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

A CORPUS BASED STUDY ON L2 ENGLISH HYPOTHETICAL CONDITIONALS

Yunus Emre AKBANA

MASTER OF ARTS

ADANA / 2011

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

A CORPUS BASED STUDY ON L2 ENGLISH HYPOTHETICAL CONDITIONALS

Yunus Emre AKBANA

Supervisor: Asst. Prof. Dr. Cem CAN

MASTER OF ARTS

ADANA/2011

To Çukurova University Institute of Social Sciences,

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

Chairperson: Asst. Prof. Dr. Cem CAN
(Supervisor)

Member of Examining Committee: Prof. Dr. Hatice SOFU

Member of Examining Committee: Prof. Dr. Tahir BALCI

I certify that this thesis conforms to the formal standards of the Institute of Social Sciences.
….. / …. / 2011

Prof. Dr. Azmi YALÇIN
Director of Institute

P.S: The uncited usage of the reports, charts, figures, and photographs in this thesis, whether original or quoted for mother sources, is subject to the Law of Works of Arts and Thought.No: 5846
NOT: Bu tezde kullanılan ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunu'ndaki hükümlerine tabidir.

ÖZET

İKİNCİ DİL OLARAK İNGİLİZCEDEKİ VARSAYIMSAL ŞART KİPİ ÜZERİNE DERLEM TEMELLİ BİR ÇALIŞMA

Yunus Emre AKBANA

Yüksek Lisans Tezi, İngiliz Dili Eğitimi Anabilim Dalı
Danışman: Yrd. Doç. Dr. Cem CAN
Kasım 2011, 134 sayfa

Granger (2009) öğrenici derleminin yerini dört dilbilim araştırma alanının kesişim noktasında görmektedir: Derlem Dilbilim, Dilbilim Kuramları, İkinci Dil Edinimi ve Yabancı Dil Öğretimi (s. 13). Ayrıca, Granger (2002) öğrenici derleminin alandaki önemini şöyle vurgulamaktadır: Derlem Dilbilimin temel kurallarını ve yöntemlerini kullanarak yabancı dil veya ikinci dil edinimi çalışmalarının çeşitli amaçlarda kullanılmasını ve yabancı dil eğitimini geliştirmek için öğrenici dili ile ilgili, ileri düzeyde açıklamalar sağlanmasını hedefler (s. 4). Araştırmacılar öğrenici derlemini ilgilendikleri araştırma alanlarında kullanabilirler. Araştırmacılar "İngilizce aradilini etkileyebilecek olan çeşitli değişkenleri incelerken dil araştırma alanlarına bol miktarda veri sağlayan Uluslararası Öğrenici İngilizcesi Derlemi' ni (ICLE) yaygın olarak kullanmaktadırlar." (Can, 2010, s.16)

Dilbilgisel değişkenleri inceleyen çalışmalar arasında bağımlı tümcecikler karmaşık yapılar olarak yer almaktadır. Bağımlı tümcecikler üç ana başlık altında incelenmektedir; sıfat tümcecikleri, zarf tümcecikleri ve ad tümcecikleri. Bu çalışmada zarf tümcecikleri ele alınmıştır. Bu tümcecikler de kendi içlerinde sınıflandırıldığında; şart (eğer, şayet), neden-sonuç (çünkü) ve diğer tümceciklerden (ör: sonra, önce, -e karşın, -e kadar, olduğu gibi, olduğundan, -den dolayı) yapıları yer almaktadır (Biber, 2006, s.77). Bu çalışmada araştırma konusu olan yapı ise şart (şayet) kipi olarak belirlenmiştir.

Bu çalışmanın temel amacı, Amerikalı ve Türk katılımcıların anadillerindeki varsayımsal şart tümcelerini kullanımlarını ve Japon ve Türk katılımcıların ikinci dili olarak İngilizcedeki kullanımlarını karşılaştırarak ortaya çıkan az ve/ veya fazla kullanımın herhangi bir olası aradil özelliği gösterip göstermediğini ortaya koymaktır.

Bu amaçla, dört derlemde -Louvain Anadili İngilizce Olanların Komposizyonları Derlemi (LOCNESS), Türk Üniversite Derlemi (TUC) ve Uluslararası Öğrenici İngilizcesi Derlemi' nin (ICLE) alt derlemlerinden olan Uluslararası Türk Öğrenici İngilizcesi Derlemi (TICLE) ve Japon Öğrenici İngilizcesi Derlemi (JPICLE)- incelenen gerekçelendirme içerikli yazılı anlatımlardaki varsayımsal şart tümceleri incelenmiştir. Çalışmanın sonucunda, incelenen dört derlemin nitel sonuçları, Japon ve Türk öğrenicilerin yazılı anlatımlarında kullanmış oldukları varsayımsal şart tümcecikleri yapılarından belirli yardımcı eylemlerin olası bir İngilizce aradil özelliğinin var olabileceğini gösterdiği düşünülmektedir.

Anahtar Kelimeler: Öğrenici Derlemi, Varsayımsal Şart Tümcecikleri, Aradil, Derlembilim, LOCNESS, ICLE, JPICLE, TICLE TUC.

ABSTRACT

A CORPUS BASED STUDY ON L2 ENGLISH HYPOTHETICAL CONDITIONALS

Yunus Emre AKBANA

Master of Arts, English Language Teaching
Supervisor: Asst. Prof. Dr. Cem CAN
November 2011, 134 pages

Granger (2009) considers the place of learner corpus (LC) studies at the crossroads of four language research areas: corpus linguistics, linguistic theory, second language acquisition and foreign language teaching (p. 13). Granger also stresses the significance of learner corpus in the field: 'using main principles, tools and methods from corpus linguistics, it aims to provide improved descriptions of learner language used for a wide range of purposes in foreign/second language acquisition research and also to improve foreign language teaching' (2002, p. 4). So, researchers can utilize the learner corpus in their interest of research areas. In examining various variables that could affect English interlanguage, "researchers commonly use International Corpus of Learner English (ICLE), which contributes to language research areas with the abundance of hands-on material." (Can, 2010, p. 16)

A particular type of linguistics complexity in grammatical variation studies is 'dependent clause' which consists of three main structural clauses; namely, relative clauses, adverbial clauses and complement clauses. These structural clauses have their sub-divisions. In this study, 'adverbial clauses' is the main theme, which has three major meaning domains as conditional (if), cause (because), and other clauses (e.g., after, before, while, until, as, since, so that) (Biber, 2006, p. 77). In adverbial clauses, hypothetical conditional sentences composed by using conditional (if), has been selected as a topic of research examination in the observed argumentative written essays.

The ultimate purpose of the study is to reveal any possible effects of interlanguage by the over and/ or underuse of hypothetical conditional clauses in argumentative essays of Turkish students written both in Turkish and English Languages by native speakers

of English and EFL Turkish students, native speakers of Turkish in the light of four corpora: Louvain Corpus of Native English Essays (LOCNESS), Turkish University Corpus (TUC), and two other subcorpora of International Corpus of Learner English (ICLE); Turkish International Corpus of Learner English (TICLE) and Japanese International Corpus of Learner English (JPICLE). In this study, the quantitative results obtained from the four corpora have discussed any possible English interlanguage property on the use of particular modal verbs in hypothetical conditional structures examined in Japanese and Turkish EFL learners' written performances.

Key Words*:* Learner Corpus, Hypothetical Conditional Clauses, Interlanguage, Corpus Linguistics, LOCNESS, ICLE, JPICLE, TICLE, TUC.

ACKNOWLEDGEMENTS

I believe that this thesis would not have been completed without the guidance and support of the open-minded and original thinker, my supervisor, Asst. Prof. Dr. Cem CAN. He has been an outstanding guide for me that he has inspired me to love Corpus Linguistics and study on this research. He has been a model scholar for me to have a perspective on how a real researcher studies.

I would like to thank Prof. Dr. Hatice SOFU and Prof. Dr. Tahir BALCI for having accepted to be the members of the examining committee and their invaluable insights on this thesis. I would like to pay tribute again to Prof. Dr. Hatice SOFU for her professional advice that she has always encouraged me to think in wider horizons. I would like to extend my sincere thank teacher of teachers, Mehmet SEYİS, who has always allocated time to advise and suggest me in my presentations.

I would like to thank Gülten GENÇOĞLU that she has always supported and encouraged me to do the initiative. Now, I am very pleased to see that her and my theses have been completed, but we would never forget our memories that we had in our trips to Adana.

The invaluable comments of my friends, Asst. Prof. Dr. Namık ÜLKERSOY, Tuba DEMİRKOL, Reyhan AĞÇAM, Bahar BALABAN, Metin KURTAR and Stefan RATHERT, have shaped the study. So, I would like to thank them with my deepest gratitude. I would like to thank Mustafa ÇOBAN that we have always exchanged ideas with one another and supported each other with an endless motivation.

I would like to extend my sincere thank my family and my home mate, M. Akif KARA, who have always trusted in me without any hesitation. They have contributed to this study by their constant support, patience and guiding.

TABLE OF CONTENTS

Pages

ÖZET ..... **iii**
ABSTRACT ..... **v**
ACKNOWLEDGEMENTS ..... **vii**
TABLE OF CONTENTS ..... **viii**
ABBREVIATIONS ..... **xi**
LIST OF TABLES ..... **xii**
LIST OF FIGURES ..... **xiv**
LIST OF APPENDICES ..... **xv**

CHAPTER I
INTRODUCTION

1.1. Background to the Study 1
1.1.1. Corpus Linguistics: A Sub-Field of Linguistics or Methodology? 2
1.2. Statement of the Problem 4
1.3. Purpose of the Study 5
1.4. Research Questions of the Study 5
1.5. Assumptions and Limitations 6
1.6. Operational Definitions 6

CHAPTER II
REVIEW OF RELATED LITERATURE

2.1. History of Corpus Linguistics 9
2.1.1. Corpus Linguistics Before the 1950s 10
2.1.1.1. Views of Generative Linguists' towards CL 13
2.1.2. Corpus Linguistics between the 1950s- 1980s 19
2.1.2.1. The First Computational Corpus-Based Linguistics 21
2.1.2.2. The First Machine Readable Corpus 21
2.1.2.3. The First Grammar Based Corpus Linguistics 22
2.1.2.4. The First Inspiring Works of Neo-Firthians 22

2.1.3. Corpus Linguistics from the1980s- Present day 23
2.2. Types of Corpora..... 24
2.2.1. Specialised Corpus 26
2.2.2. Reference Corpus 26
2.2.3. Multilingual Corpus..... 27
2.2.4. Parallel Corpus 28
2.2.5. Diachronic Corpus..... 29
2.2.6. Monitor Corpus 30
2.2.7. Pedagogic Corpus..... 30
2.2.8. Learner Corpus..... 31
2.3. From Contrastive Analysis to Contrastive Interlanguage Analysis 34
2.4. Conditional Constructions 38
2.4.1. Hypothetical Conditionals 43
2.4.1.1. Hypothetical Conditionals in Turkish..... 45
2.4.1.2. Hypothetical Conditionals in Japanese 51

CHAPTER III

METHODOLOGY

3.1. Research Design..... 55
3.2. Materials 56
3.2.1. ICLE*v2* Design Criteria 56
3.2.1.1. Task Variables..... 57
3.2.1.2. Genre..... 58
3.2.1.3. Length 59
3.2.1.4. Topic 60
3.2.1.5. Task Setting..... 61
3.2.1.6. Learner Variables 62
3.2.2. LOCNESS..... 71
3.2.3. TUC 73
3.3. Wordsmith Tools Version 5. 74
3.4. Data Analysis 75

CHAPTER IV
DATA ANALYSIS AND RESULTS

4.1. Introduction 78
4.2. Analyses and Results 78
4.2.1. Analysis and Results of "would" and "would + have + V*en*" across Type 2 and Type 3 80
4.2.2. Analysis and Results of "could" and "could + have + V*en*" across Type 2 and Type 3 85
4.2.3. Analysis and Results of "might" and "might + have + V*en*" across Type 2 and Type 3 88
4.3. Error Analysis and Results 91
4.3.1. "would" and "would + have + V*en*" Error Analyses and Results in Type 2 and 3 If and Main Clauses 94
4.3.1.1. "would" Errors in Type 2 If Clauses 95
4.3.1.2. "would" Errors in Type 2 Main Clauses 103
4.3.2. "could" and "could + have + V*en*" Error Analyses and Results in Type 2 and 3 If and Main Clauses 105
4.3.2.1. "could + have + V*en*" Errors in Type 3 If Clauses 106
4.3.3. "might" and "might + have + V*en*" Error Analyses and Results in Type 2 and 3 If and Main Clauses 107
4.3.3.1. "might" Errors in Type 2 If Clauses 108

CHAPTER V
CONCLUSION AND DISCUSSION

5.1. Introduction 111
5.2. Conclusions 111
5.3. Implications for English Language Teaching 116
5.4. Suggestions for Further Studies 117

REFERENCES 119
APPENDICES 128
CURRICULUM VITAE 133

ABBREVIATIONS

BOE: Bank of English

BNC: British National Corpus

CA: Contrastive Analysis

CIA: Contrastive Interlanguage Analysis

CL: Corpus Linguistics

DDL: Data-Driven Learning

EFL: English as a Foreign Language

ELT: English Language Teaching

ESL: English as a Second Language

ICE: International Corpus of English

ICLE: International Corpus of Learner English

JPICLE: Japanese International Corpus of Learner English

L1: First Language / Mother Tongue

L2: Second Language

LC: Learner Corpus

LOCNESS: Louvain Corpus of Native English Essays

NL: Native Language

NNS: Non-Native Speakers

NS: Native Speakers

SLA: Second Language Acquisition

TICLE: Turkish International Corpus of Learner English

TL: Target Language

TUC: Turkish University Corpus

LIST OF TABLES

Page

Table 1: Corpus Linguistics from the 1950s to the 1980s by Four Subsections 20
Table 2: Types of Corpora 25
Table 3: The Currently Existing Learner Corpora 31
Table 4: Multiple Comparison Approach 37
Table 5: Types of Conditionals 40
Table 6: The Verb Forms with Hypothetical Conditionals 44
Table 7: The Variants of –s*A* in Turkish Conditionals 45
Table 8: ICLE Task and Learner Variables 57
Table 9: Proportion of Argumentative Essays 58
Table 10: Average Essay Length in ICLE 59
Table 11: Top Ten Essay Topics in ICLE 60
Table 12: Proportion of Task Conditions in ICLE 61
Table 13: Learners' Age and Gender Distribution in ICLE 62
Table 14: CEF Results- 20 Essays Per Subcorpus in ICLE 67
Table 15: Distribution of Essays/Words Per Subcorpus in ICLE 68
Table 16: Turkish Subcorpus of ICLE 69
Table 17: Japanese Subcorpus of ICLE 69
Table 18: Selected Proportion of LOCNESS Data Distribution 73
Table 19: Selected Proportion of TUC Data Distribution 73
Table 20: All Variables Analyzed Across Four Corpora- TICLE, JPICLE, LOCNESS, TUC- with Their Abbreviations Processed in SPSS 76
Table 21: Overall Frequency and Ratio Analysis of the Target Variables in Type 2 and 3 If-Clauses among four Corpora 79
Table 22: Overall Frequency and Ratio Analysis of "would" and "would + have + V*en*" in Type 2 and 3 If-Clauses among four Corpora 81
Table 23: Overall Pearson Chi-Square Results of "would" across four Corpora 83
Table 24: Overall Pearson Chi-Square Results of "would + have + V*en*" across four Corpora 84
Table 25: Overall Frequency and Ratio analysis of "could" and "could + have + V*en*" in Type 2 and 3 If-Clauses among four Corpora 85
Table 26: Overall Pearson Chi-Square Results of "could" across four Corpora 87

Table 27: Overall Pearson Chi-Square Results of "could + have + V*en*" across four Corpora. 88
Table 28: Overall Frequency and Ratio Analysis of "might" and "might + have + V*en*" in Type 2 and 3 If-Clauses among four Corpora 89
Table 29: Overall Pearson Chi-Square Results of "might" across four Corpora..... 90
Table 30: Overall Pearson Chi-Square Results of "might + have + V*en*" across four Corpora 90
Table 31: Overall View of the Modal Verb Uses Across If and Main Clauses 91
Table 32: Overall Frequency and Ratio Analysis of Target Variable Errors in Type 2 and 3 If and Main Clauses between two Learner Corpora 92
Table 33: Overall Frequency and Ratio Analysis of "would" and "would + have + V*en*" Errors in Type 2 and 3 If and Main Clauses between two Learner Corpora.. 94
Table 34: Overall Pearson Chi-Square Results of "would" Errors in If Clauses across four Corpora..... 96
Table 35: Overall Pearson Chi-Square Results of "would" Errors in Main Clauses across four Corpora 103
Table 36: Overall Frequency and Ratio Analysis of "could" and "could + have + V*en*" Errors in Type 2 and 3 If and Main clauses between two Learner Corpora. 105
Table 37: Overall Pearson Chi-Square Results of "could + have + V*en*" Errors in If Clauses across four Corpora 106
Table 38: Overall Frequency and Ratio Analysis of "might" and "might + have + V*en*" Errors in Type 2 and 3 If and Main clauses between two Learner Corpora. 108
Table 39: Overall Pearson Chi-Square Results of "might" Errors in If Clauses across four Corpora..... 109

LIST OF FIGURES

Page

Figure 1. Multilingual Comparisons of English with Translations into Other Languages in The Oslo Multilingual Corpus 28

Figure 2. Contrastive Interlanguage Analysis..... 35

Figure 3. Multiple Comparison of L1, TL And IL Corpora. 36

Figure 4. Dependent Clause Types Across Registers 38

Figure 5. Breakdown of Finite Adverbial Clause Types Across Registers 39

Figure 6. A Screenshot of Task and Learner Variables Extracted From ICLE*v2*. 64

Figure 7. A Screenshot of Task Variables, Learner Variables and Preference of Institution Extracted From ICLE*v2*..... 65

Figure 8. A Screenshot of Concordance 'Would' within TICLE Extracted from ICLE*v2* 66

LIST OF APPENDICES

Appendix 1: The Institutions Codes Manipulated in The Study for TICLE..... 128
Appendix 2: The Institutions Codes Manipulated in The Study for JPICLE..... 129
Appendix 3: Corpus Collection Guideline for Compiling TUC..... 130
Appendix 4: The verb formation of the word "-sev" and conditional marker "–sA".. 132

CHAPTER I

INTRODUCTION

1.1. Background to the Study

Language is organic, changing and developing with each generation reflecting the era, society and dialect. As language is a lively subject, its growth and development tend to conform with the crucial needs of the era. Its social nature is stated as a "… social product deposited in the brain of each individual …" (Saussure, 1966, cited in Harris, 1993, p.18). In other words, individuals of each society are the primary sources of its language. Corpus linguistics, taking as its basis this assumption, has been one of the main research areas in language studies in the last century, specifically, in the last two or three decades. It has been widely embedded in the minds of linguists who wish to observe language as a social phenomenon "… searching through screen after screen concordance lines and wordlists generated by computer software …" (O'Keeffe and McCarthy, 2010, p. 3). Any kind of language data in small texts are collected and stored in different ways as punch cards or computer software programmes even with small bits of papers dating back to 1350s with Amarna letters. This data archive is titled as *corpus* in contemporary linguistic terminology. O'Keeffe and McCarthy (2010, p. 3) suggest that corpus is "… a big collection of smaller texts. "

Corpus linguistics, which has a long history, can be defined as "the study of language which serves with its real life examples of sources" (McEnery and Wilson 2001, p.1). Since corpus includes real life examples of sources, it serves with the natural data on use of language. This natural storage of language data has paved the way for scholars and linguists to make their research on. There are many advantages of using such data in research that will be touched upon by the related studies in the following sections.

