000 words. That is, rather than obsolete frequencies, frequencies of two corpora were given per 1000 words. This method is called normalizing, and according to Lindquist (2009) it should always be used when frequency of two corpora of different sizes are compared so that results from a corpus of a different size will be comparable with the results of another corpus of different size. For example:

Table 16: Example of Norming

File Name	Tag	Concordance	Word count	Normed Frequency
con.adv.009	rp++++	1	360	2,777777778
con.adv.010	rp++++	5	699	7,153075823
con.adv.011	rp++++	3	497	6,036217304
con.adv.012	rp++++	1	541	1,848428835

Obtained frequencies were assessed and compared with Log likelihood ratio to find statistical differences, if any, between the frequency of nouns, verbs, and adjectives.

4.2. Results

The purpose of the study was to analyze and compare written corpora of two groups of students, those in the experimental group and those in the control group. First, lexical richness of two parallel corpora were analyzed and compared. Three types of nouns in the noun category, three types of adjective categories, finally, general verbs category were analyzed and compared.

4.2.1. General Lexical Richness

Lexical richness is generally measured to assess lexical proficiency level of a child or a student. In such assessment, lexical richness is generally compared with an external reference point such as age, gender or culture. In this study, a treatment on reading-writing relationship by using text mining and imitating strategies as prewriting activities was taken as a reference point to compare lexical richness of samples produced by the experimental group and the control group.

Lexical richness of the samples is evaluated by using a type-token ratio measure (TTR). In TTR measure, a number of different words (types) over the total number of words (token) in a text are calculated. Due to criticism that holds this ratio is text length dependent, or, in other words, longer texts have less type token ratio than shorter texts, many adjustments on TTR measure have been proposed. One of these adjustments is “*Mean Segmental TTR*” (MSTTR) which was proposed by Engber in 1995. This model calculates the mean TTR of consecutive text sections of equal length. Therefore, the result may give more reliable results. The other adjustment is “*theoretical vocabulary model*” which was proposed by Broeder, Extra and Van Hout in 1986. In this model, similar number of words from texts are chosen randomly to calculate TTR measure, so that text length will be equal which helps eliminate the criticism that holds longer texts have less TTR than shorter texts by having the equal word counts in each essays. Depending on this model, 200 words of each essay were chosen randomly in order to make the text length equal. By doing this, it is believed that TTR, if performed on randomly sampled corpus of equal length, will give consistent result for the corpus under analysis.

Type token ratio of each essay assessed and data from the experimental group and the control group were compared with SPSS 16. Independent samples T-test.

Table 17: Type Token Ratio of Essays: Group Statistics

Group Statistics				
group	N	Mean	Std. Deviation	Std. Error Mean
Typetokenratio experimental	59	,5919	,06021	,00784
control	39	,5967	,06563	,01051

Table 18: Independent Samples T-test Results : Type Token Ratio of Essays:

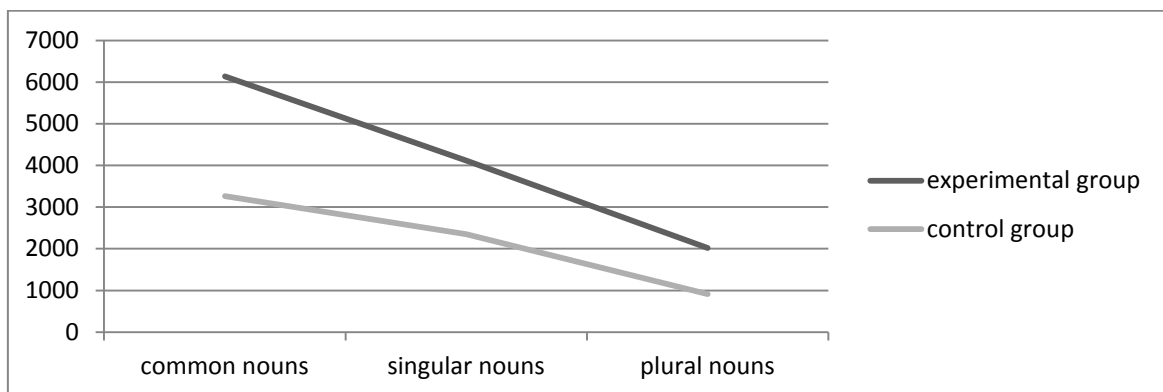
		Independent Samples Test								
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
TypeToken ratio	Equal variances assumed	,407	,525	-,373	96	,710	-,00480	,01288	-,03037	,02077
	Equal variances not assumed			-,366	76,532	,715	-,00480	,01311	-,03091	,02131

It has been found that there is no statistical significant difference between two groups in terms of lexical richness.

4.2.2. The Use of Nouns in Frequency and Form: Singular or Plural

Nouns are one of the most frequent words that were used both in essays of the experimental group and the control group. Out of 25027 words with 6134 concordances, common nouns constitute 24,5 %of corpus of the experimental group. Percentage of singular nouns, with 4115 concordances, is 16,5 and %8 of corpus is plural nouns, with 2019 concordances; while, in the control group common nouns are %20,5,with 3260 concordances, singular nouns are %14,8, with 2351 concordances, and plural nouns are %5,7, with 909 concordance.

Figure 3: Noun distribution according to its subclasses.



The purpose of the research was to find possible impact of reading strategies on the use of lexical items in student writing. Students in the experimental group were instructed on how to search for and analyze valuable information in model essays before they start producing their own texts. They were taught to mine their texts, in other words break their model essays into pieces to understand how words are used and how they can use these words in their essays. However, students in the control group were only instructed on how to write an essay without using any reading passages as a pre-writing activity. Frequency of nouns in two corpus of students were computed with log-likelihood ration in order to compare and see the results of the treatment.

Table 19: Log-likelihood ratio (LL) of overall frequency of Nouns among the experimental and the control group corpora.

Item	O1	%1	O2	%2	LL
Common Nouns	6134	24.51	3260	20.60 +	65.54
Singular Nouns	4115	16.44	2351	14.86 +	15.55
Plural Nouns	2019	8.07	909	5.74 +	75.21

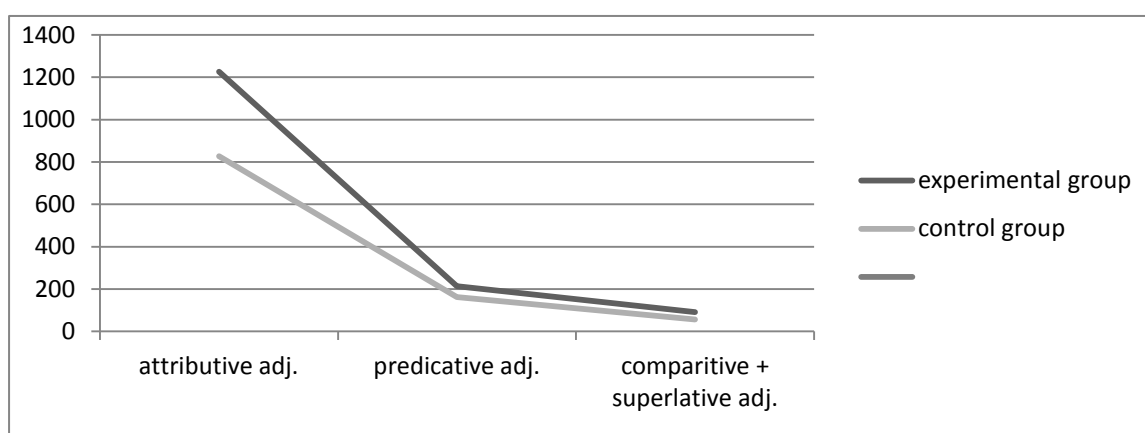
Writing samples of two groups of participants were analyzed. Concordances of singular common noun, singular noun + nominalization and plural noun + nominalization were found. The results of Log-likelihood ratio showed that there is a significant difference between writings of students in the experimental group and the control group. (LL=65.54 p < 0.05 (critical value: 3.84) + indicates overuse in the first corpus relative to the second corpus). That is, students in the the experimental group used significantly more nouns compared to students in the the control group. Depending on the findings in the noun frequency, it can be said that reading strategies, text mining and imitating, had a fostering impact on student use of common nouns in their essays.

4.2.3 The Use of Adjectives in frequency and type: attributive, predicative and comparative + superlative adjectives.

Third research question was “How does “text mining” and “imitating” used as prewriting techniques, affect student writing in terms of use of adjectives in frequency and type: attributive, comparative and superlative adjectives.

Figure 4 shows adjective distribution according to two groups of students. As seen in Figure 4 nearly all adjective occurrences are similar in two groups.

Figure 4. adjective distribution according to its subclasses



Attributive adjectives constitute 4.9 % of corpus of the experimental group with 1226 concordances out of 25027 words while this number is 5.23% in the control group. Adjectives with predicative function constitutes 0.86% in the experimental group with 214 concordances; 1.03% in the control group. Frequencies of comparative and superlative adjectives were computed together and results showed that frequency of these adjectives is 0.36% in the experimental and 0.35% in the control group.

Table 20: Log-likelihood ratio (LL) of overall frequency of adjectives among the experimental and the control group corpora.