English corpus linguistics has taken its place among the main interests of many linguists all over the world such as Baker, Carter, Chomsky, Francis, Granger, Halliday, Hoey, Kucera, Leech, McEnery, McCarthy, O'Keeffe, Quirk, Sinclair, Svartvik, Wilson, and so many on. There have been a vast number of contributions of these

prominent scholars to the field of corpus linguistics. Recent uses of corpus linguistics have been introduced as a branch of linguistics or methodology in applied linguistics.

On one hand, corpus linguistics has been perceived as a sub-field of linguistics like the other areas of linguistic studies in sociolinguistics, forensic linguistics, pragmatics, semantics, syntax and the like. On the other hand, it has been applied as a methodology on research in applied linguistics and utilized as a reference tool for grammar books, dictionaries, course books, and quantitative and qualitative empirical studies in its history. Whether it should be considered as a branch of linguistics or an instrument in research methodology has been a theme of debate over decades in the studies of Granger (2002), McEnery and Wilson (2001), Meyer (2002), O'Keeffe and McCarthy (2010) and Sinclair (2000).

1.1.1. Corpus Linguistics: A Sub-Field of Linguistics or Methodology?

A sub-field of linguistics can be defined as "… the branch of linguistics such as sociolinguistics, psycholinguistics, forensic linguistics, syntax, semantics, phonetics, and so on …"(McEnery and Wilson, 2001, p. 2). It is difficult to say that corpus linguistics takes its place among these branches of linguistics; however, it can be utilized to support any of these branches with relevance to the interest of the research studies such as lexicography and grammar. Meyer (2002, p. 27) suggests that corpora have been proven to be valuable resources in many other disciplines of linguistics as a tool for "… creating dictionaries, studying language change and variation, understanding the process of language acquisition, and improving foreign- and second-language instruction." Many lexicographers have used corpora in their studies to explain the language-related patterning pertaining to word meaning and usage. O'Keeffe, McCarthy and Carter (2007, p. 17) point out that the pioneering work in this area of research is Sinclair's (1980) *Collins Birmingham University International Language Database* (COBUILD) *Project*. It has been quite a productive project presenting 16 dictionaries and grammar of the English language. It has been believed that the most influential output of this project is the *Collins COBUILD English Language Dictionary* (1987, 1995, 2001, 2003) and the *Collins COBUILD Grammar Patterns* series (1996, 1998). COBUILD Project (1980) has paved the way for all the major publishers "… to base their dictionaries on corpora studies to examine and

describe the grammars of English language" (cited in O'Keeffe, McCarthy and Carter, 2007, p. 17).

Corpus linguistics has provided the relationship between lexis and grammar. This relationship has been investigated in detail for language teaching by Hunston and Francis (2000). In their study, corpus-based lexico-grammatical research is the main focus. As this relationship has been so influential in grammar studies, major grammars of English are now corpus based, for example, Biber, Johansson, Leech, Conrad, and Finegan (1999), Carter and McCarthy (2006), Quirk, Greenbaum, Leech, and Svartvik (1985) and Sinclair (1990).

According to McEnery and Wilson (2001, p. 2), corpus linguistics is a methodology which does not need to have a description like any branches of linguistics such as discourse, pragmatics, phonology, morphology, syntax, semantics and so on. However, it can be considered as a sub-field of linguistics with respect to the cooperation with any of these branches of linguistics. Sinclair (2000, p. vii) states "… corpus linguists have recently been making a case for their work being treated as a sub-discipline and not just a source of information or an innovative methodology." Within the same line of reasoning, there have been many studies, several of which focus on the transformation of corpus linguistics to disciplines within linguistics. Considering what Sinclair emphasizes, McEnery and Wilson (2001, p. 2) state that a methodology can take place for any language use and a corpus-based methodology can be applied to some linguistic enquiries such as corpus-based syntax, corpus-based semantics or corpus-based pragmatics. Vice versa can also be possible, i.e. non corpus-based syntax, non-corpus-based semantics, non-corpus-based pragmatics and so on. Not only is corpus linguistics just an analysis of linguistics, but it also provides an indispensable amount of examples of language use by different groups, individuals or studies on any branch of linguistics. Thus, corpus linguistics is a sort of approach that can be shaped by the assistance of the linguistic uses.

Granger (2002) concludes the discussions on this norm by defining corpus linguistics as a "… linguistic methodology. It is neither a new branch of linguistics nor a new theory of language, but the very nature of the evidence it uses makes it a particularly powerful methodology having the potential to change perspectives on language." (p.4)

Corpus linguistics has been an approach applied in many sub-fields; however, the important question to ponder is 'how has corpus linguistics been so influential on

many research areas?' Examining the studies conducted in the previous literature of corpus linguistics may form the answer to the question. McEnery and Wilson (2001) and O'Keeffe and McCarthy (2010) have pointed out that corpus linguistics goes back to 1880s and it is a modern term now. In its history, generative linguists did not support corpus linguistics; however, it had many supporters such as Granger, Halliday, Sinclair and Svartvik especially in the 20th century and has had in the 21st century. Its history has been fitted in three periods in this current study; before 1950s (early corpus linguistics period), 1950s-1980s period, and 1980s- present day.

1.2. Statement of the Problem

There has been a numerous number of corpus based research studies (Biber, 2006; Ghadessy, Henry and Roseberry, 2001; Granger, 1996a; 1997; 1998; 1999; Granger and Meunier, 2008; Granger and Paquot, 2009; Mahlberg, 2005; O'Keeffe, McCarthy and Carter, 2007) on linguistic enquiries in due course of teaching and learning processes that teachers have encountered.

Those studies have paved the way for researchers to study any linguistic enquiry via corpus. However, there are few studies on conditionals (Römer, 2007; Dancygier and Sweetser, 2005), and hypothetical conditionals have not been investigated as a corpus based research study among Turkish university EFL students. Taking as its basis this observation, the current study focuses on difficulties Turkish EFL learners experience in using hypothetical conditionals; the data obtained from this examination are then compared with the data from Japanese EFL learners regarding the same subject. As Japanese is a verb-final language like Turkish, Japanese EFL learners' written productions have been preferred in comparability with that of Turkish EFL learners.'

Many teachers of English encounter students tending to misuse hypothetical conditionals because of its linguistic complexity. The difficulty which Turkish and Japanese EFL learners face is explained in linguistic terminology as "back-shifting. " (Ke, 2004, p. 9) In these constructions, they are supposed to use the past simple tense instead of the present simple tense and the past perfect tense instead of the past simple tense. This tense movement is called as back shifting and is used in unreal conditions, in other words hypothetical conditional statements. Back shifting is considered as a past time reference as past-past. Palmer (1990, p.170) claims that "there is both past time

and unreality, and the past tense, therefore, needs to be marked twice" (cited in Ke, 2004, p.9). That is to say, back-shifting can be achieved by using a past structure in the sentence, but it is not enough especially for Turkish EFL learners, because Banguoğlu (1990) and Hengirmen (1998) share a common view that in Turkish ; judging a sentence whether it is hypothetical or not is not only the matter of examining the structure, it needs to be examined semantically as well. In using hypothetical conditionals, English can be defined a marked language when compared to the use in Turkish; because in English, a hypothetical conditional clause can be appointed by just checking the syntax of a sentence to define hypothetical conditional clause.

1.3. Purpose of the Study

The ultimate purpose of this study is to reveal any possible effects of mother tongue and/or possible interlanguage properties to be found in the written samples extracted from JPICLE and TICLE by examining the overuse and/or underuse of hypothetical conditionals. This study also aims at comparing the usage between Japanese and Turkish EFL learners' with that of American native speakers as well as Turkish native speakers in their mother tongue by means of their written argumentative essays. By comparing the usage in both learner groups, we aim at investigating the errors in hypothetical conditionals, but as the current study is quantitative in character, we depict only the common error types committed by Japanese and Turkish EFL learners conducting a basic error analysis.

1.4. Research Questions of the Study

This study aims to find out the answers of the questions below, which are supposed to shed light on any possible learning difficulty of hypothetical conditionals.

1. What is the distribution of the hypothetical conditional sentences from most frequently use to the least within native speakers of English (American university students) and Turkish (Turkish university students) and non-native speakers of English (Japanese &Turkish university students)?

2. Do advanced Japanese and Turkish EFL learners use Type 2 and 3 if-conditional statements referring to hypothetical conditionals in the same extent as native speakers of American university students?
3. Which of the modal verb forms investigated in hypothetical conditionals are most commonly marked among the argumentative written essays in L2 English by Japanese and Turkish non-native speakers of English, in L1 English by American university students and in L1 Turkish by Turkish university students?
4. If there is a significant difference in the use of hypothetical conditionals used by Japanese & Turkish EFL learners across the American Native use and Turkish L1 use, is this due to
    a) mother tongue effect?
    b) the property of interlanguage?
5. To what extent do Japanese & Turkish non-native speakers of English use hypothetical conditionals correct in syntactic form?

1.5. Assumptions and Limitations

Analysing hypothetical conditionals should not be restricted to learner data gathered from Japanese and Turkish learners of English, and the native users of English, though, Turkish and Japanese, as verb-final languages, are very close to each other syntactically as defined by Kuno (1978, p.58). Syntactically close languages should be included in this sort of research in order to observe property of interlanguage at a broader scale. Another research done within the scope of the present study, Turkish language written data enables the researcher to observe any possible effect of mother tongue on L2 usage, however, Japanese language written data should also be gathered since it may shed light on comparing the effects of mother tongue in a wider context.

1.6. Operational Definitions

Contrastive Analysis (CA): "Contrastive analysis is a way of comparing languages in order to determine potential errors for the ultimate purpose of isolating

what needs to be learned and what does not need to be learned in a second-language-learning situation" (Gass and Selinker 2008, p.96).

Contrastive Interlanguage Analysis (CIA): According to Granger (1996b, p.43), CIA is "Contrasting and comparing the use between native and learner varieties of one and the same language" .

Corpus: "Corpus is a computerised collection of authentic texts, amenable to automatic or semi-automatic processing or analysis. The texts are selected according to explicit criteria in order to capture the regularities of a language, a language variety or a sub-language" (Tognini Bonelli, 2001, p. 55).

Corpus Linguistics (CL): According to McEnery and Wilson (2001, p.1), it is the study of language, which serves with its real life examples of sources.

Interlanguage: A linguistic system in its own right of a second language learner who has not become fully proficient yet but is approximating the target language: preserving some features of their first language, or overgeneralizing target language rules in speaking or writing the target language and creating innovations (Selinker, 1972).

English Interlanguage: According to Granger (1998, p.6), it is the English L2 data supplied from one or more different mother tongue backgrounds.

International Corpus of English (ICE): A corpus representing the varieties of English.

International Corpus of Learner English (ICLE): A learner corpus consisting of English L2 data supplied from 16 different mother tongue backgrounds.

Japanese International Corpus of Learner English (JPICLE): A learner corpus including the Japanese EFL learners' argumentative essays in English.

Learner Corpus (LC): "Systematic, computerized collections of texts produced by language learners" (Nesselhauf, 2005, p.40).

Louvain Corpus of Native English Essays (LOCNESS): A native corpus including British and American Students' written essays.

Turkish International Corpus of Learner English (TICLE): A learner corpus including the Turkish EFL learners' written essays in English.

Turkish University Corpus (TUC): A reference corpus including Turkish University Students' argumentative essays in Turkish compiled by Çukurova University.

CHAPTER II

REVIEW OF RELATED LITERATURE

2.1. History of Corpus Linguistics

There is a common viewpoint that corpus has been a very young subject area of research interest; however, the proof in literature sets forth the existence of corpus by referring to the first concordancing biblical texts. In addition, corpus has been even used in archaeological corpora pertaining to BC 1330s. At the very early times of corpus history, corpus studies took part in the studies of the interests of monks who prepared indexes and prepared concordance lists for bibles. It was the works manipulated roughly about A.D. 600s, but a very first concordance bible was produced in 1230 by Hugo De Saint Cher (Hugo de Sancto Charo) (http://en.wikipedia.org/wiki/Bible_concordance). However, O'Keeffe and McCarthy (2010) state that since then, there have been many other biblical concordance studies including Cruden's 1737 "A Complete Concordance to the Holy Scriptures" and Strong's 1890 "Exhaustive Concordance of the Bible" (p.3).

Chronologically, the history of corpus can be divided into three main parts in this study. The first one refers to corpus linguistics before the 1950s. Corpus data was mainly processed on bits of papers at those times. Large teams of scholars or monks were employed to process the data for corpus studies. This early period of corpus linguistics describes the linguistics before the Chomskyan Revolution, which was the starting point for the second period in corpus history. Chomksy has been a prominent scholar with Transformational-Generative Grammar (TGG) theory to trigger the minds of linguists on many debates over decades on the nature of data in linguistics. The generative grammarians, mainly Chomsky, did not support corpus as a native speaker would never need a corpus. The opposing views towards corpus have formed a new period in the history of corpus linguistics. That is, the first period is different from the second in terms of the views of scholars as stated in a sort of consensus in many studies of researchers such as Dodd's (2000); McEnery and Wilson's (2001) and Meyer's (2002).

The second period is of relevance between the 1950s and 1980s. In the late 1950s, the first computer-generated concordances appeared by using punch-card

technology. However, throughout this period corpus linguistics attracted less interest relatively to the first and, of course, the third period in its history. There was not as much research related to the second period as conducted in the last period cycle (the 1980s – present day corpus period). By the development of computational data, corpus has moved from the first period on bits of papers to punch cards, and subsequently, to the third period to machine readable corpora, besides, it has become a very effective approach in language teaching and applied linguistics with many studies (McEnery and Wilson, 2001, pp. 7-21).

In the third period, corpus has moved from thousands of word collections to billions by the advance of computer technology. Finding a word or a phrase in billions even in thousands of sources and many different genres has become achievable in seconds via computer technology. As McEnery and Wilson (2001, p. 17) suggest, it is also a very sophisticated development when compared to its early times since it is much cheaper and time-saving to process, search and retrieve the data. Below is a detailed history of the three periods mentioned briefly above.

2.1.1. Corpus Linguistics Before 1950s

The origin of corpus goes back not only to the biblical studies, but also, to governmental issues that Cılız (2010, p. 2) suggests "one of the first samples of corpus even dates back to BC.1350s with Amarna letters. They form a diplomatic archive of correspondence between the Egyptian administration, and its representatives written in Akkadian." However, it was also touched upon in the literary works of Shakespeare as in Becket's 1787 *A Concordance to Shakespeare.* In language studies corpus approach began "… with a large collection of recorded utterances from some language corpus. The corpus was subjected to a clear, stepwise, bottom-up strategy of analysis" as Harris (1993, p. 27) summarised.

As McEnery and Wilson (2001, p. 3) claim it is very difficult to find the term 'Corpus Linguistics' in this period (before 1950) since it was not formed in a clear cut research area and popular in language studies. It is called 'Early Corpus Linguistics' referring to the use of corpora prior to the 1950s. Below is a brief overview on several studies conducted predating the 1950s.

In modern day: Corpus linguistics has been accepted as *electronic corpus linguistics*, but it does not mean that there were no language corpora before computers.

In 1755, Dr. Samuel Johnson already provided the first comprehensive dictionary of English. It was the result of nine years of paper-based corpus work between 1746- 1755. It was the product of the application through endless slips of paper logging samples of usage from the period 1560 to 1660. 173 years later, there was another lexicographical corpus used for the *Oxford English Dictionary* (originally entitled *New English Dictionary*). It took fifty years to complete, since it had many stages to be produced as it was done manually. Subsequently, in 1928, the first edition was completed and had accumulated over four million citation slips from two thousand volunteer readers. It was a very large team of work to its time and with no help of the computer (O'Keeffe and McCarthy, 2010, p.4). However, not having access to a computer, the researchers of the time were in depth with very difficult workloads, so as a solution, they received assistance from others. One of the researchers of the time, Murray (1977), solved the problem. Murray's many children did much of the work of alphabetising and sorting the slips. Before computers, there were not only lexicographical studies, but also, dialect studies, such as *The Existing Phonology of English Dialects* provided by Ellis (1889). In this study, over 800 researchers participated in the project to provide information and assistance to the core researcher team (Svartvik, 2007, pp. 12, 13).

In first language acquisition research, between 1876 and 1926, the studies were carried out with the data obtained from parental diaries including the child's utterances. As McEnery and Wilson (2001, p.3) cited, researchers, as Preyer (1889) and Stern (1924), were the main linguists who carried out many comments on those language acquisition studies, but now, the views of Ingram (1978 cited in McEnery and Wilson, 2001, p.3) have been put forward in accordance with the language acquisition studies and there are electronic recording programmes (which also transcribes dialogues/monologues automatically) instead of parental diaries recordings.

At the Early Corpus Linguistics period, there were some corpora compiled large in size by several studies. In the studies of Baker, Hardie, & McEnery (2006, p.50), McEnery and Wilson (2001, p.3) it is emphasized that one of the largest corpora at those times was of Kaeding's (1897) who used a large corpus of German – some 11 million words – to collate frequency distributions and sequences of letters in German. The corpus was impressive in terms of size alone for its time and even now , it can be compared favourably with some modern corpora in terms of size alone.

Corpus was popular in second language pedagogy in the first half of the twentieth century as it provided many vocabulary lists for foreign learners, and it

became as a movement at that time. As cited in McEnery and Wilson (2001, p.4), Bongers (1947); Fries and Traver (1940); Thorndike (1921) and Palmer (1933) have been the important studies in corpus linguistics for that time.

Eaton (1940) carried out a corpus-based study comparing the frequency of word meanings in Italian, German, French and Dutch. This demonstrates that comparative linguistics had also a strong link with corpus linguistics. Since 1990s, there have been many corpus-based contrastive analyses such as Johansson's (1997) and Johansson and Oksefjell's (1998) by analysing grammar in their corpus based studies. (McEnery and Wilson, 2001, p.4)

Comparative linguistics has had a strong link with 'International Corpus of English' (ICE) by corpus-based studies as ICE represents the varieties of English from around the world. The availability of such a richness of varieties has accumulated for comparing the usage of any language related patterning in different dialects and societies all over the world. Over twenty countries where English is the first language or an official second language are included in ICE. The current list of participant countries are; Australia, Cameroon, Canada, East Africa (Kenya, Malawi, Tanzania), Fiji, Great Britain (parsed), Hong Kong, India, Ireland, Jamaica, Kenya, Malta, Malaysia, New Zealand, Nigeria, Pakistan, Philippines, Sierra Leone, Singapore, South Africa, Sri Lanka, Trinidad and Tobago, USA. The project began in 1990 with the primary aim of collecting material for comparative studies of English worldwide. (http://en.wikipedia.org/wiki/International_Corpus_of_English)

In the early periods of corpus linguistics, as it was very difficult for scholars and linguists to compile a corpus or carry out a corpus based study, they needed a considerable amount of working hours to process the data. Therefore, they were not interested in corpus linguistics. However, taking the reliability and validity of real life language input into consideration, they were engaged in many different types of texts and language areas. The researchers got expertise in their interest of areas by collecting language data and forming corpus with the support of the foundations of the university presses and institutions they worked for. Therefore, researchers at each branch of linguistics adapted their studies to their expertise research areas. Each of their studies inspired to each other, hence, the modern day corpora have their roots in those valuable studies carried out before 1950s. As O'Keeffe and McCarthy (2010, p.4) state and confirm others in the terminology of corpus linguistics, those linguists supporting corpus linguistics are regarded as "structural linguists". Later, in 1950s generative

grammarians did not support corpus linguistics and their views were based on Chomsky's criticisms of the corpus. In the words of McEnery and Wilson (2001, p.4), it is stressed that corpus linguistics was widespread for many years until 1950s, which was the period when corpus linguistics fell into disgrace by one of the most influential linguists ever; Chomsky.

2.1.1.1. Views of Generative Linguists' Towards CL

Chomsky's criticisms on corpus linguistics are the fundamental points of forming the generative linguists' views towards corpus linguistics. Chomsky's criticisms have shed light on the foresight that have directed the current modern corpus so effective in size, comparability, providing lots of data extracts in seconds and in a cheap way. The criticisms firstly caused debates between rationalists and empiricists. Rationalists supported artificially induced observations, whereas empiricists supported naturally occurring observations. Debates on corpus linguistics have been of relevance to two groups; namely, empiricist and rationalist linguists. Chomsky's criticisms against the empiricist linguists have received responses from other linguists who supported the real life examples of natural data.

Rationalist theory is based on introspective judgements; it may be reflections and claims of a native speaker of a language. Empiricist approach is characterised with the observation of natural data by the vehicle of corpus. They both have advantages and disadvantages, which will be discussed in the following sections. However, to be sure of the debate for now; Chomsky supports the rationalist theories (McEnery and Wilson (2001, pp.5, 6). For a more detailed explanation of rationalism, the study of Chomsky (2006) introduces rationalism in a more widely context.

Chomsky (1988 cited in McEnery and Wilson, 2001, pp. 5, 6) rejected the corpus as a source of evidence in linguistics enquiry by psychological reality and cognitively plausible models of language. What Chomsky emphasizes is that a linguist does never need a corpus, because a linguist must model the language competence rather than performance by his/her own introspection. "Chomsky's initial criticism of corpus is that it is by its very nature a collection of externalised utterances; it is solely the performance data and, as such, it must be a poor guide to modelling linguistic competence." In other words, Chomsky supports the idea that it was competence rather than any performance that the linguist was trying to model, because performance is just

a poor mirror of competence. Gass and Selinker (2008) define competence as being the representations of abstractions, i.e. the knowledge of the underlying structure of language and performance as being a part of putting language knowledge to use at a given point in time i.e. the language used. Competence is our tacit, internalised knowledge of a language and performance is the external evidence of language competence and its usage on particular situations when, especially, other factors may affect its form rather than competence. These factors might be the short-term limitations or the speaker's being drunk. What Chomsky meant is that "… performance cannot be the complete reflection of our competence …", and as the matter is within the linguists; they are more primarily interested in competence not in performance regarding to a kind of tool, corpus. The linguists get assistance of their competence by making use of their intuitions, and as a consequence, they need intuition rather than a performance based tool, corpus (McEnery and Wilson, 2001, p.6).

Corpus approach has to count the words and sentences in a language determining the frequencies. A natural language has infinite sentences, which sheds light on the idea that all the sentences in a language cannot be gathered and counted by the medium of corpus. The related assumption here related with Chomsky's criticism is that 'Language is finite' for what corpus linguistics serves. The number of sentences in a language is uncountable, so the number of sentences in a natural language is infinite as well. Below is a short example of recursion, i.e. repeatedly rules of the production of sentence (McEnery and Wilson, 2001, p. 9).

Example 1.

'The official news agency carried excerpts from a speech by Brezhnev at a Kremlin dinner for visiting Cambodian leader Heng Samrin'.

The above sentence is taken from the Associated Press Corpus. Below is a broken sentence by McEnery and Wilson (2001, p. 9)

from a speech
PP : Prep NP
NP : AT N

from a speech by Brezhnev
PP : Prep NP
NP : AT N PP
PP : Prep NP

NP : PropN

from a speech by Brezhnev at a Kremlin dinner
PP : Prep NP
NP : AT N PP
PP : Prep NP
NP : PropN PP
PP : Prep NP
NP : AT PropN N

This example can be continued at this level of repetition cycle, so recursion can go on infinitely with these rules, that is to say, with finite rules, infinite sentences! Language is uncountable; therefore, no finite corpus can represent the language in every detail that occurs. As no corpus can represent the language in every piece of information, Sinclair (1991, p. 18) suggests that a corpus should be as large as possible.

Corpus serves also with the frequencies of any linguistic enquiry. However, Chomsky does not support the idea that a linguist would need a mere list of language items to illustrate how many times a language related item occurs. In McEnery and Wilson (2001, p. 10) Chomsky's claim is displayed that one can easily support the idea that *'I live in New York'* is more likely than *'I live in Dayton Ohio'* as the former can be said by more people than the latter. As an inference, Chomsky regarded this as a major failing of early corpus linguistics. Fillmore (1992, cited in McEnery and Wilson, 2001, p. 10) suggests that corpus linguistics is just a matter of collection of words and sentences even in zillions, so the linguist can find facts from the secondary facts from the primary facts. Both support the idea that introspection can achieve the work of any corpora. Chomsky (1984, p.44, cited in McEnery and Wilson, 2001, p. 11) defines the power of introspection by a short sentence that "if you sit and think for a few minutes, you're just flooded with relevant data." At this point, it is very useful to provide an exchange for a clearer view of Chomsky on his supposed power of introspection:

Chomsky : The verb *perform* cannot be used with mass word objects: one can *perform a task,* but one cannot *perform a labour.*

Hatcher : How do you know, if you don't use a corpus and have not studied the verb *perform?*

> Chomsky : How do I know? Because I am a native speaker of the English language.
>
> (Hill, 1962, p. 29, cited in McEnery and Wilson, 2001, p. 11)

McEnery and Wilson (2001, p.11) claim that Chomsky was not right, because British National Corpus reveals several results of '*perform magic'*, therefore, being a native-speaker and making use of its "… intuition merely allowed Chomsky to be wrong." For all that views, intuition can sometimes help us saving time instead of searching corpus. As a response to one of Chomsky's main criticisms, we can easily accept that being a native speaker of a language or linguist, i.e. intuition, does not serve for being the sole explicandum of linguistics. However, to depict the views on this norm, following is a conversation between Robert Lees and W.Nelson Francis that Jan Aarts says to his students every year at the University of Nijmegen.

> In 1962, when I was in the early stages of collecting the Brown Standard Corpus of American English, I met Professor Robert Lees at current interests; I said that I had a grant from the U.S. Office of Education to compile a million-word corpus of present-day American English for computer use. He looked at me in amazement and asked, 'Why in the world are you doing that?' I said something about finding out the true facts about English grammar. I have never forgotten his reply: 'that is a complete waste of your time and the government's money. You are a native speaker of English; in ten minutes you can produce more illustrations of any point in English grammar than you will find in many millions of words of random text.
>
> Francis (1982, pp.7,8 cited in Svartvik, 2007, p.19)

Sometimes the corpus we are using cannot be enough to identify whether a phrase is grammatical or ungrammatical, because it is finite. Then, linguist's intuition shines through the part of decision as it is encouraged by the competence of the linguist. For example, a sentence *'He shines Tony books'* is not grammatical. The BNC does not include such a sentence, and then intuition has the control on the process of giving a decision. However, it would be wrong to deduce that it's a true grammatical statement by taking into consideration some constructions like; *'He gives Keith the stare … , … he*

owes Demster a lot of money … , … he pushes Andy down into the ferns …' Those constructions incorporate the same syntactical items, but the construction *'He shines'* is not followed by a proper name in the BNC, though the above listed examples occur. This is to stress again that as language is non-finite and a corpus is finite, the problem exists and we cannot completely decline intuition (McEnery and Wilson, 2001, pp. 12, 13).

The criticisms made against the use of corpora can be summarized in three aspects. Firstly, since corpus is performance, it should not be modelled. According to Chomsky, it is not the work of a linguist to study by the enumeration and description of performance via corpus, instead the linguist should rely on and perform by his/her introspection and explanation of linguistic competence. Secondly, if the opposite situation is held as an aim for linguists to enumerate and explain by corpus, it seems impossible because, languages are non-finite. Finally, a corpus could not even be the best mere solution for our grammatical or structural judgements.

McEnery and Wilson (2001, p.13) summarise the debates on first language acquisition perspective that on the contrary partly introspection is beneficial for language uses; introspection is no valid use in language acquisition studies of children. How can somebody expect from a 19-month-old child to have a state of mind on differentiating a verb is a verb or a noun is a noun? In order to achieve this, the child must have metalinguistic knowledge. However, it would not be acceptable for a child who is at one-word stage of language to have metalinguistic awareness and develop it. We must accept that a non-corpus linguist uses introspection having a process that may not be all systematic and the linguist can devise extremes. However, a corpus-linguist can easily come up with all the relevant data to the discourse she/he is studying on with the reflection of frequency-based data. On one hand, a corpus linguist has to incline to the corpus; on the other hand, non-corpus linguist is to have his/her own imagination. At this point, Fillmore (1992) provides satire in describing the non-corpus linguist and corpus linguist by the following:

> Armchair linguistics does not have a good name in some linguistics circles. A caricature of the armchair linguist is something like this. He sits in a deep soft comfortable chair, with his eyes closed and his hands clasped behind his head. Once in a while he opens his eyes, sits up abruptly shouting, 'Wow, what a neat fact!', grabs his pencil, and writes

> something down. Then, he paces around for a few hours in the excitement of having come still closer to knowing what language is really like. (There isn't anybody exactly like this, but there are some approximations). Corpus linguistics does not have a good name in some linguistics circles. A caricature of the corpus linguist is something like this. He has all of the primary facts that he needs, in the form of a corpus of approximately one zillion running words, and he sees his job as that of deriving secondary facts from his primary facts. At the moment he is busy determining the relative frequencies of the eleven parts of speech as the first word of a sentence versus as the second word of a sentence. (There isn't anybody exactly like this, but there are some approximations). These two don't speak to each other very often, but when they do, the corpus linguist says to the armchair linguist, 'Why should I think that what you tell me is true?', and the armchair linguist says to the corpus linguist, 'Why should I think that what you tell me is interesting?'
>
> Fillmore (1992, p.35, cited in Svartvik, 2007, pp. 11,12)

There is another criticism of Chomsky against the use of corpus linguistics '95 per cent of the utterances in natural language are ungrammatical.' However, Chomsky's claim has turned out to be wrong by Labov (1969, p. 201 cited in McEnery and Wilson, 2001, p.16) who had experience upon spoken corpus data 'the great majority of utterances in all contexts are complete sentences'. The corpus is not a junk of ungrammatical constructions; therefore, the corpus-based enquiry may not be as invalid as it was originally supposed to be so.

There is also a different viewpoint on Early Corpus Linguistics different from Chomsky's, which is suggested by the concept of 'pseudo-procedure' by Abercrombie (1965, cited in McEnery and Wilson, 2001, p. 18). Abercrombie suggests processing data would be slow, expensive and prone to error– before the computerization of the data, i.e. the 1950s. Therefore, early corpus linguistics remains more expensive, more time consuming, less accurate and, basically, less feasible.

McEnery and Wilson (2001, p.19) sum up the controversial debates on the early corpus linguistics and Chomsky linguistics very briefly with the following sentences:

> The argument being made here is that, in abandoning the corpus-based approach, linguistics, if we can speak idiomatically, threw the baby out with the bath-water. The problems Chomsky rightly highlighted were believed to be fundamental to the corpus itself, rather than being fundamental to the approach taken to the corpus by the post-Bloomfieldian linguistics. In other words, if you think language is finite, then your interpretation of the findings in a corpus may reflect that – if we can change the interpretation of the findings in a corpus to match the verities Chomsky revealed, then the natural data provided by the corpus can be a rich and a powerful tool for the linguist. But we must understand what we are doing when we are looking in a corpus and building one.

The important point to either use or form a corpus is the purpose of the user and the richness of the corpus. All in all, both the artificial and natural data should be the yardstick in corpus linguistics by the mixture of introspective and observational procedures. After the Early Corpus Linguistics period, the second phase in the history of corpus linguistics started, namely, CL between the 1950s and the 1980s.

2.1.2. Corpus Linguistics between the 1950s- 1980s

The arguments above shed light on the historical reasons of moving from early corpus linguistics to a period, which is regarded by McEnery and Wilson (2001, p. 4) as an unpopular period, because it has been called "… unpopular not because of any terminology but because the second period underwent a very slow development and many linguists abandoned their studies or supports." Still, there were several linguists who continued their research and studies trying to find solutions. The basic solution has been processing the data by computer, which is touched upon the section 2.1.3.

This period covers the studies conducted roughly from the 1950s to the early 1980s. There is a common belief about the corpus-linguistics that the early corpus remained until the 1950s and then it was abandoned for about thirty years. In fact, it is not true, because several prominent scholars went on their studies on corpus-based methodologies by 1950s. Their work is divided into four margins: the work carried out by computing, multilingual linguistics respect to mechanolinguistics, work on English grammar and the last one by the neo- Firthians. Each of these works is touched upon in

Table 1 and the following subchapters.

Table 1

Corpus Linguistics from the 1950s to the 1980s by four subsections (Adapted from McEnery and Wilson, 2001, pp. 20-24)

	Computing CL	Machine Readable CL	Study on English Grammar	Neo-Firthians
Pioneer(s)	Roberto BUSA	Alphonse Juiland	Quirk, Svartvik, Francis and Kucera	Firth, Sinclair, Hallidays, Hoey
Field of Study	Computational corpus-based linguistics	Sampling techniques, Text processing on corpus with different genres	Use of English Grammar	Examination of complete texts and an open-ended corpus
Date of study/ (studies)	1949- 1967	1956	1960, 1980	About 1950s
Name of the corpus/ corpus project	The first study with IBM computers	Multilingual Machine-Readable Corpus	Survey of English Usage (SEU), International Corpus of English Project, Brown Corpus(BC), London-Oslo Bergen(LOB)	Cobuild Project, Bank of English(BoE)

As Table 1 illustrates, there have been four margins and the studies displayed have been a sort of primary corpus based studies conducted by the prominent scholars.

Those scholars have paved the way to the other corpus linguists to conduct their studies in their own field of interests. Below are the detailed explanations of each margin.