Item	O1	%1	O2	%2	LL
Attributive adjectives	1226	4.90	828	5.23-	2.13
Predicative adjectives	214	0.86	163	1.03-	3.17

Comparitive +Superlative adjectives	91	0.36	56	0.35 +	0.03

Log-likelihood ratio results indicated that there is no significant difference between the experimental and the control group in terms of use of adjectives in their written productions (LL = 2.13, 3.17, 0.03 < 3.84)

4.2.4 . The Use of Verbs in Frequency and Type: Base Form of a Verb, Past Tense Verb and Progressive Verb.

Verbs have been investigated in writing samples in terms of base form of a verb, past tense verbs and progressive verbs. Base form of verbs constitute 7,86 % of corpus of the experimental group with 1967 concordances while this number is 7,09 % in the corpus of the control group with 1122 concordances. Students in the experimental group used past tense verbs 966 times, % 3,86 of corpus, while students in the control group used it 633 times, % 4 of their corpus. Progressive verbs were used 286 times (%1,14) by the the experimental group; 129 times (%0.82) by the the control group.

Figure 5: General verb distribution according to its subclasses

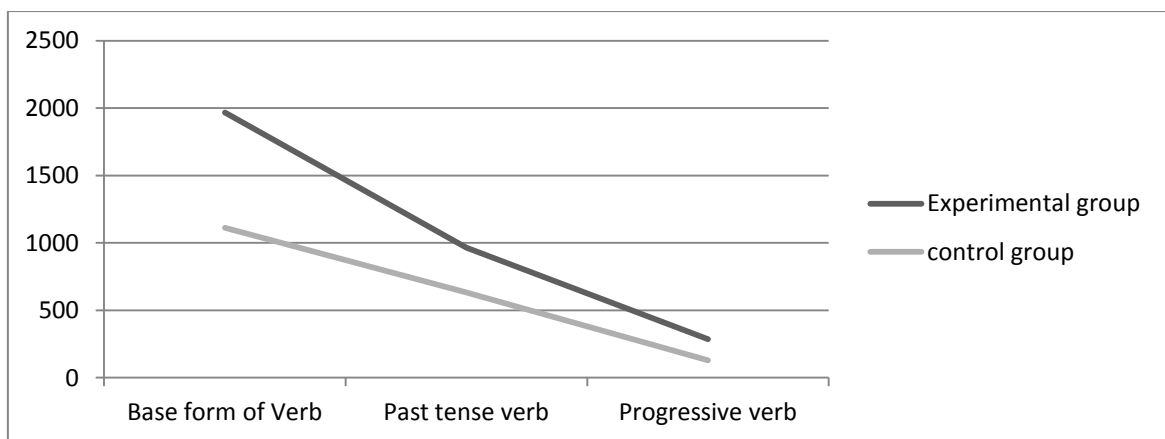


Table 21: Log-likelihood ratio (LL) of overall frequency of verbs among the experimental and the control group corpora.

Item	O1	%1	O2	%2	LL
Base form of verbs	1967	7.86	1122	7.09 +	7.66
Past tense verb	966	3.86	633	4.00 -	0.48
Progressive verb	286	1.14	129	0.82 +	10.55

Log-likelihood ratio showed that there is statistically significant difference between two groups of students in terms of use of base form of verbs ($LL=7.66 > 3.84$) and use of progressive verb ($LL= 10.55 > 3.84$) in their essays. However, it was found that there is no statistically significant difference between two groups in terms of use of past tense verbs.

4.2.5. General Success in Terms of Grading

The last research question was related to general success of students in writing in terms of essay grading. Two independent raters graded papers according to Essay Scoring Rubric (developed by Oshima, Houge and Buttler in 2006. To measure of internal consistency of two raters, in other words, to measure how closely related grading of two raters as a group, Cronbach's Alpha was found by using Spss 16. in gradings of two raters.

Table 22: Reliability Statistics for two raters.

Reliability Statistics

Cronbach's Alpha	N of Items
.949	2

As it seen above, the alpha coefficient for the two items is .949, suggesting that the items have relatively high internal consistency.

Table 23: Independent Sample T-test result: Group Statistics for General Success Rates of Students

Group Statistics

group	N	Mean	Std. Deviation	Std. Error Mean
rater1 control	30	62,87	9,047	1,652
the experimental	30	73,33	5,833	1,065
rater2 control	30	60,73	10,654	1,945
the experimental	30	74,33	7,576	1,383

Table 24: Independent Sample T-test result General Success Rates of Students

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
rater1	Equal variances assumed	7,602	,008	-5,326	58	,000	-10,467	1,965	-14,401	-6,533
	Equal variances not assumed			-5,326	49,558	,000	-10,467	1,965	-14,415	-6,518
rater2	Equal variances assumed	4,282	,043	-5,698	58	,000	-13,600	2,387	-18,378	-8,822
	Equal variances not assumed			-5,698	52,357	,000	-13,600	2,387	-18,389	-8,811

Independent samples T-test indicated that there is significant difference between general grading of papers in the control group and the experimental group. Mean of two group shows that students in the experimental group have higher grades. This means that treatment on the experimental group worked on students' general success.

CHAPTER FIVE

5. CONCLUSION

5.1. Introduction

This study was carried out to investigate the impact of reading strategies, “text mining and imitating”, on EFL learners’ , majoring in English, improvement in vocabulary use in their essays. The participants of this study were first year students attending the first year of English Language and Literature Department of Karadeniz Technical University. Writing skill, just like speaking skill, when compared to the other three skills requires more time and practice to be improved. It is thought that in order to improve student proficiency in the use of parts of speech (nouns, verbs, adjectives), the influence of text-mining and imitating was considered to be worth focusing on. The participants of the the experimental group were trained on using text mining and imitating reading strategies. In order to collect the data that would be concentrated on to make the conclusions for the study, students are asked to write two kinds of essays: advantage and disadvantage and descriptive essays. In accordance with the data gathered through the essays, the research questions below were targeted to be answered:

How does “text mining” and “imitating”, used as prewriting techniques, affect student writing in terms of:

1. The general lexical richness of the written output in terms of type/token ratio?
2. The use of nouns in frequency and form: singular or plural?
3. The use of adjectives in frequency and type: attributive, predicative comparative + superlative adjectives?
4. The use of verbs in frequency of base form of verbs, past tense verbs, and progressive verbs?
5. General success of student in terms of essay grading?

5.2. Conclusions

The data collected by essays were analyzed so as to find out the answers of the research questions that were determined before the start of the study.

1. How does “text mining” and “imitating”, used as prewriting activities, affect student writing in terms of: The general lexical richness of the written output in terms of type/token ratio?

The purpose of this study was to examine the relationship between lexical richness and reading model essays by using text mining and imitating strategies as a prewriting activity to write essays order to shed light on the impact of these strategies on lexical richness.

The statistical analysis of two parallel corpora indicated that there is no significant difference between two groups in terms of type token ratio. As lexical richness is directly related with the quality of written and spoken language produced by learners of second language, at the start of the study, it was assumed that students in the experimental group will have better lexical richness ratio. The “mean” of statistical analysis of the experimental group is higher than that of the control group, which shows there is a slight difference between these two groups. Having similar findings, Laufer (1991: 445) examined written compositions of advanced L2 learners of English and found no improvement in lexical richness over two semesters. Upon this result, Laufer proposed the “active vocabulary threshold hypothesis” which assumes that “passive vocabulary knowledge may continue to develop throughout the lifespan, but “our productive lexicon will grow only until it reaches the average level of the group in which we are required to function.”

From a pedagogical point of view, it can be said that as model essays provide learners with already written input, it will be beneficial for them to dig out valuable information and imitate certain phrases or chunks by adjusting them in their own context.

2. How does “text mining” and “imitating” ,used as prewriting activities, affect student writing in terms of use of nouns in frequency and form: singular or plural?

The result showed that there is a statistically significant difference in terms of general nouns between the experimental group and the control group. When zoomed in on

finer-grained subclasses, statistical analysis also showed that there is significant difference between two groups in terms of use of plural and singular nouns. It can be concluded that reading strategies such as text mining and imitating help learners use more nouns in their essays.

This result shows that it will be beneficial for teachers of English to provide their learners with model essays by using text mining and imitating strategies as a pre-writing activity.

3. *How does “text mining” and “imitating”, used as prewriting techniques, affect student writing in terms of use of adjectives in frequency and type: attributive, predicative and comparative + superlative adjectives?*

The results of the statistical analysis showed no significant difference between two groups in terms of use of attributive, predicative and comparative+ superlative adjectives.

4. *How does “text mining” and “imitating”, used as prewriting activities, affect student writing in terms of use of verbs in frequency of base form of a verb, past tense verb and progressive verb.*

Statistical analysis of general verb category showed a significant difference in terms of use of base form of verbs and progressive verbs between two groups of students. Depending on the data, it can be concluded that students using text-mining and imitating strategies as a prewriting activity, used more verbs than those in the control group.

Concordances of all parts of speech were found and analyzed in order to find, if any, differences between the experimental group and the control group. It can be concluded that most categories of parts of speech are more frequent in essay corpora of the experimental group. Treatment for this research seems to work to a certain extent. However, if applied in a longer time, it carries the potential of yielding to more significant results.

5. *The last research question was how “text mining” and “imitating”, used as prewriting activities, affect students’ writings in terms of general success in writing.*

Results of the analysis showed that students in the experimental group have better grades than students in the control group. Independent samples T-test showed that there is significant difference in terms of grading between these two groups. This result can be interpreted as follows. Students studying on model texts and trying to imitate certain structures in their essays succeeded to write better.