2.1.2.1. The First Computational Corpus-Based Linguistics

For a broadly corpus-based approach, Busa was the prominent scholar of computational corpus-based linguistics. Busa had a scholarly interest in the works of St. Thomas Aquinas, an Italian theologian and philosopher, who lived between 1225-1274. Busa had 10,000 sentences on cards with one sentence written on each. When Busa realised that it was very difficult to analyse them to find the word *in*, it was his interest at that time. Busa visited Thomas J. Watson Sr of IBM in New York and Watson helped Busa to do research by having processed cards into punch cards that those IBM computers guided by searching and retrieving from the word-by-word basis. Busa's project was in process from 1949 to 1967 developed 15,600,000 words related to medieaval philosophy and 5,000,000 words of the corpus were based on four languages; German, Russian, Aramaic and Nabatean (McEnery and Wilson, 2001, p.20).

2.1.2.2. The First Machine Readable Corpus

Mechano is described in Oxford English Dictionary (Version 2.0.3; 51.5 for Mac) as being mechanical, relating to a mechanical source, hence, mechanolinguistics is the branch of linguistics carried out mechanically or by the help of machine in corpus linguistics.

Juiland is the pioneer name in mechanolinguistics. Juiland's work represented in 1956 as a multilingual corpus is also as famous and inspiring as Busa's project which was finished in 1967. What is different from Busa's is that there were sampling techniques, different variety of text-based corpora on different genres. Moreover, in his study, Juliand used corpus annotation in the very early version and referred to different authors as for the different types of texts. The main contribution was developing the machine-readable corpus on those above -mentioned categories. Juiland had a very big goal of creating a machine-readable contrastive corpus of 500,000 words at that time. Juiland succeeded to some extent by including corpora of four languages as Romanian, Spanish, French, and Chinese in 1956. Juiland can be called the inspiring father of ICE and may be even the International Corpus of Learner English (ICLE), which was

thought about 35-45 years before those corpora came out with being multilingual or dispersed corpora. At that time, it was also West (1953, cited in McEnery and Wilson, 2001, p. 21) who was studying on computer-based corpus with the help of a number of human analysists. Juiland's computer-based corpus was a more specific to its own criteria. Busa and Juiland can be called the first engagers in the modern version of corpus (McEnery and Wilson, 2001, p. 21).

2.1.2.3. The First Grammar Based Corpus Linguistics

In 1960, it was Quirk who planned the Survey of English Usage (SEU); in addition, Francis and Kucera started their work on Brown Corpus (BC). They focused on the usages of English language on corpus-based studies. However, Jan Svartvik is a striking name on building the exact SEU and BC to transmit to London-Lund Corpus. In 1975, Svartvik succeeded in computerizing the SEU and in the 1990s, BNC became popular with its large index of spoken texts, which was a lack of the SEU. SEU had another goal in addition to the usage of English; it would be a grammar training reference for academicians on their analysis of English. Later, Leech developed it and started a very large scaled corpus project, which included even the most famous ones as Lancaster-Oslo-Bergen (LOB) and BNC. Quirk was also interested in SEU and in the mid-1980s, he founded International Corpus of English Project. All in all, SEU was very popular from the 1950s to 1980s which inspired several linguists to find new projects, in addition, it was the work of neo-Firthians which was similarly influential like SEU (McEnery and Wilson, 2001, p. 22).

2.1.2.4. The First Inspiring Works of Neo-Firthians

Neo-Firthians are known as the followers of John Ropert Firth. Among these are Sinclair, Hallidays, and Hoey. Firth called corpus an 'Attested language' (1957, cited in McEnery and Wilson, 2001, p. 23). As a result, Firth introduced the term 'collocation' to corpus linguistics and the linguists in Prague used this term between the 1930s and 1972 as "automation". Sinclair's study (1991) reveals how Firth has been so influential. The most exciting studies took place by the neo-Firthians – especially by Sinclair- are Cobuild Project conducted in Birmingham University and its associated project the Bank of English (BoE). Nowadays, according to Can (2010, p.18), today's largest

language corpora are BoE and BNC. The project of BoE started in 1991 and has reached the number of 524 millions of words, meanwhile, BNC serves with 100 million words collected between the years 1991-1994.

The work of neo-Firthians is different from those of Juiland's and SEU on the aspect of examination of complete texts and a kind of open-ended corpus, because the others are based on sampling and being epitomists for corpus (McEnery and Wilson, 2001, p. 23).

2.1.3. Corpus Linguistics from the 1980s- Present-day

Halliday (2007, pp. 314-315) states that "… the modern corpus is an extremely powerful theoretical instrument for linguistic research …" As this is so powerful, McEnery and Wilson (2001, p. 23), claims that the observation of Abercrombie was mostly suitable and was right at the time, however, it is not so now. At the modern time, the word corpus is synonymous with the term machine-readable corpus. The computer has put away all the pseudo-procedure effects by its great ability to search for, retrieve, sort, and calculate data in text form or digitized speech (after Leech 1991). Pseudo-procedures observation came up with three problems as being slow, expensive and prone to error, but now, (after the 1980s with the improvements of SEU and neo-Firthians) with the help of the computer, data is processed extremely fast and users can retrieve over millions of words with a mouse click and it is much more cheaper compared to the time when linguists had to hire an army of analysts to process data. It serves with almost total accuracy data according to the needs of linguists, teachers, and learners. By the assistance of concordance soft wares, therefore, the computer has invalidated the pseudo-procedure criticism.

Corpus linguists produced and developed their corpus based studies annually even between the years 1950s and 1980s. Now, there are many sophisticated corpora which are far beyond the expectations of the early corpus linguistics period, particularly, BC, BNC, BoE, ICE and ICLE. ICLE is different from other corpora in dispersion with being the leading of methodology for learners who can apply it for their own learning and it is a way, which addresses the learner as a tool of data-driven learning.

Chambers (2010), Gilquin and Granger (2010) as well as Meyer (2002) state that learner corpora help second language or foreign language learners to find the true

lexicogrammar functions for example via grammar mistakes (e.g. Longman's Learner Corpus which is a dictionary). At this point, not only learners can get benefit from learner corpora, but also teachers use them to discover what kind of difficulties learners are faced with. Hence, they can develop their teaching strategies by taking into consideration the learner corpora as a way of guidance into their own curriculum. DDL methodology is in a triangular crossroads of Learner Corpora, Computer Assisted Language Learning (CALL) and Corpus Methodology. Learner corpora on par, with data-driven-learning methodology serve for the same aim claiming that 'the students learn via learner corpus'. However, in DDL methodology, students must be taught to use concordancing programmes to use corpora, that is why computer-assisted language learning (CALL) is in the triangular cross-road. Students can also learn by worldwide web (www) using blogs, wikis, social networks, learning platforms in addition to corpora as a tool in learning environment; that's the way of DDL, which is applied, at a broader scale.

2.2. Types of Corpora

As the field of corpus linguistics has contributed a lot to the language research area, corpus has been categorised into several types serving in accordance with the needs of each research area. As corpus is a kind of text collection tool, it is wise to illustrate the types of text collection before touching upon the types of corpora. Atkins, Clear, and Ostler (1992, p. 1) refer four types of text collection; the first one is ***archive***, which is storage of electronic texts not linked in a principled way. The second one is ***electronic text library***, which is the repository of texts with generally principled formats serving for different contents but not serving with basic sections. The third type of collection of a text is ***corpus*** regarded as a subset of electronic text library, which is to be designed with explicit criteria having clear-cut constraints of sections within it for a specific purpose. The last one is ***subcorpus*** referred as sub category of a corpus or a component of a specific corpus. When these four types of text collection are placed in a hierarchical order for doing the cluster analysis; it can be briefly outlined that an archive is the leading term for electronic text library containing the data of a corpus, which can be divided into different subcorpora. In this study, corpus and subcorpus have been preferred as the selection of yardstick. Leech (1991, p. 11) states the significance of types of corpora since a corpus must have a particular representativeness function in its

design criteria no matter how large or small it is. Taking as its basis the compilation of corpora, corpus has been mainly classified into two registers; spoken and written registers and at least eight types of corpora can be utilised under these two registers. Table 2 illustrates those types in clear.

Table 2

Types of Corpora (Adapted from Baker, Hardie, and McEnery (2006, p. 49) and Hunston, (2002, pp. 14-16)

SPOKEN & WRITTEN REGISTERS	
Specialised Corpus	Reference Corpus
Multilingual Corpus	Parallel Corpus
Diachronic Corpus	Monitor Corpus
Pedagogic Corpus	Learner Corpus

As corpus is "a large and principled collection of natural texts" (Biber et al. 1998, p. 4), it can be collected through texts or oral data which is to be typed and the corpus can be based on either the written data or spoken data (or sometimes both). The data collected through any kind of written registers (books, journals, newspapers, articles, electronic texts, etc.) and transformed into corpus archive by being classified into each genres (i.e. principles) is written corpus. However, natural data can be observed by any oral sources such as conversations, dialogues, telephone conversations etc. Spoken corpus is consisted of above-mentioned data transcribed into written formats specifically being classified into different principles. Ghadessy, Henry & Roseberry (2001, pp. 178, 179) state that a corpus must have representativeness and it must have a relationship with the language genre it is used in.

Baker, Hardie, and McEnery (2006) have indicated seven corpora categories used in corpus linguistics. These corpora have been observed, taking into account the purpose of or the rationale behind compiling a corpus. In addition to these seven types of corpora Hunston (2002, pp. 14-16) added one more corpus type; pedagogic corpus.

According to these studies, corpora can be categorised as specialised, reference, multilingual, parallel, learner, diachronic, monitor and pedagogic.

2.2.1. Specialised Corpus

Cılız (2010, p. 16) defines specialised corpus compiled for use in a particular field of research and claims that specialised corpus has a purpose of providing authentic examples. For example, a medical English corpus can be attributed as a specialised corpus since it aims to exemplify medical English. These corpora have been regarded also as domain-specific or specific corpora. These corpora can be used to compare and contrast similarities and differences of usage patterns across different fields, for example; corpus of English for Academic Purposes (EAP) can be compared with English for Specific Purposes (ESP) or English for Medical Purposes (EMP).

To widen the context of specialised corpora, it can be suggested that a study on dialects is most probably based on spoken data and it is domain-specific to specialised corpus. Another example of domain-specific to specialised corpus can be raised that dictionaries are used as specialised corpora. More examples are given by Baker, Hardie and McEnery (2006) as Guangzhou Petroleum English Corpus (GPEC) – listing 411,612 words of petroleum English-, Jiao Tong University Corpus for English in Science and Technology – including roughly one million words of English serving with the genres of science and technology.

2.2.2. Reference Corpus

Reference corpus has a representativeness feature, which is essential to specialised corpus. However, it is a general corpus representing (a variety of) a language as a whole stated by Biber, et al (1998, p. 161). Hunston (2002, p. 15) states that reference corpora are also attributed as general corpora which can be based on both spoken or written registers, because these kinds of corpora can be utilized in language learning and translation studies in the norm of a native corpus which is much larger than specialised corpora. Reference corpus deals more with frequency-based corpus compared to specialised corpus. Baker, Hardie and McEnery (2006, p. 137) explain the need for a reference corpus in using frequency-based techniques to analyse a text or a set of texts to compare them by these words; “... some word or form is more common in

a particular text than is normally expected. The basis for the comparison is often a larger set of texts drawn from a wider range of genres and/or sources." Consequently, the above-mentioned larger dataset is often called a reference corpus. McEnery, Xiao, & Tonio, (2006, p. 310) regard British National Corpus (BNC) and The Freiburg - LOB Corpus of British English (FLOB) as reference corpora and prefer them as basis of selection in their study. As a reference corpus, the written component of BNC has been preferred as the basis of research by Bondi (2010) in her book edited with Scott (2010) "Keyness in Texts."

2.2.3. Multilingual Corpus

Hunston (2002, p.15) defines multilingual corpora as consisting of two or more corpora, for example, English and German, or different varieties of a language, for example, British English and Chicano English. They are designed within the same lines to be compared with the same proportion of design criteria. The varieties of a language can be compared to each other. For the comparison of different languages, translators or learners can compare them to identify differences or equivalents in the same language. ICE is a comparable corpus consisting of one million words of each of World Englishes (different varieties of English). If the corpus contains texts in one language, it is attributed as monolingual corpus. At that point, Lawson (2001, p. 282) defines multilingual corpus in more detail and specifies the difference between the two and the need for multilingual corpus. "Examining a monolingual corpus can obviously only provide information about that language. To compare two languages or a group of languages, a multilingual corpus must be consulted".

Baker, Hardie and McEnery (2006, p. 18) suggest Enabling Minority Language Engineering (EMILLE) corpus as an example of multilingual corpus. However, a parallel corpus is a specific type of multilingual corpus in which there is a defined relationship between the texts in different languages (usually, the texts are direct translations of one another). Lawson (2001) has confirmed the view that parallel corpus is a subcorpus of multilingual corpus by the following words: "There is a lesser, but steadily growing body of multilingual corpus material, of which the two main types are parallel and comparable corpora."

Aijmer (2008, p. 277) suggests two examples of multilingual corpora: Pedant Corpus (PC) including Swedish, English, German and French parallel texts. The main

language of PC is Swedish that provides the chance for the user to compare any combinations of Swedish-English, English-Swedish, Swedish-German and the like. The other corpus suggested is *Oslo Multilingual Corpus* consisting of English original texts and their translations into four languages: German, Norwegian, Dutch and Portuguese illustrated in Figure 1.

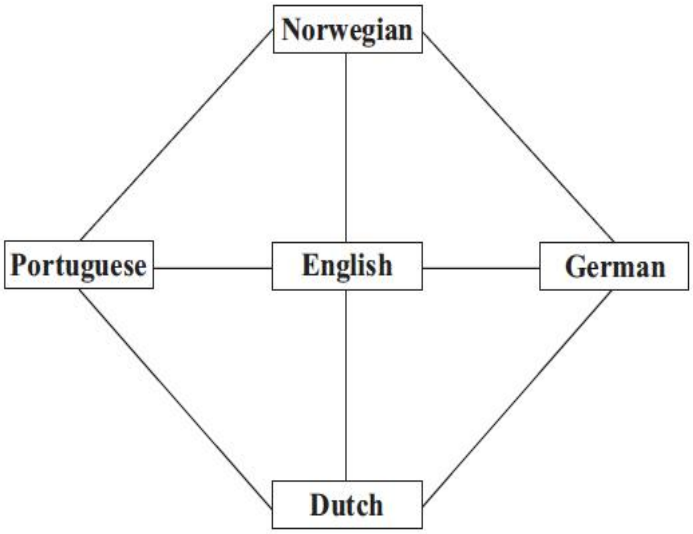


Figure 1. Multilingual comparisons of English with translations into other languages in the Oslo Multilingual Corpus (Aijmer, 2009, p. 277)

2.2.4. Parallel Corpus

As stated in Multilingual Corpus Section, parallel corpus can be considered as a subcorpus of multilingual corpus. It enables the translation opportunity of two or more different languages from each other. According to Dias, Madeira, Pereira Lopes (2005, p. 27) "in corpus linguistics the term is traditionally employed to refer to bi- or multilingual (sentence-wise) aligned text corpora with text tokens which are translations of one another.' The types of texts, which constitute parallel corpus, are usually technical manuals, government information leaflets, and political texts as they are generally translated.

It is labour-intensive for compilation of translations in databases. For this problem, Baobao, Danielsson, Teubert (2005, p. 132) point out a feasible solution, by

‘... recent advances, especially in parallel corpus research (Gale 1991; Tufis 2001; Wu 1994; cited in Baobao, Danielsson, Teubert, 2005, p.132), automatic identification of translation units and their target equivalents, from existing authentic translations by methods providing a list of bilingual translation units.’

2.2.5. Diachronic Corpus

Baker, Hardie and McEnery (2006, p. 56) define diachronic corpus that it is carefully built in order to be representative of a language or language variety over a particular period of time. Thus, it is possible for researchers to track linguistic changes within a particular period studying linguistic enquiry belonging to different periods of time.

Meyer (2002, p.46) states “with diachronic corpora (i.e. corpora used to study historical periods of English), the time-frame for texts is somewhat easier to determine, since the various historical periods of English are fairly well defined”. Although it seems easy, Meyer states that there would be some complications; Rissanen (1992, cited in Meyer 2002, p. 46) has remarked that Helsinki corpus has been divided into 100 year sub periods under the title of Old and Early Middle English sections of corpus. The time- frames determined in the Helsinki Corpus has shown that diachronic corpora not only cover the historical periods of English but also an evidence of how significant events took place during those places.

Although diachronic corpus seems easy to build or study on, Eckkrammer (2005, p. 20) mentions the importance of representativeness and comparability which a diachronic corpus seems to be more demanding in. It should have a certain representativeness that can be easily distinguished. For example, diachronic and synchronic corpora can be a matter of complexity. However, Cılız (2010, p. 18) compares the diachronic and synchronic corpora in historical linguistics aspect and the differences. Diachronic corpora are called historical corpora as diachronic corpora consists of the data spread to different particular periods of time that American National Corpus can be a good model of that type of corpora. However, synchronic corpora have been attributed as contemporary corpora mainly based on the language use at a certain period of time that Brown and Lancaster- Oslo-Bergen can be examples of this type of corpora.

2.2.6. Monitor Corpus

A monitor corpus is a corpus growing over time and it is dynamic, as opposed to a static corpus, which does not change in size once it has been built. According to Baker, Hardie and McEnery (2006, pp. 64-65), monitor corpora are useful in that they provide the means to monitor language change over time. That is to say it has two main features; the first one can be explained as monitor corpus is growing over time, and, it is so called as dynamic as opposed to static corpus. The second feature can be explained as this corpus has the feature of "monitoring language", it is sometimes so called. As stated previously in section 2. 2. 2. , Bank of English (BoE) can be attributed both as a reference corpus and a monitor corpus.

Meyer (2002, p.15) explains the reason lying behind why monitor corpus has occurred by defining it as "… a large corpus that is not static and fixed but that is constantly being updated to reflect the fact that new words and meanings are always being added to English." This type of corpus turns out to be a reliable one to study lexical items, hence, monitor corpus constitutes for 'being the philosophy of the Collins COBUILD Project at Birmingham University in England, which has led many dictionaries based on two monitor corpora: the Birmingham Corpus and BoE Corpus.'

Teubert (2005, p.103) states a different view on the use of monitor corpora besides traditional corpora out of printed texts " the Internet as virtual corpus becomes more important day by day; from it, domain-specific special corpora can be derived and continuously updated (in the sense of a monitor corpus) using thesaurus-based filtering techniques".

2.2.7. Pedagogic Corpus

A pedagogic corpus is made up of all the course books and reading materials that a learner has used in addition to the types that they have heard. Hunston (2002, p. 16) defines it as "all the language a learner has been exposed to." Hunston suggests the use of such corpora in corpus-based comparative studies to check the natural occurring forms of English that the learner has been presented in the materials with a reference corpus. In accordance with the purposes of this type of corpus, Reguzzoni (2008) has studied pedagogical corpus on investigating the materials used in an Italian Maritime

School aiming to identify the main lexical characteristics of the language, finding the 'right' items to focus on and devising well-targeted learning tasks.

2.2.8. Learner Corpus

The core of this study is based on "Learner Corpus", so far, it is mentioned above that there are seven other main types of corpora (specialised, reference, multilingual, parallel, monitor and pedagogic corpora). This type of corpus-Learner Corpus-has been a new type in the field and is different from the others on several aspects, but it can be adapted and studied in-line with others.

James (1992, p. 190) underlines the significance of learner corpora by uttering "the really authentic texts for foreign language learning are not those produced by native speakers for native speakers, but those produced by learners themselves" (cited in Baker, Hardie and McEnery, 2006, p. 103). Learner corpora provide the relationship (overuse, underuse, error analysis and so like) of any linguistic enquiry produced by learners and native speakers of that language for researchers.

There are some existing learner corpora, some of which are commercial or academic and based on one or more languages.

Table 3

The currently existing learner corpora. Adapted from Pravec (2002, pp. 82, 83) and Nesselhauf (2005, pp. 129,130)

Name of Corpus	Type of Corpus	Location of Corpus	Language Corpus
CLC	Commercial	England	Various
HKUST	Academic	University of Science & Technology, Hong Kong	Cantonese
ICLE	Academic	University of Louvain La-Neuve,Belgium	Various
JEFLL	Academic	Meikai University, Japan	Japanese
JPU	Academic	University of Pecs, Hungary	Hungarian

(Table 3. Continued)

LLC	Commercial	England	Various
MELD	Academic	Montlair State University, USA	Various
PELCRA	Academic	University of Lodz, Poland	Polish
TSLC	Academic	Hong Kong University, Hong Kong	Cantonese
USE	Academic	Uppsala University, Sweden	Swedish
CUP	Commercial	England	Various
SST	Academic	Meikai University, Japan	Japanese
LINDSEI	Academic	University of Louvain La-Neuve,Belgium	Various
FRIDA	Academic	University of Louvain La-Neuve,Belgium	French

ICLE: The corpus utilized in this study.

Table 3 illustrates the currently existing 14 learner corpora three of which are commercial. In this study, one of the other eleven academic learner corpora have been preferred, namely, International Corpus of Leaner English (ICLE). Granger (2009, p. 12) defines learner corpus (LC) studies being 'situated at the crossroads of four interrelated subjects: corpus linguistics, linguistic theory, second language acquisition, and foreign language teaching' (cited in Can, 2010, p. 16). "A LC, like International Corpus of Learner English (ICLE), contributes to these fields with the abundance of hands-on material enabling the researchers to examine various variables that could affect English interlanguage" (Can, 2010, p. 16). In this study, International Corpus of Learner English (ICLE), one of the biggest modern academic learner corpora, has been preferred as the selection of learner corpora. There are two biggest learner corpora to date that Nesselhauf (2005, p. 44) claims as the Hong Kong University of Science and Technology (HKUST) Learner Corpus with currently about 25 million words of Chinese learner of English, and the TeleNex Student Corpus with currently about 3

million words of the same learner variety. As these are the biggest learner corpora, there is a big gap in the number of words that other corpora have as containing between 600.000 and one million words. Nesselhauf (2005, p. 44) continues to state the only currently existing learner corpora that include more than one learner variety besides ICLE are two commercial corpora: the Longman Learner Corpus and the Cambridge University Press Learner Corpus.

Granger (2002, p. 4) stresses the significance of learner corpus research by creating an influential link between corpus linguistics and foreign/second language research. Corpus linguistics provides us main principles, tools and methods. We can use these in our corpus studies that are shaped by the main goal of corpus linguistics; to serve description of learner language in detail. By taking this goal as a basis, corpus linguistics can be applied for a great variety of purposes in foreign/second language acquisition as well as foreign language teaching by the material of learner corpus.

ICLE has been carried out for the purposes of triggering major developments in the field by taking into consideration the non- availability of learner corpora. By taking this purpose in mind, ICLE was assumed to have an impact as a new stage in the evolution of EFL research. This corpus has had two versions (ICLE*v1* & ICLE*v2*) up to now and was run first by Granger from Université catholique de Louvain, Belgium as the director of the project in 1990. This project was handled a team of 11 international universities and as a result of the study, the first version of this learner corpus- ICLEI*v1*- was released with a CD-ROM in 2002 (Granger, Dagneux, Meunier, F. eds. 2002) including written data supplied from 11 mother tongue backgrounds: Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Polish, Russian, Spanish and Swedish. When ICLE*v1* was first released, the researchers were able to "enjoy the first harvest in the form of an ICLE CD-ROM." (Tono, 2003, p. 800) ICLE has grown and ICLE*v2* has been made available to public research following Sinclair's (1991) guidance that "a corpus should be as large as possible, and should keep on growing." Data collection continued uninterruptedly after the release of the first CD-ROM – indeed is still continuing today, with several new national subcorpora still under construction (Granger, Dagneaux, Meunier and Paquot, 2009, p. 1). ICLE*v2* has written data supplied from five mother tongue backgrounds in addition to ICLE*v1*: Chinese, Japanese, Norwegian, Turkish and Tswana. Granger (1998) defines the L2 data supplied from the different mother tongue backgrounds as "total interlanguage". It can be used in modern day by the data provided from ICLE (p. 6). This definition is in line

with what Selinker (1972) suggested as an assumption and "… may be one of the few points agreed on by all SLA researchers; interlanguage - at least partially different from the native language and the target L2 – is a linguistic system in its own right." (cited in Lakshmanan and Selinker, 2001, p. 395) In other words, the English interlanguages taking place in ICLE project have their own system based on a basic research design at the process of compiling the learner corpus. The importance of the learner language has long been recognized, and in that state of affairs, there have been many studies by the material of the ICLE project that Bikelienė (2008, p. 30) supports the existence of the ICLE project which has facilitated the research into interlanguage offering reliable data based on advanced students' written essays.

2.3. From Contrastive Analysis to Contrastive Interlanguage Analysis

Fries (1945, p.9) states that "the most effective materials are those that are based upon a scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner" (cited in Powell, 1998, p.1). In Contrastive Analysis, the potential problems that a learner of L2 could face can be identified leading to a better understanding of interference derived from L1. Gass and Selinker (2008, p. 96) point out that "Contrastive analysis is a way of comparing languages in order to determine potential errors for the ultimate purpose of isolating what needs to be learned and what does not need to be learned in a second-language-learning situation." In this way, the differences or similarities have been analysed. The process of language learning, however, has been perceived to include the construction of IL, a 'transitional competence' illustrating the dynamic nature of the learner's developing system according to Powell (1998, p.4). Since the conducted CA studies have paved the way for SLA researchers to do their studies on CA, the limitations of CA have been realized. At that point, for example, Pery-Woodley (1990, p. 143, cited in cited in Granger, 1996b, p. 43) stresses the need for CA research in a wider context of "comparing/contrasting what non-native and native speakers of a language do in a comparable situation." Moreover, Selinker (1989, p. 285) stresses the significance of the impressive potential in this "new type of CA", whose object is to compare parallel native language and IL data (cited in Granger, 1996b, p. 43).

The new way of CA has emerged as a need for contrasting and comparing the use in different varieties of the same language under the terms of Contrastive

Interlanguage Analysis (CIA). The main need has been derived from the collected written data of L2 forming ICLE. That is to say, Granger has developed the CIA model. Granger (1996b) compares the CA and CIA in simple terms "Unlike classical CA, CIA does not establish comparisons between two different languages but between native and learner varieties of one and the same language" (p. 43) and represents CIA in Figure 2.

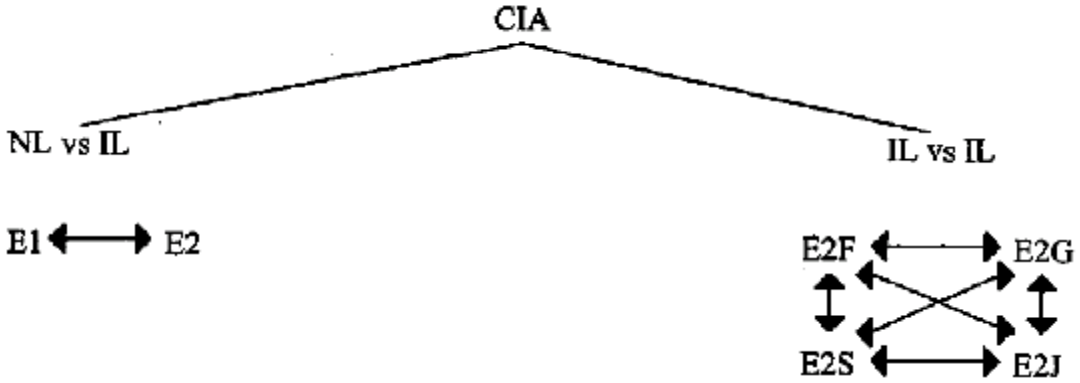


Figure 2. Contrastive Interlanguage Analysis (Granger, 1996b, p. 43)

CIA consists of two types of comparison; the first one is between Native Language (NL) and Interlanguage (IL). It displays the comparison of native and non-native varieties of one and the same language. In this figure, E1 has been contrasted English as first language and E2 as foreign language. The second type of comparison is shown between IL and IL; however, it includes the comparison of different interlanguages of the same language. In Figure 2, E2F refers to English of French learners, E2G stands for English of German learners, E2S for English of Swedish learners and E2J points English of Japanese learners. It can be applied with more interlanguages (Granger, 1996b, p. 44). Hence, ICLE has been preferred in CIA studies as the main material for many studies. In that sense, CIA also generated the need for one more thing to be applied in corpus based studies that Krzeszowski (1990, p. 206) stresses the importance of observing errors quantitatively and in order to determine " … such an error one has to perform a quantitative contrastive study of texts written by native users of a particular language and by a non-native user of the same language and compare the frequencies of use of the investigated forms" (cited in Granger, 1996b, p. 45). It triggered the need for a native / reference corpus which was LOCNESS. Therefore, it can be claimed that CIA has its roots derived from Selinker's (1972) suggestion that SLA studies should be based on IL, NL and Target Language (TL)

productions.

Tono (2004) reflected a multiple comparison of IL, TL, L1 and L2. Tono utilized ‘textbook corpus’ in his study and categorized it in TL. In our study, ‘LOCNESS’ has been called upon to provide an explanation for differences between IL and TL. Adapted from Tono’s proposal to our current study, the IL and TL relationship can be explained that “standard reference (e.g, the BNC, LOCNESS) and learner corpora have all roles to play in a fuller and proper exploration of learner language, a method which we may refer to as the “multimethod comparison” approach.” Figure 3 illustrates this point in a diagram.

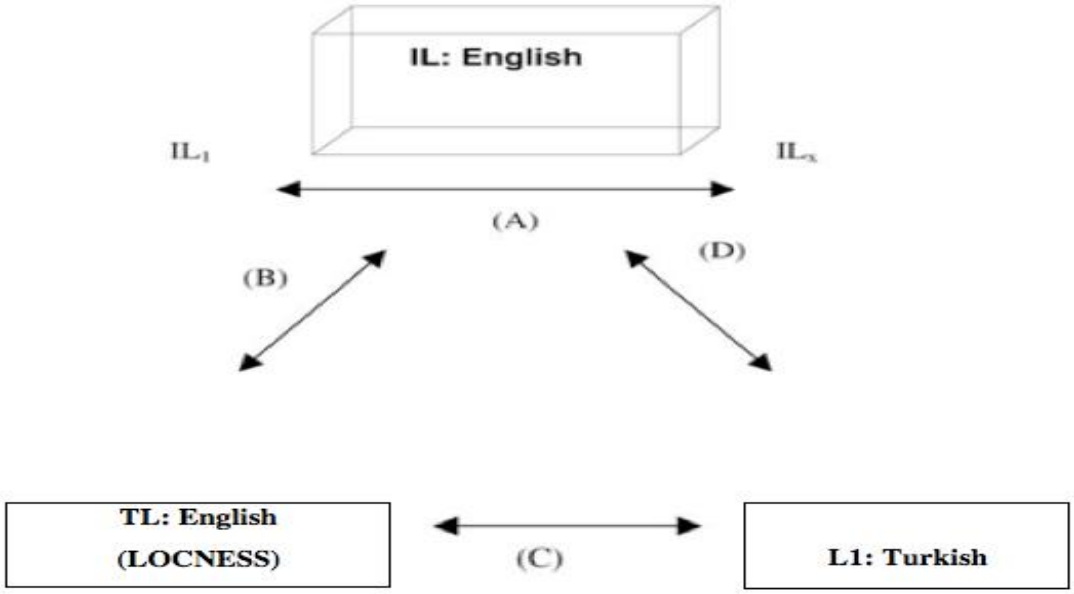


Figure 3. Multiple comparisons of L1, TL and IL corpora. (Tono, 2004, p. 53)

“IL1 ↔ ILx” in Figure 1 refers to “IL1: TICLE ↔ Ilx: JPICLE”. Tono (2004) suggests that these IL-IL comparisons can be of several different types, depending on the learner variables; however, as the learner variables are all same in both of the data of subcorpora, it strengthens the IL-IL comparisons of the present study. A comparison between L2 corpora and TL corpora can also be made. (see [B] in Figure 3) One can use either a general standard corpus such as the BNC to investigate differences in, or use a more comparable corpus of native-speaker texts, e.g., LOCNESS in ICLE, to compare like with like. Tono (2004) refers to this type of comparison as IL-TL comparison.

TL corpora are compared with L1 corpora (TL-L1 comparison, cf. I in Figure 3) in order to describe the target adult grammar system and identify potential causes of L1

transfer. This analysis should be combined with L2 corpus analysis. TL-L1 comparison could provide significant information on the influence of the source language on the acquisition of the target language.

A fourth type of comparison is between IL corpora and L1 mother tongue corpora (L1-IL comparison, cf. [D] in Figure 1). In the current study, to apply this procedure, TUC has been designed as L1 mother tongue corpus. Japanese L1 mother tongue corpus has not been utilized on the procedure; however, as Turkish and Japanese languages are so close to each other syntactically, it is possible to comment on the data in form in this study. If the researcher of the study were bilingual both in Turkish and Japanese, the current study would be semantically comparable as well.

Tono (2004) states that L1 corpora can provide information on features of the L2 learners' native language by which it is easy to understand potential sources of L1-related errors or overuse/underuse phenomena. L1-IL comparisons provide fundamental data in this area. Each comparison type is illustrated in Table 4:

Table 4

Multiple comparison approach (Tono, 2004, p. 54)

Comparison	Description
IL-IL comparison	Comparison between different stages if ILs or ILs by learners with different L1 backgrounds.
IL-TL comparison	Comparisons between learner corpora and target language corpora (i.e. ELT textbook corpora in the present study, or general native corpora).
TL-L1 comparison	Comparisons between target language corpora and L1 mother tongue corpora (to identify potential causes of L1 transfer).
L1-IL comparison	Comparisons between L1 corpora and learner corpora (to identify L1-related errors or overuse/underuse phenomena).
Il-L1-Tl comparison	Combination of the above comparisons (to identify the complex relationship between IL, L1 and TL corpora on L2 learners' error patterns or overuse/underuse phenomena).

2.4. Conditional Constructions

Dependent clauses consisting of three main structural clauses – relative clauses, adverbial clauses and complement clauses – are ‘a type of linguistic complexity’ in grammatical variation studies. Biber (2006, p. 72) states “… relative clauses are more common in written university registers when compared to adverbial and complement clauses that are more common in spoken university registers.” Figure 4 illustrates the overall distribution of the three structural clauses; relative, adverbial and complement clauses. Complement and adverbial clauses are considerably more common than relative clauses in spoken university registers (classroom teaching, class management, labs, office hours, study groups, service encounters) which require more oral communication than written university registers.