5.3. Limitations and Suggestions

In this study, impact of reading strategies on writing performance of students, on lexical items in specific, was investigated. Taking number of students (59 in the experimental; 39 in the control group) into account, the results of the study cannot be generalized. However, this study may give a reference point for further research. It is believed that replications of this study with more students at different age levels and educational backgrounds in various ELT contexts such as compulsory service English classes and preparatory English courses will contribute to the field. This study lasted two months, if it were longer, it may have given more significant results as means of two groups indicate that students in the experimental group use higher rate of parts of speech.

Further research might be conducted to shed light on the use of all categories of parts of speech and discourse markers.

REFERENCES

- Aart, J. and Granger, S. (1998), **Tag sequences in learner corpora: a key to interlanguage grammar and discourse**. In Granger, S. (Ed.) *Learner English on Computer*. Addison Wesley Longman: 132-141.
- Albrechtsen, Dorte et al. (2008), **Vocabulary and writing in a first and second language: Process and development**, Basingstoke: Palgrave Macmillan.
- Armbruster, B. Bonnie et al. (2003), **Put reading First: The research building blocks for teaching children to read: Kindergarten through grade 3**, Jessup, MD: National Institute for Literacy.
- Aston, G. (2001), **Learning with Corpora**. Bologna: CLUEB and Houston, TX: Athelstan. In Campoy, Mari Carmen et al. (2010) *Corpus-Based Approaches to English Language Teaching*.
- Baker, P. (2010), **Sociolinguistics and Corpus Linguistics**, Edinburgh University Press.
- Bazerman, C. (1988), **Shaping written knowledge**, Madison, WI: University of Wisconsin Press.
- Bennet, G.R. (2010), **Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers**, Michigan ELT.
- Biancarosa, G., and Snow, C. (2004), **Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York**, Washington, DC: Alliance for Excellence in Education.
- Biber, D. (1986), *Spoken and Written Textual Dimensions in English. Resolving the Contradictory Findings*, **Linguistic Society of America**.
- _____ (2006), "Stance in spoken and written university registers", **Journal of English for Academic Purposes**, 5(2), 97–116.
- Biber, Douglas et al. (1998), **Corpus Linguistics. Investigating Language Structure and Use**, Edinburgh, U.K: Cambridge University Press.
- Bhatia, V.K. (2004), **Worlds of Written Discourse: A Genre-Based View**, London: Continuum.
- Brown, H. D. (1994), **Teaching by principles: An interactive approach to language pedagogy**, Englewood Cliffs, NJ: Prentice Hall Regents.

- Campoy, Mari Carmen et al. (2010), **Corpus-Based Approaches to English Language Teaching**, New York: Continuum International Publishing Group.
- Carrell, P. and J. Carson (1997), “Extensive and Intensive Reading in an EAP setting”, **English for Specific Purposes**, 16 (1), 47-60.
- Casanave, C.P (2004), **Controversies in Second Language Writing: Dilemmas and Decision in Research and Instruction**, Ann Arbor, MI: University of Michigan Press.
- Celce-Murcia, Marinna et al. (1999), **The grammar book: An ESL /EFL teacher's course (2nd ed.)**, Boston: Heinle & Heinle.
- Christmas, B. (2011), **The role of brainstorming in improving students writing performance in the EFL classroom**
<http://sydney.edu.au/cet/docs/research/The%20Role%20of%20Brainstorming%20in%20Improving%20Students%20Writing.pdf> (12.10. 2012).
- Cunningham, P.M. and Allington, R.L. (2003), **Classrooms that work: They can all read and write**, 3rd edition, Boston: Pearson Education, Inc.
- Cooper, J.D. (1993), **Literacy: Helping Children Construct Meaning** (2nd ed), Boston: Houghton Mifflin.
- Crossley Scott A. et al. (2010), Predicting lexical proficiency in language learner texts using computational indices, **Language Testing**.
- Durkin, D. (1988), **Teaching them to read** (5th ed.). Boston: Allyn & Bacon.
- Ellis, N. C. (2002), “Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition”, **Studies in Second Language Acquisition**, 24, 143–188.
- Elgort, I and Warren P. (2014), “L2 Vocabulary Learning from Reading: Explicit and Tacit Knowledge and The Role of a Learner and Item Variables”, **A journal of Research in Language Studies. Language Learning**, 64 (2), 365–414
- Engber, C. (1995), “The relationship of lexical proficiency to the quality of ESL compositions”, **Journal of L2 Writing**, 4 (2), 138-155.

- Eschholz, P. A. (1980), **The prose models approach: Using products in the process.** in: **T. R.** Donovan, and B. W. McClelland, (Eds.) *Eight Approaches to Teaching Composition*. Urban IL: Council of Teachers of English.
- Erçetin, G. (2012), The role of text comprehension, L2 reading proficiency, incidental vs. intentional vocabulary exposure in intake, gain, and retention of target vocabulary. Published Master Thesis. Retrieved from **Turkish Council of Higher Education National Database**.
- Escaribano, P.D. (1999), *Teaching writing through reading: a text-centered approach*, **Universidad Politécnica de Madrid**.
- Eschholz P.A. (1980), **The prose models approach: Using products in the process.** In T.R. Donovan and B.W. McClelland (Eds.), *Eight Approaches to Teaching Composition*. Urbana: National Council of Teachers of English.
- Flowerdew, L. (1998), **Applying corpus linguistics to pedagogy: A critical evaluation**, Hong Kong University of Science and Technology, “*International Journal of Corpus Linguistics*”, 14(3), John Benjamins Publishing Company.
- Fitzgerald, J. and Shanahan, T. (2000), Reading and writing relations and their development, **Education Psychologist**, 35(1), 39-50.
- Gorell, D. (1987) “Freedom to Write through Imitation”, **Journal of Basic Writing**, 6 (2).
- Grabe, W. (1997), **Discourse analysis and reading instruction**, In T. Miller (Ed.), *Functional approaches to written texts: Classroom applications*, (2-15). Washington, DC: USA.
- _____ (2002), *Using Discourse Patterns to Improve Reading Comprehension* **JALT 2002 AT SHIZUOKA Conference Proceedings**.
- Graham, S., et al. (2005), “Improving the Writing Performance, Knowledge and Motivation of Struggling Young Writers: The Effect of Self-Regulated Strategy Development”, **Contemporary Educational Psychology**, 30, 207-241
- Granger, S., & Tyson, S. (1996), “Connector usage in the English essay writing of native and non-native EFL speakers of English”, **World Englishes**, 15(1), 17-27.

- _____ (1998), 'The computer learner corpus: a versatile new source of data for SLA research', pp. 3-18. In S. Granger (Ed.) **Learner English on Computer**, London: Longman.
- _____ (2002), 'A bird's-eye view of learner corpus research' in S. Granger, J. Hung and S. Petch-Tyson (Eds.) **Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching**, pp. 3–33. Amsterdam: John Benjamins.
- _____ (2003), 'Practical applications of learner corpora' in B. Lewandowska Tomaszczyk (ed.) **Practical Applications in Language and Computers**, pp. 291-302. Frankfurt: Peter Lang.
- Greene, S. (1991), "Mining Text in Reading to Write", **Journal of Advanced Composition** <http://www.jaconlinejournal.com/archives/vol12.1/greene-mining.pdf> (19.02.2013)
- Hyland, K. (2002), "Genre: Language, context and literacy", **Annual Review of Applied Linguistics**, 22, 113–135.
- _____ (2003), **Second Language Writing, Fourth Edition**, Cambridge University Press
- _____ (2009), **Teaching and Researching Writing**, Second Ed, Applied Linguistic in Action Series, Longman Pearson Publication.
- Ito, F. (2011), "L2 Reading-writing correlation in Japanese EFL High School Students" **Gunma National College of Technology**, *The Language Teacher* 35 (5).
- Joybrato M. and Jan-Marc R. (2006), "Rethinking Applied Corpus Linguistics from a Language- pedagogical Perspective: New Departures in Learner Corpus Research", **Applied Corpus Linguistics and Learner Corpus Research**.
- Johansson, V (2008), **Lexical diversity and lexical density in speech and writing: a developmental perspective**, Lund University, Department of Linguistics and Phonetics Working Papers 53 (61).
- Kamrul, H. and Akhand, M. (2010), "Approaches to Writing in EFL/ESL Context: Balancing Product and Process in Writing Class at Tertiary Level", **Journal of Nelta** 15 (1).

- Kennedy, C. and Bolitho, R. (1985), **English for Specific Purposes**, Hong Kong: Macmillan.
- Krashen, S. (1985), **The Input Hypothesis: Issues and Implications**, Harlow: Longman.
- _____ (1989), **Language Acquisition and Language Education**, Prentice Hall International.
- _____ (1993a), **The power of reading**, Englewood, CO: Libraries Unlimited.
- _____ (1993b), **We Learn to Write by Reading, But Writing can Make You Smarter**, University of Southern California.
- _____ (2004), **The Power of Reading, Insights from the Research**, Second edition, Libraries Unlimited Publication.
- Krashen, S. and Lee, S. (2004), "Competence in Foreign Language Writing: Progress and Lacunae, 12 (2).
<http://jaltcue-sig.org/files/OnCUE/12.2content.pdf> (15.12.2012)
- Kroll, B. (1990), **Second language writing: research insights for the classroom**, Cambridge: Cambridge University Press. In Kamrul, H. and Akhand, M. (2010), **Approaches to Writing in EFL/ESL Context, Balancing Product and Process in Writing Class at Tertiary Level**.
- Laufer, B. (1990), "Ease and difficulty in vocabulary learning: some teaching implications", **Foreign Language Annals**, 23, 147-156.
- _____ (1991), "The Development of L2 Lexis in the expression of advanced learner", **Modern Language Journal**, 75 (4), 440-448.
- Lee, S. H. & Muncie, J. (2006), "From receptive to productive: improving ESL learners' use of vocabulary in a post-reading composition task", **TESOL Quarterly**, 40(2), 295-320.
- Leech, G.(1997), "Teaching and language corpora: a convergence", In A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles (Eds.), **Teaching and Language Corpora**, Harlow: Addison Wesley Longman.
- Lewis, M. (1993), "The lexical approach: The state of ELT and the way forward." Hove, England: **Language Teaching Publications**.

- Lindquist, H. (2009), **Corpus Linguistics and the Description of English**, Edinburg: Edinburg University Press.
- McCarthy, P.M., and Jarvis, S. (2010), “MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment”, **Behavior Research Methods**, 42, 381_392. doi: 10.3758/BRM.42.2.381.
- Myles, F. (2005), “Interlanguage corpora and second language acquisition research”, **Second Language Research**, 21(4), 373-391.
- Nassaji, H. (2006), “The relationship between depth of vocabulary knowledge and L2 learners’ lexical inferencing strategy use and success”, **The Modern Language Journal**, 90(3), 387-401.
- Nation, I. S. P. (1990), **Teaching and Learning Vocabulary**. New York: Newbury House.
- Nelson, N., and Calfee, R. C. (Eds.). (1998), **The reading-writing connection. 97th yearbook of the National Society for the Study of Education**, Chicago: National Society for the Study of Education.
- Nagy, W. E., Anderson, et al. (1987),” Learning word meanings from context during normal reading”, **American Educational Research Journal**, 24, 237-270 *Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- O’Keefe A., et al. (2007), **From Corpus to Classroom: Language Use and Language Teaching**, Cambridge: Cambridge University Press
- Oshima, A. and Houge, A (2006), **Writing Academic English**, Fourth edition, Longman Pearson.
- Özçelik, M (1996), A Study of teaching writing through reading to low level pre-school students, published Master Thesis. Retrieved from **Turkish Higher Education National Theses Database**.
- Parodi, G. (2006), “Reading–writing connections: Discourse-oriented research”, **Reading and Writing Interdisciplinary Journal**, July 1-26.
- Paton, M. Q. (2002), **Qualitative research & evaluation methods**, Thousand Oaks: Sage Publication.

- Peat, B. (2006), "Integrating writing and d research skills: Development and testing of a rubric to measure student outcomes", **Journal of Public Affairs Education**, 12, 295–311.
- Perin, D (2006), *Assessing the Reading- Writing Relation in Adult Literacy Students*, **Reading Psychology**: Teachers College, Columbia University.
- Ponniah, J, R. (2011), "Incidental Acquisition of Vocabulary by Reading", **The Reading Matrix**, 11 (2).
- Priz, A.M.P (2008), Reasoning figuratively in early EFL: Some implications for the development of vocabulary, in Boers, F and Lindstromberg., S. (Eds). **Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology**, Mouton de Gruyter, Berlin, Newyork.
- Pulido, D. (2004), "The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity?", **Language Learning**, 54, 469-523.
- Qian, D. D. (2002), "Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective", **Language Learning**, 52(3), 513-536.
- Raimes, A (1983), **Techniques in Teaching Writing**. England: Oxford University Press.
- Rayson P and Garside, R. (2000), "Comparing corpora using frequency profiling. In Proceedings of the workshop on Comparing corpora", 9 (1–6), **Association for Computational Linguistics**.
- Sahillioğlu, F. et al., (2012), "Effectiveness of Two Corpus tools on Turkish Students' Learning English Grammar", **E-Journal of New World Sciences Academy (NWSA)**, 7(1), 230-240.
- Santoso, A. (2010), *Scaffolding an EFL effective writing class in a hybrid learning community*. Unpublished Ph.D. dissertation. Queensland: **Queensland University of Technology**. In Seifoori, Zohreh (2012) *A Profile of an Effective EFL Writing Teacher (A Technology-based approach)*.
- Schellekens, P. (2007), **The Oxford ESOL Handbook**. NY: Oxford
- Seidlhofer, B. (2002), *Pedagogy and local learner corpora: working with learning-driven*

- data. In: S. Granger/J. Hung/S. Petch-Tyson (eds.).**Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching**. Amsterdam: John Benjamins. 213-234.
- Seifoori, Z. (2012), A Profile of an Effective EFL Writing Teacher (A Technology-based Approach) Ph.D. in TEFL, Department of English Language Tabriz Branch, **Islamic Azad University**, Tabriz, 51575-5311, Iran
- Silva, T (1990), **Second Language Compositon Instruction: Developments, Issues and Directions in ESL In B. Kroll (Ed.)**, Second language writing: Insights for the classroom (pp. 11–23).Cambridge, UK: Cambridge University Press
- Shanahan,T.(1984), “Nature of the reading-writing relation: An Exploratory multivariate analysis”, **Journal of Educational Psychology**, 76,466-477.
- Shin C, H. (2010), “EFL Learners' Consciousness-Raising through a Corpus Based Approach”, **English Teaching**, 65(1), 57-86.
- Smith, F. (1988), **Joining the Literacy Club**, Portsmouth, NH: Heinemann.
- Steele, V.(1992), **Product and process writing: a comparison**, Rowley: Newbury House.
- _____ (2004), Product and process writing. Retrieved on 5th Sept. 2010 from <http://www.englishonline.org.cn/en/teachers/workshops/teaching-writing/teaching-tips/product-process>
- Tierney, R., and Shanahan, T. (1991), Research on the reading-writing relationship: Interactions, transactions, and outcomes, In R. Barr, M. Kamil, P. Mosenthal, and D. Pearson (Eds.), **The handbook of reading research**, 2, 246–280, New York, NY: Longman.
- Tsai, Jui-min (2006), “Connecting Reading and Writing in College EFL Courses Ohio State University”, **The Internet TESL Journal**, XII (12).
- Wai-King Tsang (1996), “Comparing the effects of reading and writing on writing performance”, **Applied Linguistics**, 17(2), 210-233.
- Waring,R. and Nation, P. (2004), “Second Language Reading and Incidental Vocabulary Learning”, **Angels on the English-Speaking World**, 4.
- Weigle, S. C. (2002), **Assesing Writing**, NY:Cambridge University Press.

Wilkins, D.A. (1992), **Second- Language Learning and Teaching**, Great Britain: The Chaucer Press.

Wolfe-Quintero, I., and Kim, H.Y. (1998), **Second language development in writing: Measures of fluency, accuracy, and complexity**, Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Zhang, Y. (2013), “A corpus based analysis of lexical richness of Beijing Mandarin speakers: variable identification and model construction”, Elsevier, **Language Sciences**.

APPENDICES

APPENDIX-1: SAMPLE LESSON PLAN FOR THE EXPERIMENTAL GROUP

Allocated Time: 1,5 hour

Place: Class 201

Objectives:

- ✓ Consider the meaning of readings and also work with the vocabulary needed for writing a descriptive essay.
- ✓ Establish and understand the importance of vivid descriptions
- ✓ Introduce & review descriptive writing
- ✓ Take notes on descriptive writing
- ✓ Make the connection between figurative language and strong descriptions
- ✓ Write an original descriptive essay depending on their analysis of a model descriptive essay.

Reading for Writing:

1. Analyzing Model Essays

Reading and analyzing a model essay for description “Cherries for My Grandma” by Geoffrey Canada, and My Grandfather by Gary Soto (see p. 23 in Reason to Write)

1.1 General Understanding the Reading

Do the exercises in p.24, 25 in Reason to Write.

1.2 Working with the Language

Identify Synonyms in p. 26 in Reason to Write.

1.3 Describing Characters

Read the list of adjectives and write each adjective next to the person. p.27 in Reason to Write

Pre-Writing Activity

Read the excerpt “My Grandfather” by Gary Soto and try to see how the writer described his grandfather.

1. Recognizing Detailed Descriptions

Read each adjective and decide whether it could be used to describe Soto’s grandfather. P.28 in Reason To Write.

2. Writing Detailed Descriptions

In your notebook, write at least three characteristics of a person who influenced you . See p.29 in Reason to Write

Structured Writing Focus

Write a five paragraph descriptive essay about a childhood relationship with a person who has had a powerful influence over you. Give specific examples and detailed descriptions of things you did together and conversations you had. Show how this relationship helped you to become the person you are.

Preparing the First Draft

For Analyzing Essay Structure and identifying parts of the essay see pages 31 and 32 in Reason to Write.

Analyze the student essay second draft and answer the questions see pages 33 and 34.

Organizing Your Essay

Study the block diagram of a five paragraph essay to plan a first draft of your essay. In your notebook, draw your own diagram and write your notes in each of its sections. See page 38.