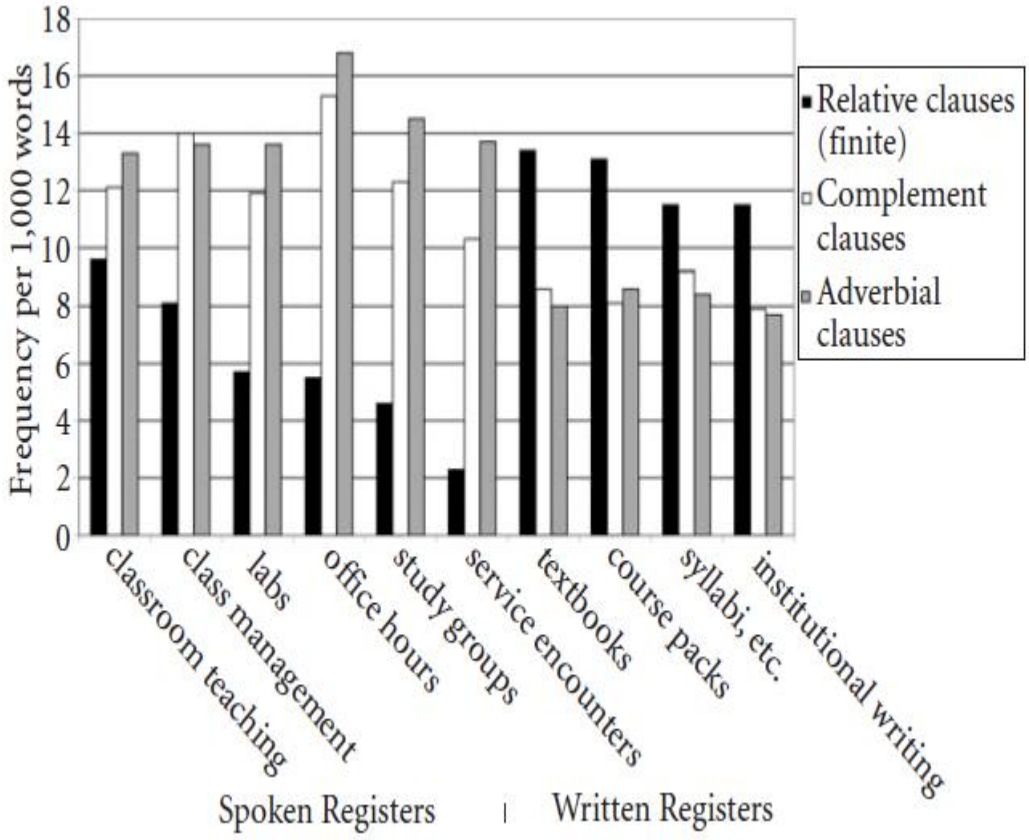


Figure 4. Dependent clause types across registers. (Biber, 2006, p. 73)

These three structural dependent clauses have their sub-categories as well. The major focus of the study is on the category of adverbial clauses. Adverbial clauses

consist of three major meaning domains; conditional (if), causative (because), and other clauses (e.g., after, before, while, until, as, since, so that) (Biber, 2006, p. 77). Conditional (if) include fourteen types, one of which is the subject of this research, namely, hypothetical conditionals.

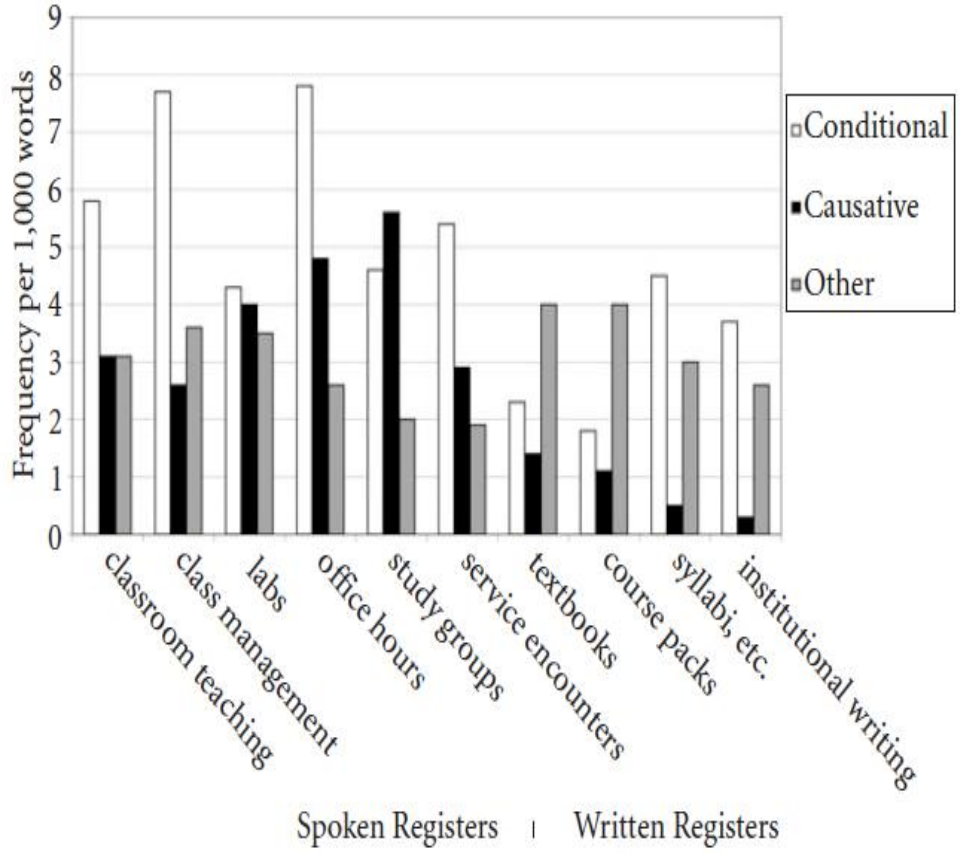


Figure 5. Breakdown of finite adverbial clause types across registers. (Biber, 2006, p. 77)

Figure 5 displays that conditional adverbial clauses has been preferred more in spoken university registers; however, in written university register, this type of clauses is the most common in syllabi and institutional writing. Quirk, Greenbaum, Leech & Svartvik (1985, p. 1089) state that many subordinators used for conditionals are ‘if’, ‘unless’, ‘as long as’, ‘so long as’, ‘assuming (that)’, ‘given (that)<formal>’, ‘in case’, ‘in the event that’, ‘just so (that)’, ‘<informal>’, ‘on condition (that)’, ‘provided (that)’, ‘providing (that)’, ‘supposing (that)’. The two simple subordinators of these are ‘if’ and ‘unless.’ Quirk et al. (1985) explain fourteen types of conditionals as illustrated in Table 5.

Table 5

Types of conditionals. (Adapted from Quirk et al., 1985)

Types of Clauses	Description and Example
Direct	The situation in the MC is directly contingent of on that of the conditional clause. e.g. If you put the baby down, it will scream.
Indirect	The condition is not related to the situation in the MC. e. g. She is far too considerate, if I may say so.
Open	This type of conditionals is neutral: they leave unresolved the question of the fulfilment or nonfulfillment of the condition, and hence also the truth of the proposition expressed by the MC. e. g. If Colin is in London, he is undoubtedly staying at the Hilton.
Hypothetical	This condition structure conveys the speaker's belief that the condition will not be fulfilled (for future conditions), is not fulfilled (for present conditions), or was not fulfilled (for past conditions), and hence the probable or certain falsify of the proposition expressed by the MC. e. g. If he changed his opinions, he would be a more likeable person. (He probably will not change his opinions.) They would be here with us if they had the time. (They presumably do not have the time.) If you had listened to me, you wouldn't have made so many mistakes. (You certainly did not listen to me.)
Rhetorical	These conditionals give the appearance of expressing an open condition, but they actually make a strong assertion. There are two types of rhetorical conditionals; a) If the proposition in the MC is patently absurd, the proposition in the conditional clause is shown to be false. e. g. If they're Irish, I'm the Pope. (Since I'm obviously not the Pope, they're certainly not Irish.) b) If the proposition in the conditional clause patently true, the proposition in the MC is shown to be true. The if- clause is positioned finally. e. g. He is ninety if he is a day. (If you will agree that he is at least a day old, perhaps you will take my word that he is ninety.)

(Table 5. Continued)

Conditional-Concessive	'Even if-' is used in the conditional forms of concessive; combining the concessive force of 'even' with the conditional force of 'if.' e. g. Even if you dislike ancient monuments, Warwick Castle is worth a visit.
The overlap between conditional and concession, already noted with 'even if', is particularly marked with in two types of adverbial clauses termed; Alternative & Universal Conditional-Concessive	
a) Alternative Conditional-Concessive	The correlative sequence whether … or (whether) is an alternative condition in that logically it combines the conditional meaning of 'if' with the disjunctive meaning of 'either … or', thus, it coordinates two subordinate clauses. e. g. Whether Martin pays for the broken vase or (whether) he replaces it with a new vase, I am not inviting him again.
b) Universal Conditional-Concessive	The alternative conditional-concessive clause gives a choice between two (or occasionally more) stated conditions, normally in sharp opposition, whereas the universal conditional-concessive clause indicates a free choice from any number of conditions. More often it is constructed by '–wh.' e. g. Whatever I said to them, I cannot keep them quiet. (Any number of choices) Wherever you live, you can keep a horse. (You can keep a horse at any place you may live)
Clauses of Contrast	Introduced by 'whereas', 'while', 'whilst (the less common one found in British English)', the contrastive meaning can be emphasized by correlative antithetic conjuncts such as 'in contrast' and 'by contrast' when the contrastive clause is initial. e. g. Mr. Larson teaches physics, while Mr. Corby teaches chemistry. I ignore them, whereas my husband is always worried about what they think of us.
Clauses of Exception	Introduced by several subordinators blending exception with condition: 'but', 'but that (formal)', 'only', 'except (that); less frequently excepting (that) and save (that) formal.' e. g. I would pay you now, except I do not have any money on me. Nothing would satisfy the child but that I place her on my laptop.

(Table 5. Continued	
Reason Clauses	There are many types of reason clauses including direct and indirect reason clauses. The related type is circumstantial clause, which combines reason with a condition that is assumed to be fulfilled or about to be fulfilled. e.g. Since the weather has improved, the game will be held as planned.
Adverbial/Adjunct Clauses of Comparison	They are also introduced as predication adjuncts and by 'as if', 'as though' and 'like.' If the comparison is factual, the verb in the comparison follows the normal rules for temporal reference. e. g. He looks as if he is getting better. If the comparison is hypothetical (implying lack of reality) a subjunctive or hypothetical past may be used as an alternative. e. g. She treats me as if I am/I was/ I were a stranger.
Subordinators with finite clauses	Of the subordinators that are conditional; 'if', 'unless', 'just so', 'as long as', 'provided that', 'in case', 'given that', 'assuming that', 'supposing that.' The clause can be finite. e. g. If you want some more, you should ask me.
Subordinate clauses with nonfinite clauses	Of the subordinators that are conditional, 'only if' and 'unless' represent non-finite clauses (mainly *–ed* participle clauses) and verbless clauses. e. g. The grass will grow more quickly if watered regularly. Unless otherwise instructed, you should leave by back exit. If wet, the pipe will not give you a good smoke. It has little taste unless hot.
Negative Condition	This type of clauses can be introduced by 'and', 'or', 'if not', 'unless' and can be used verbless with 'and' and 'or.' e. g. Give me some money, or (else) I will shoot. If you do not give me some money I will shoot. Unless you give me money, I will shoot. Make a move, and I will shoot! Do not make a move, or I will shoot.
Adverbial Conditional Clause	In categorizing adverbials, contingency has been preferred as a selection of category that includes cause, reason, purpose, result, condition and concession. For condition segment, it is linked to result in indicating the circumstances in which the result would be achieved. e. g. If he reads the book carefully, he will acquire some knowledge of metaphysics.

2.4.1. Hypothetical Conditionals

This study addresses the hypothetical conditionals consisting of a syntactically subordinate if-clause and a matrix clause. Taylor (1997, p. 301) defines conditionals in a three way dimension; factual, counterfactual and hypothetical conditionals. The first one refers to open conditions in Table 5 that the content of the if-clause is presumed to be the case and the second conditional, counterfactual, refers to the content of the if-clause to be contrary to fact. The last one which is "hypothetical conditional" stands between these two categories in which the content of the if-clause conveys possibility and no accordance with reality. Table 6 handles counterfactuals with hypothetical conditionals, in simple terms; hypothetical conditionals serve for type 2 and type 3 conditionals, counterfactual conditionals specifically refer to type 3 in accordance with hypothetical conditionals. Cheng (2002) states, " Both 'hypothetical' and 'counterfactual' conditionals share the characteristics in the content domain and the epistemic domain …"

Table 5 represents three examples of hypothetical conditionals, Quirk et al. (1985) claims that the speaker's belief is formulated on the time reference of the conditional clause. Hypothetical conditions are termed as closed, unreal, rejected, nonfactual, counterfactual and marked conditions. In the first example; for future reference (If he changed his opinions, he would be a more likeable person), the condition is contrary to expectation. In the second example illustrated in Table 5 for present reference (They would be here with us if they had the time), it is contrary to assumption. In the last example, for past reference (If you had listened to me, you would not have made so many mistakes), it is contrary to fact. Most subordinators are not used with hypothetical conditions, "If" is the main subordinator used in hypothetical conditions and it is important as a feature of hypothetical conditions that these constructions are back shifted (p, 1091, 1092). According to Dancygier (1998), and Ke (2004) back shifting occurs in form, but the meaning is referred as one tense back that the simple past tense is used to refer to the present, the past perfect tense is used to refer to the past which are illustrated by Quirk et al.(1985) as well (see Table 6). However, present simple tense is used to refer to the future. For example, 'If Jack takes his driving license, then he will buy a car.' rather than, 'If Jack will take his driving license, then he will buy a car.' In summary, Palmer (1990) termed back shifting with a past time reference

for unreal conditionals as “past-past” with a claim that explains why the past perfect tense is used in the past unreal conditionals “There is both past time and unreality, and the past tense, therefore, needs to be marked twice.” (p.170, cited in Ke, 2004, p. 9) Table 6 illustrates the verb forms with hypothetical conditions.

Table 6

The verb forms with hypothetical conditionals. (Quirk et al., 1985, p. 1092)

	Conditional Clause	Matrix Clause
Present and Future Reference	PAST If I were younger,	PAST MODAL I would study Classical Greek.
Past Reference	PAST PERFECTIVE If I had seen you,	PAST PERFECTIVE MODAL I would have invited you home.

Biber et al. (1999, p. 819, 828) state that when categorizing the semantic adverbials it is important to state that hypothetical conditionals are one of the three conditionals (open & rhetorical conditions), in addition to time, manner, place, contingency adverbials, conveying the meaning that the condition is not fulfilled. The difference in tense and modality is of high importance in hypothetical conditionals in so far that the hypothetical meaning is expressed in the subordinate clause by the past tense using ‘was’, ‘were’, ‘past perfective’, whereas in the main clause it is expressed by ‘would’ or some other modal auxiliaries such as; ‘might’, ‘could’ or ‘should’. Ziegeler (2000, p. 45) claims that ‘would’ conveys hypothetical meaning for future or past reference, for counterfactuals ‘would’ should be formed in ‘would + have + V-*en.*’ However, in Ziegeler’s (2000) diachronic corpus based study of ‘would’, it has been revealed by the light of corpus results that ‘would’ has hypothetical meaning in ‘if’ conditionals, but ‘would + have + V-*en*’ needs some contextual meaning which can be occupied by protasis in counterfactuals. Nonetheless, ‘would + have + V-*en*’ has been preferred to construct counterfactuals aiming to occur in a hypothetical division of conditionals. Dancygier (1998, cited in Kerslake, 2003, p. 216) states terminologically ‘would + V’ as weak hypothetical and ‘would + have +V-*en*’ type strong hypothetical

in conditional statements. Thus, it is observable that counterfactual statements in conditionals are also named terminologically as hypothetical conditionals.

2.4.1.1. Hypothetical Conditionals in Turkish

Kerslake (2003, p. 215) states that there are two main morphological markers used in conditional structures in Turkish. The first one is the verbal suffix *-sA*, which is used stressable in line with the normal morphophonological rules of the language. Another marker is the enclictic *ise*; in addition, one more marker is *-(y)sA* as a more common alternate form for enclictic *ise* . In other words, Kabak & Schiering (2004) rearticulate what Kerlsake (2003) points out for the use of conditionals in Turkish "… in three distinct markers: the conditional copula *ise*, its cliticized allomorph *=(y)sA* , and the conditional suffix *-sA* " (p. 233). Kornfilt (1997) states that verb is formally conditional when *–sA* is used on bare verb form and *-(y)sA* is used as a following tense/aspect/modality marker in a verb form in the grammar distribution of Turkish. As these morphological markers are used with verbs, their forms can change with reference to the subject pronoun that the verb is used. The forms of morphological markers can vary in terms of the referred use of subject pronouns. The variants of *–sA* used in regard to its subject pronouns have been illustrated at Table 7 below.

Table 7

The variants of –sA in Turkish conditionals. (Adapted from Banguoğlu, 1990 and Kabak & Schiering, 2004)

Subject Pronoun	Present Tense (- *A*r + se + person marker) Future Tense (-ecek + -s*A* + person marker) Past Tense (- *A*r + se + (y) di / -miş + person marker / -di (y) + -s*A* + person marker)						
I	- sem			- seydim	- saydım		
		-sam		- seymişim	-saymışım	- diysem	- dıysam
You	- sen			- seydin	- saydın		
		- san		-seymişsin	-saymışın		
She/	- se			- seydi	- saydı		
He/		- sa	- -	- seymiş	-saymış		

It/			seler	salar		
Passive						

(Table 7. Continued)

We	- sek		- seydik	- saydık		
		- sak	-seymişiz	-saymışız		
You	-		- seydiniz	- saydınız		
	seniz	- sanız	-seymişsiniz	-saymışsanız		
They	- seler		- selerdi	-salardı	-seydiler	-
		- salar	- seymişler	-saymışlar		saydılar

Table 7 illustrates the conditional suffixes used in Turkish. In this study, we have searched the concordance lists out of TUC by searching these suffixes. It is clearer in Appendix 4 that all these suffixes are formulated with the verb “-sev” as an an example.

Kornfilt (1997) states that there are two types of conditional adverbial constructions; the main type involves a nominalized subordinate clause illustrated in the following example:

Example 2.

(303) Hasan [[kitab-I san -a ver -dig -im] takdir -de] çok kız -acak
Hasan book-Acc. you-Dat. give -FNom -l.sg. case –locovery angry -Past
"Hasan will get very angry if (in case) I give you the book"
(p. 74)

Kornfilt (1997, p. 74) states another type of conditional is formed with the postclitic copular morpheme sequence -(y)-sA which means “if”. Like all postclitics, this morpheme undergoes Vowel Harmony, but it is exceptional with respect to word stress, which is placed on the syllable preceding the suffix; the suffix follows the tense suffix and precedes the subject agreement suffix:

Example 3.