APPENDIX-2: SAMPLE LESSON PLAN FOR THE CONTROL GROUP

Allocated Time: 1,5 hour

Objectives:

- ✓ Establish and understand the importance of vivid descriptions
- ✓ Take notes on descriptive writing
- ✓ Make the connection between figurative language and strong descriptions
- ✓ Write an original descriptive essay depending on their analysis of a model descriptive essay.

Structuring a Descriptive Essay

A descriptive essay simply describes something or someone by appealing to the reader's senses: sight, sound, touch, smell and taste. Here are the basic steps to writing an effective descriptive essay:

1. Select a subject

Observation is the key to writing a good description. For example, if you are writing about a place, go there and take notes on the sights, sounds, and smells. A descriptive essay paints a picture for the reader, using descriptive devices and the senses. Create a thesis statement that informs the reader who or what you are describing. Examples: "The wooden roller coaster in Coney Island is a work of art." "My bedroom is an ocean sanctuary."

2. Select dominant details

Select only the details that support the dominant impression (your thesis statement).

3. Organize details

The paragraphs in a descriptive essay can be structured spatially (from top to bottom or from near to far) or chronologically (time order) or from general to specific. Descriptive essays can also use other patterns of organization such as narrative or exemplification.

4. Use descriptive words

Do not use vague words or generalities (such as good, nice, bad, or beautiful). Be specific and use sensory, descriptive words (adjectives). For example: I ate a good dinner. OR I devoured a steaming hot, cheese-filled pepperoni pizza for dinner. Provide sensory details: Smells that are in the air (the aroma of freshly brewed coffee) Sounds (traffic, honking horns) Sights (“The sun scattered tiny diamonds across dew-covered grass as it peeked out from beyond the horizon.”) Touch (“The texture of the adobe hut’s walls resembled coarse sandpaper.”) Taste: sweet, sour, salty, bitter, tart (“Giant goose bumps formed on my tongue when I accidentally bit into a sliver of lemon.”)

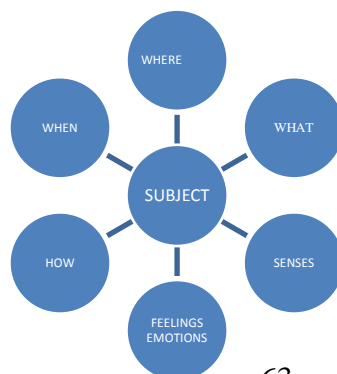
5. Draw a logical conclusion

The conclusion may also use descriptive words; however, make certain the conclusion is logical and relevant.

Pre Writing Activity:

First students choose a topic to describe. Students spend time thinking about the qualities they want to describe about a person, a place or an event. Students brainstorm about all the details associated with the topic. Students take notes on where is the object located? Where does the person live? Students consider memories, feelings, and ideas that subject evokes, aside from physical characteristics of their subject. Memory and emotion play an important role in conveying the subject’s significance. After brain storming, students put these details into logical sequence and create paragraphs for a descriptive essay.

Filling in Graphic Organizers to Draft a Descriptive Essay



Writing Task: Write a five paragraph descriptive essay about a childhood relationship with a person who has had a powerful influence over you. Give specific examples and detailed descriptions of things you did together and conversations you had. Show how this relationship helped you to become the person you are.

**APPENDIX-3: ESSAY SCORING RUBRIC FOR WRITING DEVELOPED
BY OSHIMA, HOUGE AND BUTTLER (2006)**

Scoring Rubric: Essays		
	Maximum Score	Actual Score
Format—5 points Title centered (2), first line of each paragraph indented (1), margins on both sides (1), text double-spaced (1)		
Total	5	
Mechanics—5 points Punctuation: periods, commas, semicolons, quotation marks (3), capitalization (1), spelling (1)		
Total	5	
Content—20 points The essay fulfills the requirements of the assignment. The essay is interesting to read. The essay shows that the writer used care and thought.	5 5 10	— — —
Total	20	
Organization—45 points The essay follows the outline, and it has an introduction, a body, and a conclusion. Introduction: The introduction ends with the thesis statement. Body Each paragraph of the body discusses a new point and begins with a clear topic sentence. Each paragraph has specific supporting material: facts, examples, quotations, paraphrased or summarized information, etc. Each paragraph has unity. Each paragraph has coherence. Transitions are used to link paragraphs. Conclusion: The conclusion summarizes the main points or paraphrases the thesis statement, begins with a conclusion signal, and leaves the reader with the writer's final thoughts on the topic.	5 5 5 10 5 5 5 5 5	— — — — — — — — —
Total	35	
Grammar and Sentence Structure—25 points Estimate a grammar and sentence structure score.	25	
Grand Total	100	

© 2006 Pearson Education, Inc. Duplication for classroom use is permitted.

APPENDIX-4: TAG DESCRIPTIONS BY DOUG BIBER, 15 JUNE, 1993

There are five tag fields, separated by a plus sign (+). For most words, only one or two of the fields are used. The primary grammatical category of a word is usually marked in the first tag field; many of these first-field tags are identical to tags used in the LOB tag set. In the case of adjectives (TAG = JJ), nouns (TAG = NN), and verbs (TAG = VB), the tag ?? can appear in Tag Field 4 to mark words that were not in the dictionary; in these cases, the grammatical category is assigned based on morphology and the surrounding context. The tags xvbn and xvbnx in Field 4 mark a word as being a past participle form, regardless of function. Thus, some adjectives, nouns, and base verb forms are marked as xvbn. All past tense verbs, perfect aspect verbs, and passive verbs have this tag. The tag xvbnx is used to mark cases where the grammatical function (e.g., perfect or passive) has been identified with a very high degree of accuracy from the context; the tag xvbn is used for cases where the assigned grammatical function is less certain. The tags xvbg and xvbgx in Field 4 mark a word as being a present participle form, regardless of function. Thus, some adjectives and nouns are marked as xvbg. All present progressive verbs have this tag. The tag xvbgx is used to mark cases where the grammatical function has been identified with a very high degree of accuracy; the tag xvbg is used for cases where the assigned grammatical function is less certain.

Tag Sequence
:+clp+++ colon + clause punctuation
;+clp+++ semi-colon + clause punctuation
?+clp+++ question mark + clause punctuation
!+clp+++ exclamation mark + clause punctuation
,++++ comma
-++++ dash
"++++ double quote mark
'++++ single quote mark
(++++ left parenthesis
)++++ right parenthesis
\$++++ dollar sign

%++++	percent sign
&fo++++	formula symbols
&fw++++	foreign word
abl++++	pre-qualifier (rather, such)
abn++++	pre-quantifier (all, half)
abx++++	pre-quantifier/double conjunction (both)
ap++++	post-determiner (many, more, most, only, other, own, same, ...)
aps++++	(others)
at++++	singular indefinite article (a, an)
ati++++	singular definite article (the, no)
cc++++	coordinating conjunction (and, but, or)
cc+cls+++	coordinating conjunction + clausal connector
cc+phrs+++	coordinating conjunction + phrasal connector
cc"++++	multi-word coordinating conjunction (as well as)
cc++neg++	coordinating conjunction + + negation (nor)
cd++++	cardinal number (2, 3, 4, two, three, four, hundred, ...)
cd+date+++	cardinal number + date (year only)
cd1++++	cardinal number: 1, one
cd1s++++	cardinal number: ones
cds++++	cardinal plural (tens, hundreds, thousands)
od++++	ordinal number (1st, 2nd, first, second, ...)
cs+cnd+++	subordinating conjunction + conditional (if, unless)
cs+con+++	subordinating conjunction + concessive (although, though)
cs+cos+++	subordinating conjunction + causative (because)
cs+who+++	subordinating conjunction + WH word (whether)
cs+sub+++	subordinating conjunction + other (as, except, until, ...)

cs"++++	multi-word subordinating conjunction (in that, so that, ...)
dt+dem+++	determiner + demonstrative (this,that,these,those modifying N)
dt+pdem+++	determiner + demonstrative pronoun (this, that, these, those)
dti++++	singular or plural determiner (any, enough, some)
dt++++	other singular determiner (another, each)
dtx++++	determiner/double conjunction (either)
ex+pex+++	existential there
in++++	preposition
in+ppvb+++	preposition + prepositional verb (account for, join in, ...)
in+pl+++	preposition + place marker (above, behind, beside, ...)
in"++++	multi-word perposition (as to, away from, instead of, ...)
in+strn+++	preposition + stranded
jj+atrb+++	adjective + attributive function
jj+atrb++xvbg+	adjective + attributive function + + -ing form
jj+atrb++xvbn+	adjective + attributive function + + past participle form
jj+pred+++	adjective + predicative function
jj++++	adjective + indeterminate function
jjb+atrb+++	attributive-only adjective + attributive (chief, entire)
jjr+atrb+++	comparative adjective + attributive function
jjr+pred+++	comparative adjective + predicative function
jjt+atrb+++	superlative adjective + attributive function

All modal forms can be marked as 0 in Field 5 (e.g., md+prd+++0) to show that they are contracted forms (e.g., 'll, 've)

md+nec+++	modal + necessity (ought, should, must)
md+pos+++	modal + possibility (can, may, might, could)
md+prd+++	modal + prediction (will, would, shall)

md"++pmd"++	modal + + multi-word periphrastic modal (e.g., be going to)
nn++++	singular common noun
nn+nom+++	singular noun + nominalization
nvbg+++xvbg+	singular noun + + + -ing form
nn+++xvbn+	singular noun + + + past participle form
nns++++	plural common noun
nns+nom+++	plural noun + nominalization
nnu++++	unit of measurement (lb, kg, ...)
np++++	singular proper noun
nps++++	plural proper noun
npl++++	locative noun
npt++++	singular titular noun
npts++++	plural titular noun
nr++++	singular adverbial noun (east, west, today, home, ...)
nrs++++	plural adverbial noun