(304) Hasan [kitab-I san -a ver -ir -**se**-m] çok kız -acak
Hasan book-Ace. you-Dat. give-Aor.-**if** -l.sg. very angry-Past
"Hasan will get very angry if I give you the book"
(Kornfilt, 1997, p. 74)

Göknel (2010) and Özsoy (1999) explain the uses of conditionals in two parts; 'if clause' and 'main clause.' Whether the event specified in the main clause will be realized or not is dependent on the condition specified in the conditional clause. In an 'if clause', the supposition is either real or unreal classified according to the referred time. Kerslake (2003, p. 215) explains the traditional approach towards conditionals in Turkish, as in English and other languages, focusing on the classic philosophical distinction between 'real' conditions and 'unreal' or 'hypothetical' conditions. The category of 'unreal' or 'hypothetical' conditions is subdivided into 'remote' and 'unfulfilled' (or 'impossible').

Real	*-(y)sA*
Remote	*-sA*
Unfulfilled	*-sA-yDI* (= *-sA* + past copula)

Kerslake (2003, p.215)

These unreal conditionals have been explained by Göknel (2010) as in the following.

In a *Present Future Unreal (Contrary to Fact) Supposition*, "V- (se, sa) – (pers)" verb structure is used in the conditional clause, and "used to" is used in the main clause. The present and unreal conditional structure expresses that the condition is not likely to be fulfilled and that consequently the action of the main clause is not to be realized. In present and future irrealis structures, the verb of the conditional clause is marked with the conditional suffix *–sA* and the person marker that agrees with the subject of the clause; the verb of the main clause is marked with the aorist suffix *–Ar* (V- ir, ır, ür, ur, er, ar) – and the person suffix. In these structures, the conjunction *eğer* is not used.

Example 4.

- Yağmur yağ-**sa,** havadaki bu sıkıntı kaybol-**ur.**

If it **would** rain, **this heaviness** in the air **would** lift. (But it won't rain.)

In those cases in which the predicate of the conditional clause is a noun or an adjective, the auxiliary *ol-* is used as the verbal element.

Example 5.

- Zengin ol-**sa-m,** uzun bir tatil yap**ardım**.

If I **were** rich, **I would** take a long vacation. (But I am not rich, therefore I will not take a vacation.)

Example 6.

- Evde ol-**sa** kapıyı aç-**ar-dı.**

If he **were** at home, **he would** open the door. (He is not at home now.)

Example 7.

- Sizin yerinizde **olsam**, başka bir şekilde hareket **ederim**.

If I **were** in your place, **I would** act differently. (Underhill, 1976, p.414)

The verbal element *–ol* is used in the examples extracted to above show the formulation is in the present tense. However, if it was formulated with the past copula –yd*I*, in other words; *-olsaYdI*, the meaning and the formulation would compeletely. In these situtations where the verbal element shows a past meaning, it has been categorised under the past hypothetical conditional and Type 3 if-conditional statement. Hengirmen (1998) stresses the importance of the first verbal element in hypothetical conditionals as the determining the tense of the sentence. The example from TUC has been extracted to below to show a Type 2 if-conditional hypothetical statement for present meaning.

Example 8.

Eşitliği yaşarsak anca mezarda yaşarız.

If we experienced the equality, we could just get it under the sod.

(Retrieved from TUC-KSU-106)

According to the participant from TUC data, the example above explains a hypothetical situtation which is referring to something that could not happen any time. In the example below extracted from TUC again illustrates a clear example of a Type 3 if-conditional hypothetical statement.

Example 9.

Uçmayı düşlemeseydi Hezarfen Çelebi, uçakları bulabilir miydi,
havayolunu kullanabilir miydi insan?
If Hezarfen Çelebi had not dreamt of flying, would the human being have innovated planes, used airlines?

(Retrieved from TUC-CU-012)

The first verbal element "düşlemeseydi" shows a past copula *–eYdI* which determines the second and main verbal element of the sentence that the sentence is composed imaginatively talking about a real situation, but an unreal supposition. That is to say, this sentence is categorised as Type 3 if-conditional hypothetical statement. The present unreal suppositions can also be used to express future regret.

Example 10.

Yarın tatil ol-**sa** pikniğe gider-**di-k.**
If tomorrow **were** a holiday, **we would** go for a picnic.

To form an *Past Unreal Supposition*, "V- (se, sa)/ y/- (di, dı) – (pers)" verb structure is used in the if clause, and the "used to" is used in the main clause. The past unreal conditional structure expresses the fact that since the condition was not fulfilled, the event in the main clause did not happen. The verb of the conditional clause is marked with the conditional suffix and the past clitic *–sA+ydI* and the necessary person marker. The verb of the main clause is marked with the aorist *–Ar* or the future *–(y)AcAK*, the past clitic *–DI* and the necessary person marker.

Example 11.

(**Eğer** sen) Dün akşam toplantıya gel-**se-ydi-n,** muhakkak ki çok **sıkıl-ır-dı-n.**
If **you had come** to the meeting last night, you **would have** definitely **been** very **bored**.

In the structure given above, the sequence *–Ar-dI* expresses an unrealized possibility.

Example 12.

Dün onu gör-se/y/-di-i̲m, onunla konuş-**ur-du-u̲m.**

If I had seen him yesterday, I **would have talked** to him.

(I didn't see him, so I didn't talk to him.)

Example 13.

O araba o kadar pahalı ol-ma-sa**/y/-dı**, muhtemelen satın al-a-**bil-ir-di-ı̲m**.

If that car **hadn't been** so expensive, probably, I **could have bought** it.

In the structure given above, the sequence *-bil-ir-di* expresses a high probability which can be translated by using "could."

Example 14.

Çabucak dur-ma-**sa/y/-dı-ı̲m**, adam fena halde yaralan-a**-bil-ir-di.**

If I hadn't stopped suddenly, the man **might have been** badly **injured**.

In the structure given above, the sequence *-bil-ir-di* expresses a probability which can be translated by using "might."

Dancygier's typology (1998, cited in Kerslake, 2003, p. 216) has indicated that the most recent development of the standard treatment of conditionals having a basis on the monograph of Sweetser (1990, cited in Kerslake, 2003, p.216) into a main bipartite division; "predictive" and "non-predictive conditionals." In this study, the essential concern is on predictive conditionals as they are interrelated with the content of the clauses, expressing a casual link between two events or states as in the example shown below.

Example 15.

a) If it rains, the match will be cancelled.
b) If it rained, the match would be cancelled.
c) If it had rained, the match would have been cancelled.

Adapted from Kerslake (2003, p. 216)

Kerslake (2003, p. 216) has stated the significance of the three sentences uttered above in the following comments. The occurrence of the event in the matrix clause is predictable from the occurrence of the event in the conditional clause. The difference between (a), on the one hand, and (b) and (c), on the other is that in (a) the speaker's epistemic stance towards the fulfilment of the condition is neutral, whereas in (b) and (c) the speaker has adopted an evidentially 'distanced' stance towards the fulfilment of the condition, presenting it as unlikely in (b) and impossible in (c). Syntactically, the regular sequencing of the verb forms in predicative conditionals is stated in *p* and *q*. Semantically, they are also stated under the term of 'backshift': the verb forms referring to a time prior to that they indicate in simple sentences. Back-shifting is sometimes attributed as "modal distancing", "expression of unreality", and "distance from present reality" by other scholars (Salsbury, 2000 cited in Ke, 2004). This can be observed in the protasis of (a), where 'present' tense refers to a future time, and in both clauses of (b) and (c), where past and ultra-past marking respectively, on the verb in the protasis and the modal in the apodasis, are uses to encode hypothetical future and past situations.

2.4.1.2 Hypothetical Conditionals in Japanese

Allen et al. (2007) terms Japanese and Turkish as verb –framed languages, while English as satellite-framed. The reason for terming these three languages in this way is because Turkish and Japanese have an agglutinative syntactic system. Kuno (1978) defines Japanese in the category of verb-final languages and claims among the various word order patterns; the subject-initial sentence pattern is the most common one. Kuno (1978, p. 58) demonstrates the proof for the SOV word order by uttering "in a large-scale sentence-pattern count of modern Japanese journalistic writings it was found that sentences with SOV word order occur seventeen times more frequently than sentences with OSV order". Lehmann (1973), having examined the morphological structure of languages that have objects before verbs, namely OV languages, has found that many of these languages are agglutinative. Turkish and Japanese take place as one of the examples of consistent OV languages besides Quechua and Sanketi (a dialect of Tamil). Conversely, in the morphology of consistent VO languages English is a SVO language (Diessel, 2001, Lehmann, 1973, Kurzon, 2008). That is to say, syntactically Turkish and Japanese have the same word order system as "SOV". In hypothetical conditionals of Japanese language, the use has been explained by Kaiser et al. (2001) as in below:

-ba is used for conditionals in Japanese. In the example below, it is illustrated in the main clause that the past/perfect ending **-ta** (often in the form **-te ita**) is used often with an expression of conjecture or guessing. The conditional combination indicates a hypothetical (i.e. unrealized) condition *if ... would have*. Also, **-tara** can be used in the same sense.

Example 16.

シートベルトをしていれ【ば】、助かった。

shito beruto o shite ire**ba** tasukat**ta**

If he had been wearing a seatbelt, he would have lived.

Kaiser et al. (2001, p.32)

When *-tara* is translated as 'when', it can basically be replaced by -ba. However, this is not always possible due to the restrictions on controllability of the verb in the case of –ba. In -tara sentences, main clause can freely indicate the speaker's intention (requests, hortative, etc.). Also, when the main clause indicates a past event, it can express a hypothetical condition. *–tara* has a colloquial ring, and therefore in expository and academic prose, etc. -ba is used instead. After *-ba ii*, it indicates a hypothetical condition *would be good if*. Here, *ga* can also be used.

Example 17.

一緒に過ごすボーイフレンドがいればいいんです【けど】.

issho ni sugosu bOifurendo ga ire**ba ii** n dt.•su kt-do

It would be nice to have a boyfriend to spend (Christmas) with.

Kaiser et al. (2001, p.210)

-(y)ö mon(o)-nara indicate a hypothetical condition *if you were to do anything like*.

Example 18.

悠長に酒など飲んでい【ようものなら】、まったく取材にならない。

yiicho ni sake nado nonde i*yo mono nara* mattakn shnzai ni naranai

[Foreign correspondent:] If you were to hang around idly having drinks

etc., you would not be able to collect any news.

Kaiser et al. (2001, p. 252)

Example 19.

うかつに手を出そうもの【なら】、大けがしかねない。

ukatsu ni te o das**o mono nara**, okega shi-kanenai

If you were to dabble [in the stock market], you might get your hands burned badly.

Kaiser et al. (2001, p. 274)

Usually, the main clause (if it takes place) precedes the conditional clause with regard to the time. The example below and above illustrates that Japanese employs *could, might* in form and meaning which ensures the comparability of the enquiry with the other languages in the study (English, Turkish).

Example 20.

三つの願いがかなう【なら】、何を望むか。

mittsu no negai **ga** kanau **nara** nani o nozomu ka

If you could have three wishes fulfilled, what would you ask for?

Kaiser et al. (2001, p. 271)

Preceding the predicative conditionals, *-tara* and *–ba* can be used within real and unreal conditionals. In such sentences, *-tara* and *–ba* can be replaced by each other. The example given below is counterfactual to what is said at the time of speaking. *-tara* and *–ba* have the same function in these statements.

Example 21.

お金があったら / あれば、買えるのに。

o-kane-ga at-tara / are-ba ka-e-ru noni

Param olsaydı, (satın) alabilirdim.

If I had had money, I could have (bought).

Zülkadiroğlu (2009, p. 92)

In the example extracted below, ***-tara*** and ***–ba*** has the same semantic contribution to the sentence, which is counter to the fact.

Example 22.

もっと早く来たら / 来れば、間に合った。

motto hayaku ki-tara / kure-ba maniat-ta

Daha erken gelseydik, yetişirdik.

If we had come earlier, we would have arrived.

Zülkadiroğlu (2009, pp. 93-94)

CHAPTER III

METHODOLOGY

3.1. Research Design

This particular study has been carried out as a mixed type of descriptive quantitative research design to depict the misuse, overuse and underuse of hypothetical conditionals by the materials; LOCNESS, Japanese & Turkish subcorpora of ICLE and TUC. The hypothetical conditionals have been examined in two types of full conditional statements; Type 2 and Type 3. The use of "would", "might" and "could" have been preferred to be analysed at the category of Type 2 conditional statements. The use of "would + have + Ven", "could + have + Ven" and "might + have + Ven" have been preferred to be investigated at the category of Type 3 conditional statements. There are two methodologies having been developed so as to uncover those above uttered differences across the four materials; Contrastive Interlanguage Analysis (CIA) and Computer-aided Error Analysis (CEA). As English Interlanguage varieties show major differences within each other, which can be compared and contrasted with L1 varieties in frequencies of some language patterns overused or underused, CIA has been preferred to depict the use quantitative in character. (Granger et al., 2009, p.40)

Mainly, CIA has been used in this study based on Data Driven Learning Approach to cover similarities and differences on the frequency patterns. Error Analysis (EA) has been used to depict the errors quantitative in character made by Japanese and Turkish EFL learners as well. On one hand, comparisons of L1-L2 reveal the features of non-nativeness of learner writing, which is a matter of questioning of over- and underuse of linguistic items or structures as a question of downright errors at an advanced level writing. On the other hand, L2-L2 comparisons pave the way to assess whether these features are due to the "… influence of the learner's mother tongue (L1 dependent) or common point for all learners, whatever their mother tongue background, and are therefore likely to be developmental rather than interlingual or be due to other causes, such as teaching methods." (Granger et al., 2009, pp. 40, 41)

3.2. Materials

There have been four materials, namely, JPICLE, TICLE, LOCNESS and TUC, utilized in the current study. The main part of material has been taken from ICLE; JPICLE and TICLE. As pointed out in chapter three, ICLE has two version; ICLE and ICLE*v2*. In this study, ICLE*v2*, which is up to date, has been preferred. The learner corpora; JPICLE and TICLE, have in-line comparable criteria with the other two corpora; LOCNESS and TUC, though, these two learner corpora have not been compiled exactly on the same basis of design criteria. The design criteria of ICLE has been explained in detail below, compared and contrasted with that of LOCNESS and TUC.

3.2.1. ICLE*v2* Design Criteria

Design criteria are very important in the case of learner data because there is so much variation in the field of EFL/ESL. A random collection of heterogeneous learner data does not qualify as a learner corpus (Granger, 2002, p. 9). The requirements set at the beginning of the ICLE project were the following:

Learners: young adults (university undergraduates); advanced proficiency level; learners of EFL rather than as an ESL.

Language: Academic writing (mainly argumentative); 200,000 words per subcorpus.

In Table 8 other variables have been depicted which can be used as search criteria since these variables are included in the ICLE*v2* database. Granger et al. (2009, p. 3) state that “One variable, which undoubtedly plays a crucial role but which we have not been able to record, is the teaching methodology and pedagogical materials to which learners have been exposed”.

Table 8

ICLE Task and Learner Variables. (Granger et al., 2009, p. 4)

International Corpus of Learner English	
Task Variables	Learner Variables
Medium	Age
Genre	Gender
Field	Mother tongue
Length	Region
Topic	Other FLs
Task Setting	Stay in English-speaking country
	Learning context
	Proficiency level

As Table 8 illustrates, there are two main variables; Task and Learner variables. Task variables have been formed by six sub-variables; medium, genre, field, length, topic and task setting. Learner variables have been made up of eight sub-variables; age, gender, mother tongue, region, other foreign languages, stay in English-speaking country, learning context and proficiency level.

3.2.1.1. Task Variables

In the ICLE project, task variables were determined by medium as writing, genre as academic essay, field as general English rather than ESP and length as between 500 and 1,000 words. Topic and task setting have been determined by specific guidelines. All those task variables have been processed on the database of ICLE software that the users can easily retrieve and concord the enquiry for which study and following are the detailed overall analysis of several task variables (Granger et al., 2009, p. 5).

3.2.1.2. Genre

91% of the essays are academic (mainly argumentative) and 9% literary texts. Table 9 illustrates those texts among the 16 learner subcorpora.

Table 9

Proportion of argumentative essays. (Granger et al., 2009, p. 5)

National Corpus (Subcorpora of ICLE)	Argumentative
Bulgarian	100%
Chinese	100%
Czech	81%
Dutch	96%
Finnish	92%
French	85%
German	97%
Italian	34[1]%
Japanese	100%
Norwegian	98%
Polish	99%
Russian	100%
Spanish	79%
Swedish	85%
Turkish	100%
Tswana	100%
ICLE*v2*	91%

As Table 9 illustrates the overall proportion of the text types; it is clear that six national learner subcorpora have been constituted of only argumentative essays. In this current study, the preferred learner corpora, namely Japanese and Turkish, are compiled

of argumentative essays. The other 12 national learner corpora have been formed of both argumentative and literary texts, but with a very small percentage of literary texts.

3.2.1.3. Length

The contributors have been asked to write at least 500 words in their selected choice of topics. Most of them are argumentative essays and Table 10 illustrates the average essay length in ICLE.

Table 10

Average essay length in ICLE. (*Granger et al., 2009, p. 6)*

National Corpus (Subcorpora of ICLE)	Average Length- *per word*
Bulgarian	663
Chinese	500
Czech	830
Dutch	893
Finnish	704
French	654
German	526
Italian	572
Japanese	542
Norwegian	668
Polish	641
Russian	832
Spanish	789
Swedish	564
Turkish	713
Tswana	384
ICLE*v2*	617 words

All the essays are original collected texts and have an average length of 617 words, though differences have been observed between the subcorpora – from an

average of 384 words in the Tswana subcorpus to 893 words in the Dutch corpus. Table 10 displays the average essay length for each subcorpus.

3.2.1.4. Topic

There is a wide range of topics represented in the ICLE essays. Table 11 illustrates the list of the ten most popular topics and lists the subcorpora that contain the highest proportion among them.