NB: In the following pronoun tags, be careful of the difference between the number 1, used to mark first person, and the letter l (i.e. lower case L), used to mark reflexives.

pp1a+pp1+++	first person subject pronoun + first person pronoun
pp1a+pp1+++0	first person subject pronoun + 1st person pro. + contracted
pp1o+pp1+++	first person object pronoun + first person pronoun
pp\$+pp1+++	possessive determiner + first person pronoun (my, our)
ppl+pp1+++	singular reflexive pronoun + first person pronoun (myself)
ppls+pp1+++	plural reflexive pronoun + first person pronoun (ourselves)
pp2+pp2+++	second person pronoun + second person pronoun
pp\$+pp2+++	possessive determiner + second person pronoun (your)
ppl+pp2+++	singular reflexive pronoun + second person pronoun (yourself)

pp3a+pp3+++	third person subject pronoun + third person personal pronoun
pp3o+pp3+++	third person object pronoun + third person personal pronoun
pp3+pp3+++0	third person pronoun + 3rd person personal pro. + contracted
pp\$+pp3+++	possessive + 3rd pers. personal pro. (his, her, their)
ppl+pp3+++	sg. reflexive pronoun + 3rd pers. personal pro. (her/himself)
ppls+pp3+++	pl. reflexive pronoun + 3rd pers. personal pro. (themselves)
pp3+it+++	third person pronoun + third person impersonal pronoun (it)
pp\$+it+++	possessive determiner + third person impersonal pronoun (its)
pp\$\$++++	possessive pronoun (mine, yours, ...)
pn"++++	multi-word nominal pronoun (no one, ...)
pn++++	nominal pronoun (someone, everything, ...)

ql++++	qualifier + (as, less, more, too)
ql+amp+++	qualifier + amplifier (very)
ql+emph+++	qualifier + emphatic (most)
qlp++++	post-qualifier (enough, indeed)

All adverb forms can be marked as spl in Field 3 (e.g., rb+amp+splt++) to indicate that the adverb occurs within the auxiliary (e.g., they've probably been looking...).

rb++++	general adverb
rb"++++	multi-word adverb (at last, in general)
rb+cnj+++	adverb + conjunct (however, therefore, thus, ...)
rb++neg++	neither
rb+amp+++	adverb + amplifier (absolutely, completely, entirely, ...)
rb+down+++	adverb + downtoner (nearly, only, merely, ...)
rb+emph+++	adverb + emphatic (just, really, so, ...)
rb+hdg+++	adverb + hedge (almost, maybe, ...)

rb"+hdg"+++	multi-word adverb + hedge (kind of, sort of)
rb+phrv+++	adverb + phrasal verb (get in, wrap up, ...)
rb+pl+++	adverb + place marker (abroad, ahead, far, upstream, ...)
rb+tm+++	adverb + time marker (afterwards, again, immediately, ...)
rb+dspt+++	adverb + discourse particle (anyway, well, ...)
rbr++++	comparative adverb (better, quicker)
rbr+tm+++	comparative adverb + time marker (earlier, later, sooner, ...)
rn+pl+++	nominal adverb + place marker (here, there)
rn+tm+++	nominal adverb + time marker (now, then)
rn+dspt+++	nominal adverb + discourse particle (now)
rp++++	adverbial particle (back, in, round, up, ...)
rp+pl+++	adverbial particle + place marker (away, behind, out, ...)

tht+jcmp+++	that as dependent clause head + adjective complement
tht+ncmp+++	that as dependent clause head + noun complement
tht+vcmp+++	that as dependent clause head + verb complement
tht+rel+++	that as dependent clause head + relative clause
tht+rel+obj++	that as dep. clause head + relative clause + object position
tht+rel+subj++	that as dep. clause head + relative clause + subject position

to++++	infinitive marker
to"++++	multi-word infinitive marker (in order to)
uh++++	interjection/filler (hey, oh, ok, yes, erm ...)
vb++++	base form of verb, excluding verbs in infinitive clauses

(uninflected present tense, imperative)
vb+++xvbn+ base form of verb + + + past participle form
(e.g., cut, hit, hurt, ...)
vb+be+aux++ base form of verb + be + auxiliary verb
vb+be+vrb++ base form of verb + be + main verb
vb+bem+aux++ verb + am + auxiliary verb
vb+bem+aux++0 verb + am + auxiliary verb + + contracted ('m)
vb+bem+vrb++ verb + am + main verb
vb+bem+vrb++0 verb + am + main verb + + contracted ('m)
vb+ber+aux++ verb + are + auxiliary verb
vb+ber+aux++0 verb + are + auxiliary verb + + contracted ('re)
vb+ber+vrb++ verb + are + main verb
vb+ber+vrb++0 verb + are + main verb + + contracted ('re)
vb+do+aux++ verb + do + auxiliary verb
vb+do+vrb++ verb + do + main verb
vb+hv+aux++ verb + have + auxiliary verb
vb+hv+aux++0 verb + have + auxiliary verb + + contracted ('ve)
vb+hv+vrb++ verb + have + main verb
vb+hv+vrb++0 verb + have + main verb + + contracted ('ve)
vb+seem+++ base form of verb + seem / appear
vb+vprv+++ base form of verb + private verb (believe, feel, think, ...)
vb+vprv+tht0++ base form of verb + private verb + that deletion **
vb+vpub+++ base form of verb + public verb (assert, complain, say, ...)
vb+vpub+tht0++ base form of verb + public verb + that deletion **
vb+vsua+++ base form of verb + suasive verb (ask, command, insist, ...)

All past tense verb forms (excluding were, was, did, had) are marked as either xvbn or xvbnx; only the sequences with xvbn are listed below.

vbd+++xvbn+	past tense verb + + + past participle form
vbd+bed+aux++	past tense verb + were + auxiliary verb
vbd+bed+vr++	past tense verb + were + main verb
vbd+bedz+aux++	past tense verb + was + auxiliary verb
vbd+bedz+vr++	past tense verb + was + main verb
vbd+dod+aux++	past tense verb + did + auxiliary verb
vbd+dod+vr++	past tense verb + did + main verb
vbd+hvd+aux++	past tense verb + had + auxiliary verb
vbd+hvd+vr++	past tense verb + had + main verb
vbd+seem+++xvbn+	past tense verb + seem/appear
vbd+vprv+++xvbn+	past tense + private verb (believe, feel, think, ...)
vbd+vprv+tht0+xvbn+	past tense + private verb + that deletion **
vbd+vpub+++xvbn+	past tense + public verb (assert, complain, say, ...)
vbd+vpub+tht0+xvbn+	past tense + public verb + that deletion **
vbd+vsua+++xvbn+	past tense + suasive verb (ask, command, insist, ...)

** the tag tht0 marks the occurrence of a following that complement clause	
when the complementizer that has been deleted	
All present progressive verb forms are marked as either xvbg or xvbgx;	
only the sequences with xvbg are listed below.	
vbg+++xvbg+	present progressive verb + + + -ing form
vbg+beg+++xvbg+	present progressive verb + being
vbg+beg+aux+xvbg+	present progressive verb + being + auxiliary verb
vbg+hvg+++xvbg+	present progressive verb + having
vbg+vprv+++xvbg+	pres. prog. + private verb (believe, feel, think, ...)
vbg+vprv+tht0+xvbg+	present progressive + private verb + that deletion **

vbg+vpub++xvbg+	pres. prog. + public verb (assert, complain, say, ...)
vbg+vpub+tht0+xvbg+	present progressive + public verb + that deletion **
vbg+vsua++xvbg+	pres. prog. + suasive verb (ask, command, insist, ...)
vwbg+++xvbg+	present progressive postnominal modifier
vwbg+beg++xvbg+	present progressive postnominal modifier + being
vwbg+hvg++xvbg+	present progressive postnominal modifier + having
vwbg+vprv++xvbg+	present prog. postnom. modifier + private verb
vwbg+vpub++xvbg+	present prog. postnom. modifier + public verb

vbi++++	base form of verb in infinitive clause
vbi+vprv+++	infinitive verb + private verb (believe, feel, think, ...)
vbi+vprv+tht0++	infinitive verb + private verb + that deletion **
vbi+vpub+++	infinitive verb + public verb (assert, complain, say, ...)
vbi+vpub+tht0++	infinitive verb + public verb + that deletion **
vbi+vsua+++	infinitive verb + suasive verb (ask, command, insist, ...)

vbz++++	3rd person singular verb
vbz+bez+aux++	3rd person sg. verb + is + auxiliary verb
vbz+bez+aux++0	3rd person sg. + is + auxiliary verb + + contracted ('s)
vbz+bez+vrb++	3rd person sg. verb + is + main verb
vbz+bez+vrb++0	3rd person sg. + is + main verb + + contracted ('s)
vbz+doz+aux++	3rd person sg. verb + does + auxiliary verb
vbz+doz+vrb++	3rd person sg. verb + does + main verb
vbz+hvz+aux++	3rd person sg. verb + has + auxiliary verb
vbz+hvz+vrb++	3rd person sg. verb + has + main verb

vbz+seem+++	3rd person sg. verb + seem/appear
vbz+vprv+++	3rd person sg. + private verb (believe, feel, think, ...)
vbz+vprv+tht0++	3rd person sg. + private verb + that deletion **
vbz+vpub+++	3rd person sg. + public verb (assert, complain, say, ...)
vbz+vpub+tht0++	3rd person sg. + public verb + that deletion **
vbz+vsua+++	3rd person sg. + suasive verb (ask, command, insist, ...)

** the tag tht0 marks the occurrence of a following that complement clause when the complementizer that has been deleted

All perfect aspect verb forms and passive verb forms are marked as either xvbn or xvbnx; only the sequences with xvbn are listed below.

vprf+++xvbn+	perfect aspect verb + + + past participle form
vprf++tht0+xvbn+	perfect aspect verb + + that deletion **
vprf+ben+aux+xvbn+	perfect aspect verb + been + auxiliary verb
vprf+ben+vrb+xvbn+	perfect aspect verb + been + main verb
vpsv++agls+xvbn+	main clause passive verb + + agentless passive
vpsv++by+xvbn+	main clause passive verb + + by passive
vwbn+++xvbn+	passive postnominal modifier + + + past participle form
vwbn+vprv++xvbn+	passive postnominal modifier + private verb
vwbn+vpub++xvbn+	passive postnominal modifier + public verb
vwbn+vsua++xvbn+	passive postnominal modifier + suasive verb

wdt+who+++	WH determiner + WH word (what, whatever, whichever, ...)
wdt+who+whcl++	WH determiner + WH word + WH clause
wdt+who+whq++	WH determiner + WH word + WH question

whp+rel+obj++	WH pronoun + relative clause + object position
whp+rel+pied++	WH pronoun + relative clause + object position with prepositional fronting ('pied piping')
whp+rel+subj++	WH pronoun + relative clause + subject position
whp+who+++	WH pronoun + WH word (not a relative clause)
whp+who+whq++	WH pronoun + WH word + WH question
wrb+who+++	WH adverb (how, when, where, ...) + WH word
wrb+who+whcl++	WH adverb + WH word + WH clause
wrb+who+whq++	WH adverb + WH word + WH question
xnot++not++	not + + negation
xnot++not++0	not + + negation + + contracted form (n't)
xvbn+++xvbn+	past participle form -- indeterminate grammatical function
xvbg+++xvbg+	present participle form -- indeterminate grammatical function
zz++++	letter of the alphabet

APPENDIX-5: STUDENT DESCRIPTIVE ESSAY 1:THE EXPERIMENTAL GROUP

MY MOTHER

Do you know a person who will always with you whatever you do? Mother.. This word brings love, compassion, devotion to our minds. She is the only person who will understand and forgive everything. We are born little and helpless to this world and there is a voice relaxed us in our side. We listen her soft voice while we are sleeping and know that she is always with us. A mother kiss heals everything and in mother's look, you never see falseness. In the most complicated situations, you can ask for advice from your mother and I am sure that they have absolutely one. They are so special persons and I can easily say that when you go away from your mother, you understand her special qualities.

First of all, my mother is a good cook. If you are a student in another country and don't know about cooking, you crave for your mother. I have been studying at Karadeniz Technical University and before I came here, she had warned me about learning how to cook. I didn't know something about meal and only ate my mother meals because her meals were very special for me. When she entered to kitchen, her magic hands were doing wonders. She was choosing carefully her materials and I always believed that she added her love into her meals. Now, I can easily say that I really miss to eat her meal but I know that I never taste like her meals wherever I go.

An excellent quality of my mother is that she is relaxing. When I came to Trabzon, everything was awful for me. I had no friend and preferred to go out alone. Sometimes, I needed a friend who will talk to me and share something. When I was at Konya, there was a voice that can sooth me but it is not here now. If there is a problem in my life and I can't solve it, the only thing that I need to hear my mother voice. You can ask yourself can a voice solve all problems? If it is my mother's voice, yes it can. Although there are lots of people who can find a solution about my problems, their solutions can't always solve my problems but if she is my mother, it is impossible. Actually, I don't know how she can do this but she always succeeds it.

Beside my mother is relaxing, she is an advisor as well. Is something wrong in my life? I am sure that my mother knows it and ready to help me. When I was at high school, I wasn't a good student because I didn't like to work. It was not important for me. I

preferred to meeting my friends and surfing on the internet. My exam results are enough to pass the class but they are not enough for my mother. Firstly, she warned me fairly but I ignored her warning. After a week, she prohibited something like meeting friends. That was awful but I understand that if my mother punished me, maybe I couldn't pass university exam. Because of her punishing, I regularly studied every day.

Consequently, my mother is the most sincere and true assistant and the most valuable person in my life. Feeling her love is different from all other things in the world.

APPENDIX-6:STUDENT DESCRIPTIVE ESSAY 2: THE EXPERIMENTAL GROUP

MY MOM

The most significant person in my life is my mother who carried me nine months. She is keen on me in every way. She loves me no matter what happens. I love her three qualities: she is a good adviser, great cook and she is so beautiful.

Whenever I need to get advice I call my mom. My mother has always been ready to listen to me and she knows how to make me happy thanks to her advice. It doesn't matter whether the topic includes secret things or not, I can share my feelings and emotions with her. And she always gives importance to my happiness therefore she gives advice without judging me. A year ago I broke up with my girlfriend. I felt my heart was broken into pieces. I did not want to talk to talk or do anything except sleeping. My mother came into the room. She asked me what was happen. I told her this is personal thing and she said I'm your mother there are no personal stuff between us. I told her what happened. She told me her precious advice "if you don't love yourself, no one will love you. You should do it first then expect from others." That advice made me feel great and it helped me without my comprehension. This is why she is a good adviser.

My mother doesn't like to make foods which require lots of time and effort, but the meals that she makes are delicious. I live distant from the ottoman cuisine, but the meals that can be prepared easily. For example the food that she makes by using only bread, cheese and eggs. I want to go my hometown in order to eat that food, and it proves that my mother is a good cook.

Beauty is such an issue that each every person may give different meaning to it. And the question of who is the beautiful woman depends on the people. But for me, the most beautiful woman is my mother. Although my father was a kind of womanizer when he saw my mother's beauty, he got rid of the days in which he spent most of time with women, despite of his young age.

To sum up, there are invisible chains between us. My bond with my mother will never break into pieces. She is there when I need some advice she makes delicious meals

when I'm hungry and she is so beautiful both side. She is beyond my expectations and I thank god for gave me her.

APPENDIX-7: STUDENT ADVANTAGE DISADVANTAGE ESSAY 1: THE EXPERIMENTAL GROUP

THE BEST WOMAN IN MY LIFE

My English teacher is important for me. When I went to high school, she was my English teacher. And she influenced on my life. She is thirty years old and she has green eyes, yellow hair. So she is tall. Also, she has married and she has two daughter. Although she has two children, she has a fit body, too. I can say a lot of qualification about her. But some influenced on my life. She is hard-working, friendly and kind person.

She is hard-working person. She always showed us that she is hard-working person. For example, when I was high school, although I had a lot of teacher, I asked her questions which did not the answer. Because she had so qualification that she could answer nearly every question. I did not see her freely. Also, she often occupied with anything. So, she always improved herself by reading. Moreover, she participated a lot of conferences. She told us information which learned in there. She made me gain reading habit. I love her this qualification. Because, she showed me the most right way in my life.

Also, she is friendly person. I shared my everything about myself with teacher. Because, I knew she showed me right way. For example, when I discussed with my family, I spoken with her. She listened me carefully, and she gave me some advises. However, she never did not press on me for practicing them. So, I could speak with her about my boyfriend. And she told me I was so young such a thing and when I would go to the university, I could most people in there. Eventually, I saw her like my sister.

Moreover, she is so kind woman. She paid attention her speaking. She always wanted that we were to be kind person. And she learned us being kind human. Because, she believed that kind person is modern person. So, she always told me being kind is part of our culture. We are educated people and we have some differences from other people. When we spoken impolitely, she warned us. Thanks to her, I learned being kind person during my life. All in all, my English teacher the only person influences on my life. Thanks to her, I like English lesson, because she made me loved English. If she does not behavior such this way, maybe I could not select this department. Being English teacher would not be my only dream. I owe my teacher thanks for everything. I hope I will be teacher the same as her.

APPENDIX-8: STUDENT ADVANTAGE AND DISADVANTAGE ESSAY: THE EXPERIMENTAL GROUP

Monitoring in the Workplace

Today big companies monitor their employees' telephone calls, activities on the Web, e-mails etc. Companies do it to make provision for security, and to block unnecessary activities, and to provide savings. Some employees waste a lot time by watching videos, chatting, speaking on the phone for personal calls. To prevent these kinds of situations monitoring them can be seen acceptable. Monitoring employees in the work place has advantages but also disadvantages.

To start with advantages, monitoring in the workplace offers security. All day long, a lot of people come to the offices, shops, etc. To prevent thieving, save corporate information etc. companies use security cameras. But, during the record employers also can watch the employees' behaviors. What are they doing in office-hours: do they chat with the other employees; are they helpful to the customers or do they just sit in front of the computer and surf on the net all the time? By monitoring employees employers can find appropriate solutions to these problems. They can warn the employees to reform their behaviors; they can find the thief if there is a theft incidence. Secondly, monitoring employees electronically provides many things to the companies. For example, according to AMA's 2007 Electronic Monitoring & Surveillance Survey[1] about computer monitoring: "Employers who block access to the Web are concerned about employees visiting adult sites with sexual, romantic, or pornographic content (96%); game sites (61%); social networking sites (50%); entertainment sites (40%); shopping/auction sites (27%); and sports sites (21%). In addition, companies use URL blocks to stop employees from visiting external blogs (18%)." Thus, employers can prevent their employees from bad content, and they also prevent employees from wasting time in vain. Electronically monitoring also controls employees' phone calls and voice mails. According to AMA's survey "Six percent of employers have fired employees for misuse or private use of office phones." By monitoring the calls, employers generally try to save the money which they pay for phone bills. And monitoring the voice mails, employers warn the employees not to use the bandwidth ineffectively.

Employers can save money, time, and information by monitoring the employees but this time other problems arise. First, employees will be suspicious of their privacy. Every step they take, every web site they enter, every one they speak on the phone is watching by someone and this makes people nervous. They try to be more conscious about what they are doing, and day by day this might be frustrating. According to Electronic Monitoring: The Employees Point of view[2]: “From the employees’ point of view the most devastating effect of employees’ monitoring is the fear of losing their jobs.” And the study gives information about privacy problems: “Privacy invasion: employees of many organizations are exposed to many types of privacy invasive monitoring while earning a living. However, these activities may diminish employment moral and dignity and increase stress and may lead to lawsuit cases.” Employers should be careful about the boundaries about the employees’ privacy. Because of monitoring, employees can query the trust between employers and themselves. They may feel that they are not reliable, and this feeling decreases their will to work. The study tells: “Lowest interest in the job transforms the work place to become mistrust and hostile work environment. That makes the employees feel less satisfied with their job and start looking for other secure job opportunities.” Stress related problems are the other disadvantages of monitoring employees. The study shows: “Workers who are subjected to computer monitoring constantly feel pressure to perform and thus are more likely to succumb to stress related illness.” And this stress finally shows physical symptoms. In a study by the department of Industrial Engineering , University of Wisconsin- Madison, higher levels of stress in monitored employees resulted in an increase in somatic complaints, including a 27% increases in occurrences of pain or stiffness in shoulders, a 23% increases in occurrences of neck pressure and 21% increases in back pain experienced by employees. Such stress and stress related symptoms can create medical expenses, lost time and absenteeism. [3] Employers should be clear about the monitoring employees, and set the boundaries fairly.

In my opinion, monitoring employees are necessary for order but employers should be careful about the obdurate stance. If an employee breaks the rule, they shouldn’t fire him/her immediately. They should inform employees clearly before starting to work with their company. Employers should listen to the employees concerns about privacy and try to find solutions.

To sum up, today almost every company uses the technology for monitoring the people because of different reasons. So, employers should be clear about the monitoring employees, and set the boundaries fairly. Employers should be careful about the time if they are doing unnecessary things. Trust is the most important thing in this case. Employers and employees should be careful about it.

APPENDIX-9: STUDENT DESCRIPTIVE ESSAY 1: THE CONTROL GROUP

THE BEST MY FRIEND

For a human after her family what is the important think of her life? In my opinion, it is 'the friendship.' The friendship is a special concept. It satisfy and get out of loneliness. You can have many friends at life but there is a minimum one best friend. The best friends stand by you in all circumstances.

I met with the best my friend at preparatory school. She was a dark, thin an tall girl. At first view, I didn't like her. As I know, I come to recognize that one is wrong. That day was the beginning of our friendship.

She was a fantastic friend for me. She was a good, sympathizer, compatible and mellow girl and loved me. Of course, me too. We were shared our everything. Somebody says: a friend in need is a friend indeed or. She stood by me in my good times or in my bad times. I still do. We hold together at elementary secondary and senior high school. Yes! She and I still are friends. She worried with man that she loved after high school and very happy now.

Nowadays, she has a baby. It is a teeny, sweet boy. I'm so happy on his behalf because she deserve that. Consequently, the friendship is a something holy. It is important to set up beautiful friendships. He/She stand by you through which and thin and back up you. An advice from me, 'DO NOT FRIENDLESS.'

APPENDIX-10: STUDENT DESCRIPTIVE ESSAY 2: THE CONTROL GROUP

MY MOTHER

Life is just empty without someone who loves us. Everyone has someone who takes care of them in this world. So do I. She is my mother. To me, my mother means everything. She is my family, my support and my adviser. So I want to talk about my mother in this essay.

Her name is Nurcihan Kaya. She is 44 years old. I think that my mother is a beautiful person. She is not tall but not short, and she has straight hair and dark. She has green eyes and she has a beautiful smile. She is a very kind person. She is very lovely, friendly, patient, and she loves to help people. So she is a good example to me.

I am an only child. My father passed away when my mother was 26 years. And my mother did not marry after my father had passed away. She was very young. If she had wanted, she could have married. But she did not marry, she only thought me.. She dedicated her life to me. She is both a very good mother and a very good father to me.

Finally, I really want to say that I love my mother very much! In this life, I will make sure that I respect her and don't do anything to let her trust in me down.

**APPENDIX-11: STUDENT ADVANTAGE DISADVANTAGE ESSAY 1:
THE CONTROL GROUP**

PRIVACY AT WORK

Do you want to check your e-mails without any doubt? Do you want to enter your passwords without the fear of monitoring? Then you should insist your boss that you want privacy at work. Do you want to check your employees while they are working? Do you wonder what are they doing while you think they are doing their jobs? Then, you should have a right to monitor them and be the boss. It isn't matter whether you are the boss or the employee but there are several advantages and disadvantages of work privacy.

The first advantage is that employees feel comfortable while they are working. John Stephen Curtis who is the head of the Employees Rights Association said "Employees must have their privacy at work if you want from them to be productive. "In an interview, According to the United Kingdom's daily telegraph newspaper in 20th June. Secondly, you can hide your private passwords of your own from your employer. If you signed in some of the sites which are important for you, you don't want anyone except you look at them. When you in your lunch break, you can easily check you accounts and you can look up the things you want. Finally, there will be a trust relationship between you and your boss. Ayşe Özer, who works for Çebi A.Ş for six years, can easily look up her e-mails and accounts. Her employer loves her work but in the beginning he thinks that she won't work because of the privacy at work. Then, he realize it isn't a big thing and it isn't effect to her work. To sum up, privacy is a thing that employees need it while they are working.

On the other hand, there are several disadvantages. First of them is that, your employees will surf on internet. They can get in their Facebook accounts and forgot their work. They can chat with their friends and didn't realize the time passed. Secondly and maybe the most importantly, they can crash your network system by downloading files. 02 faced a huge compensation bill after a meltdown in its network cut off millions of customers. There was a massive bill for Tesco mobile, which uses 02 masts for its network, after their network crashed because of their employee who tried to download a funny video on internet. Finally, you employees can stole private money accounts from your customers and they can transfer it to their own accounts. I was work in a company last summer and one of my

friend stole his customer money and the authorities figure that out about 2 weeks after. He is in jail now and the company has security system for that kind of situations.

In my opinion, if you trust your employees, there can be privacy at work but you cannot trust anyone including your workers. Work privacy is something that I don't like. Your employees will pass their time by talking on the net with their friends, spend their time in their Facebook or twitter accounts. I think you have the right of monitoring them. To sum up, I'm disagree with this idea.

As a conclusion, there are advantages and disadvantages of work privacy. Maybe you think this is necessary if you want to work in your comfort zone or you can think this is really dangerous because of the circumstances which I explained in the third paragraph but it is your decision to do it or not.

APPENDIX-12: STUDENT ADVANTAGE DISADVANTAGE ESSAY 2: THE CONTROL GROUP

ADVANTAGES AND DISADVANTAGES OF WORKPLACE MONITORING

There are some advantages and disadvantages of workplace monitoring of employees by employers. The employers are duty of supervising employees. But, the monitoring of people could be achieved with education. People should be given ability of self-management and telling the truth with education. The employees should inform their employers when they saw unsafe behavior in workplaces.

There are some advantages of monitoring of employees by employers. The monitoring of employees provides the detection of deficiencies in time. The works are done in time and proceed more quickly. So, it is provided clear sailing.

There are disadvantages of monitoring of employees by employers. For example; the person can stress and he cannot concentrate to his works. Thus, works don't finish in time and some problems can occur in workplaces. People cannot relax.

In my opinion, employees should monitor by employers. Because people can neglect his works. Besides, they are interested with other things. People should not void. Because they can enter the other sites and plays the video games.

Finally, I can say that, advantages and disadvantages change person to person. For example, I can say that, there are many advantages of monitoring of employees. But he can say that, there are many disadvantages of workplace monitoring of employees by employers.

CIRRICULUM VITEA

Sevcan Bayraktar epni was born in Antakya, Hatay in 1987. She grew up in İstanbul and had her primary, secondary and high school years in İstanbul. She graduated from Şişli Anatolian High school in 2005. She graduated from Karadeniz Technical University, English Language and Literature Department in 2010. She worked in Private Ata Primary School from 2010 to 2011. She has been working both as a lecturer in Hakkari University since 2011 and as a senior researcher in scientific and technological research council of Turkey(TÜBİTAK) since 2013.