Table 11

Top ten essay topics in ICLE. (Granger et al., 2009, p. 7)

Essay Topic	Number of essays	Country of origin
Some people say that in our modern world, dominated by science, technology and industrialization, there is no longer a place for dreaming and imagination. What is your opinion?	491	29% Bulgarian
Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.	249	22% Turkish
Poverty is the cause of the HIV/AIDS epidemic in Africa.	243	100% Tswana
Marx once said that religion was the opium of the masses. If he was alive at the end of the 20th century, he would replace religion with television.	237	19% Russian
The prison system is outdated. No civilized country should punish its criminals: it should rehabilitee them.	176	32% Tswana
Discuss the advantages and disadvantages of banning smoking in restaurants.	156	100% Chinese
Discuss the advantages and disadvantages of banning using credit cards.	149	100% Chinese
Feminists have done more harm to the cause of women than good.	139	23% Russian

(Table 11. Continued)

In the words of the old song “Money is the root of all evil”.	133	22% Russian
In his novel “Animal Farm”, George Orwell wrote “All men are equal: but some are more than others”. How true is this today?	127	39% Bulgarian

As Table 11 displays, the topics represented in the ICLE design criteria are so influential and inspiring that most of them have been the issue of debates over decades, that is why so many participants of the ICLE project have written more then the required length of words (see Table 10).

3.2.1.5. Task Setting

Three task conditions have been preferred: whether the task was timed or untimed, whether it was part of an exam or not, and whether students were allowed to use reference tools to complete the task. Untimed essay is usually written at home and timed one is usually practiced under exam conditions. Following is Table 12 illustrating the rate of each of task conditions.

Table 12

Proportion of task conditions in ICLE. (Granger et al., 2009, p. 7)

Timed	Untimed	Written under exam conditions	Not written under exam conditions	With the support of reference tools	Without the support of reference tools
38%	62%	39%	61%	48%	52%

As Table 12 shows an evidence of the atmosphere in which the data obtained it seems easy for participants to write down their essays. Task setting is the last sub variable of Task Variables; apart from this, it is wise to repeat that we do not know the context of participants’ learning English as a Second Language. As a solution about the

foreign language background of the participants, each of the national subcorpus coordinators have reflected the education system of their country and may have depicted the ELT backgrounds of the learners in the ICLEv2 handbook (Granger et al., 2009).

3.2.1.6. Learner Variables

There are eight clear-cut variables applied on the data collection procedure of ICLE project: age, gender, mother tongue background, region, knowledge of other foreign languages and time spent in an English- speaking country, learning context and proficiency level. In the following Table 13 learners' age and gender are illustrated and the category national subcorpus serves for different mother tongue backgrounds.

Table 13

Learners' Age and Gender Distribution in ICLE. (Adapted from Granger et al., 2009, pp. 8, 9)

National Corpus (Subcorpora of ICLE)	Average Age	Learners' Gender	
		Percentage Female	Percentage Male
Bulgarian	20,55	83%	17%
Chinese	20,49	64%	36%
Czech	22,07	72%	28%
Dutch	20,75	73%	27%
Finnish	22,73	85%	15%
French	21,70	88%	12%
German	23,39	78%	22%
Italian	24,59	92%	8%
Japanese	20,06	73%	27%
Norwegian	23,94	74%	26%
Polish	23,39	80%	20%

(Table 13. Continued)

Russian	21,19	84%	16%
Spanish	21,72	86%	14%
Swedish	27,73	77%	23%
Turkish	22,08	81%	19%
Tswana	22,47	60%	40%
ICLE*v2*	22,30	76%	24%

In order to have a precise picture of the learners' language background the ICLE project team has recorded which other languages they speak at home. Granger, Dagneaux, Meunier and Paquot (2009, p. 10) represent a debate on the variable "learning context" whether the participants learnt English as a SL or FL, however, conclude by the following words " . . . . what is certain, is that all the participants in the ICLE corpus have learnt English primarily in a classroom setting."

All those mentioned variables could be found in their categories within the database of ICLE. Figures 6,7 and 8 display those categories with the screenshots extracted from the ICLE*v2* software. In using the ICLE*v2* software, there are three steps to reach the data. Figures 6, 7 and 8 illustrate the steps in an order. The user can reach the data and the results even not selecting any item shown in Figure 6. Figure 6 is titled as "Corpus Selection 1." At this step the user can narrow the search according to a specific file name. It makes the user to find an interlanguage production of a specific participant which has its own code like; ICLE-TR-CUK-001 belongs to the first Turkish participant from University of Çukurova. The length of words can be selected between a range specified by the user. The user can also determine age range by taping from a minimum to a maximum age of the informants.

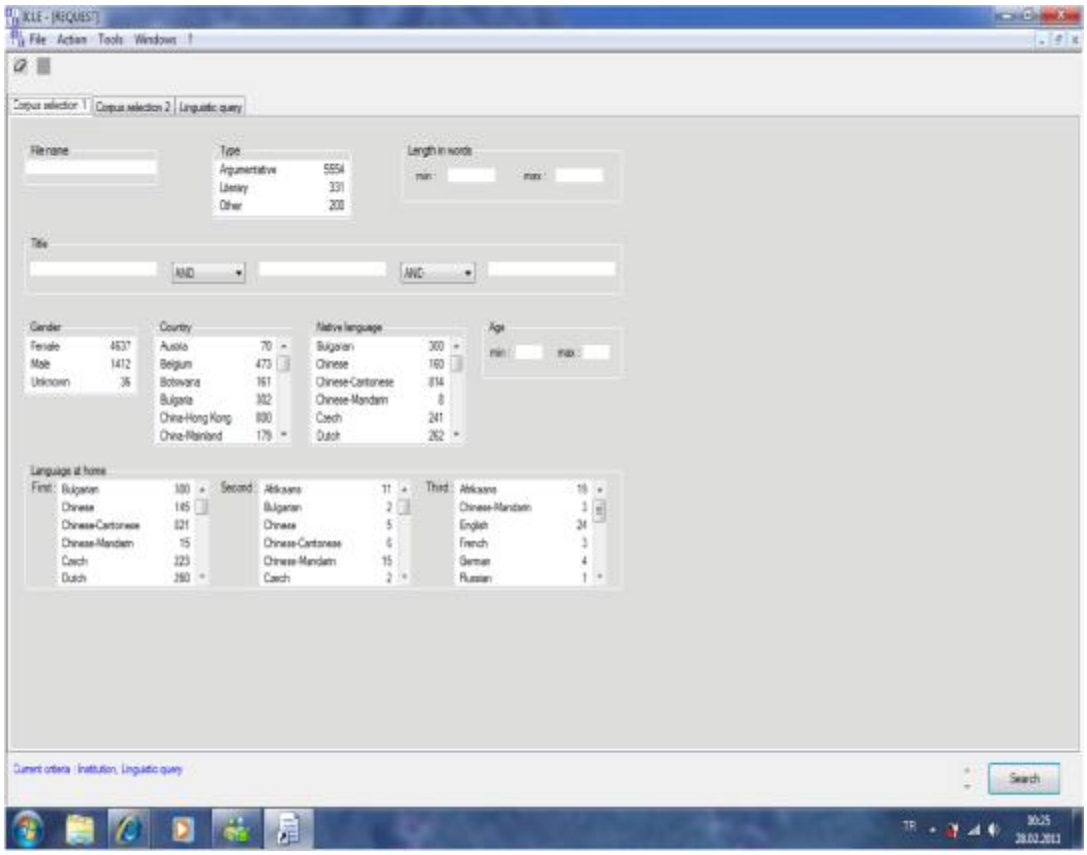


Figure 6. A screenshot of task and learner variables extracted from ICLE*v2.*

In Figure 6, the total number of each variable in this selection can be found under the titles of "Type, Gender, Country, Native Langugue, Language at Home; First, Second and Third " as well. That is to say, the user can easily compare the variables in accordance with his purpose before searching the ICLE*v2*. This step pave the way in a clear cut search for the user to switch to the second step as illustrated in Figure 7.

As the following figure shows, Figure 7 is titled as "Corpus Selection 2." In this step of determining the variable to be chosen, the user is face to ten options. Three of the options is illustrated under the title of "Other Foreign Language; First, Second and Third." In this category and the categories of "Condition; No Timed, Timed and Unknown", "Reference Tools; No, Yes, Unknown" and "Examination; No, Yes, Unknown", the total number of each item can be specified in accordance with the selected institutions. At the very down of this page, the Institutions and the number of the essays contributed of each institution can be selected. In this study, we have selected the University of Çukurova and Mersin University which are shown in blue in Figure 7.

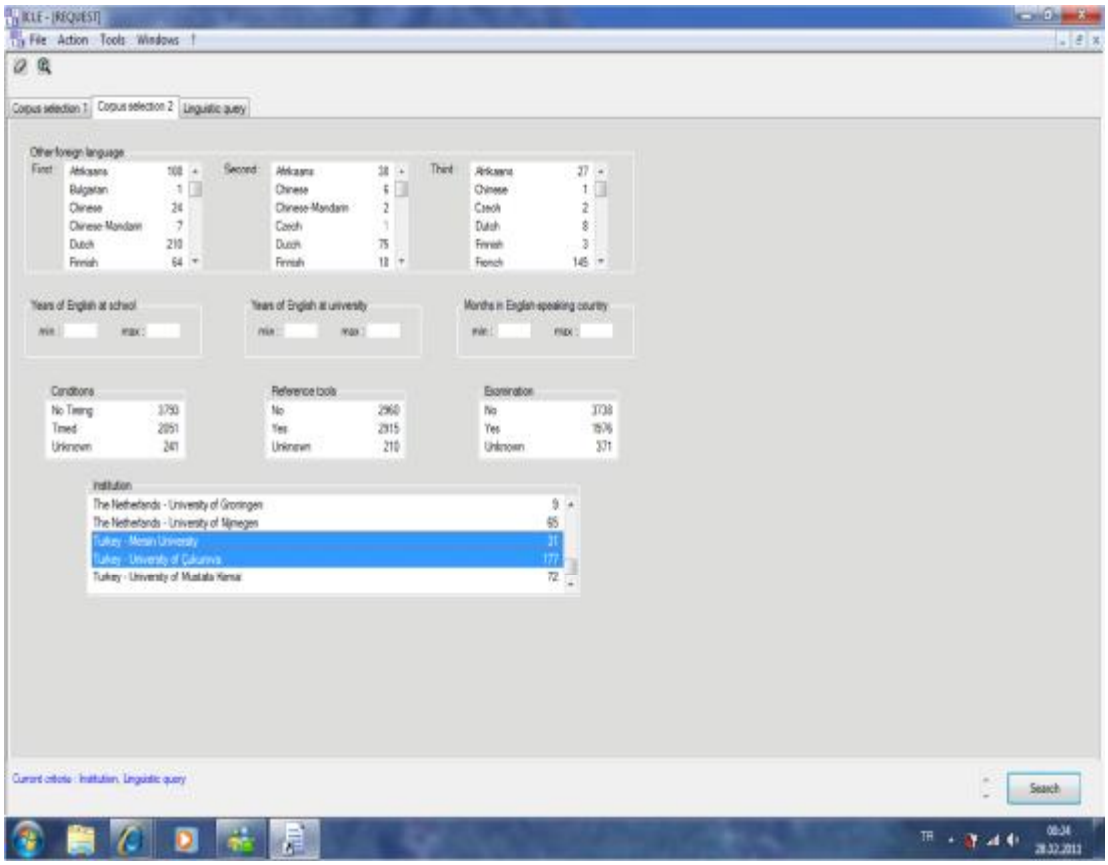


Figure 7. A screenshot of task variables, learner variables and preference of institution extracted from ICLE*v2*.

The last three variables that can be preferred as a basis of data selection are “Years of English at school”, “Years of English at university” and “Months in English-Speaking Country”. The user can determine the range of these variables by selecting the two blanks identified for each title; from minimum to maximum. As it is clear from Figure 6 and 7, we have not specified any other variables than the Institutions in the first and second corpus selection steps.

In Figure 8, we have shown the screenshot of concordance “would” in the TICLE data by selecting the two institutions, namely; University of Çukurova and Mersin University. Figure 8 displays the last step in determining the criteria of the data as “Corpus Selection 3.” This selection is divided into two sub-windows, the upper window show all the details of of the preferred variable in the former steps. In this upper sub-window, the user can see and choose the essay of a specific participant in a grid view or form view. In addition, the user can go to the chosen essay by clicking on the symbol just above the Grid View bottom; .

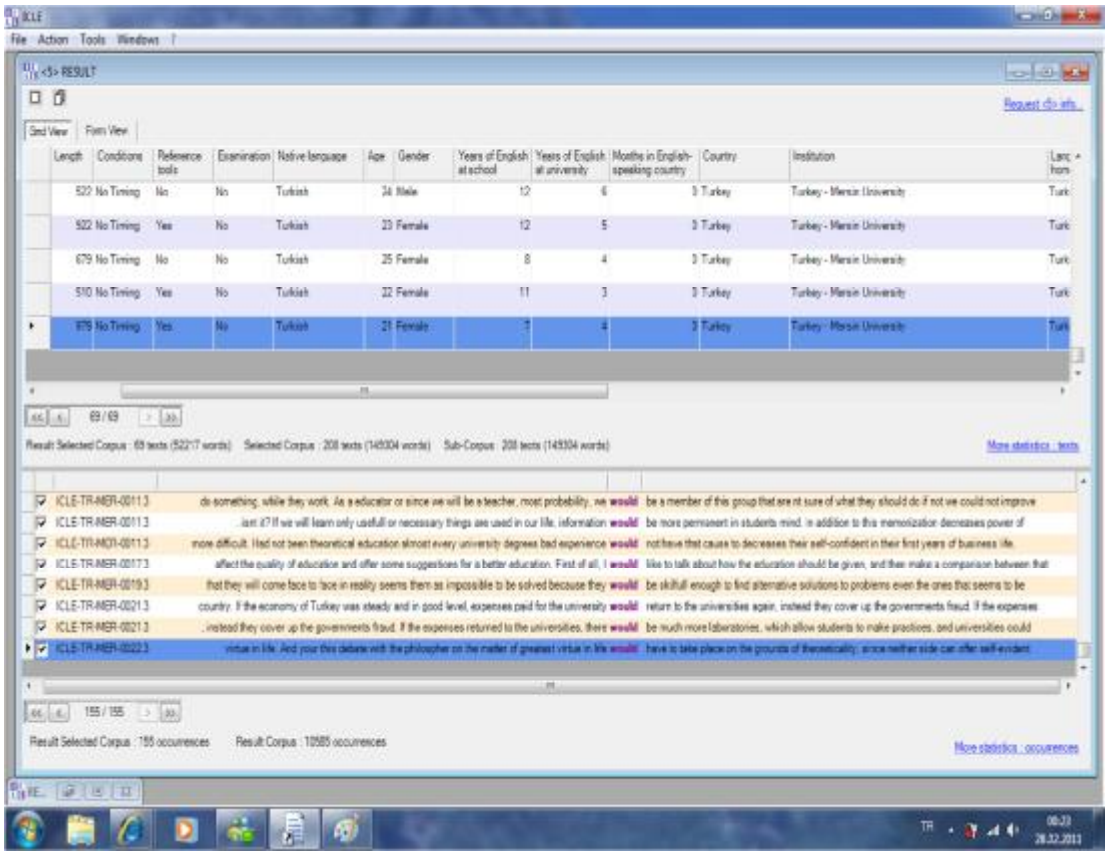


Figure 8. A screenshot of concordance '*would*' within TICLE extracted from ICLE*v2*.

The bottom window illustrates each concordance sentence of the searched linguistic query. The user can also find the number of Result Selected Corpus and Result Corpus occurrences as shown in Figure 8 for the linguistic query of "would." As Figure 6, 7 and 8 show, any user can do their research very easily by the light of the ICLE*v2* software.

The ICLE corpus project team did not asses the participants proficiency. Proficiency of the learners has been rated by a professional assessor on the basis of the "Common European Framework of Reference for Languauges (CEF)" descriptor for writing. The professional rater assessed a random of 20 essays from each of the 16 subcorpora which are detailed in Table 14.

Table 14

CEF Results- 20 essays per subcorpus in ICLE. (Granger et al., 2009, p. 12)

National Corpus (Subcorpora of ICLE)	B2 (and lower)	C1	C2	Total
Bulgarian	2	16	2	20
Chinese	19	1	0	20
Czech	11	9	0	20
Dutch	1	11	8	20
Finnish	3	8	9	20
French	3	11	6	20
German	1	12	7	20
Italian	10	9	1	20
Japanese	18	2	0	20
Norwegian	8	7	5	20
Polish	1	12	7	20
Russian	3	15	2	20
Spanish	12	8	0	20
Swedish	0	14	6	20
Turkish	18	0	2	20
Tswana	16	4	0	20
ICLE*v2*	126	139	55	320

Granger et al. (2009, p. 11) state that while 60% of the sample essays were rated as advanced (C1 or C2), the proportion is much higher in some subcorpora. Though these results need to be firmed up extensively, some of the ICLE*v2* subcorpora are rather in the higher intermediate range while others clearly qualify as advanced.

In the ICLE*v2* project there are sixteen subcorpora consisted of 6,085 essays totalling in 3,753,030 words.

Table 15

Distribution of essays/words per subcorpus in ICLE. (Granger et al., 2009, p. 25)

NATIONAL SUBCORPUS	Number of essays	Number of words
Bulgarian	302	200,194
Chinese	982	490,194
Czech	243	201,687
Dutch	263	234,723
Finnish	390	274,628
French	347	226,922
German	437	229,698
Italian	392	224,222
Japanese	366	198,241
Norwegian	317	211,725
Polish	365	233,920
Russian	276	229,584
Spanish	251	198,131
Swedish	355	200,033
Turkish	280	199,532
Tswana	519	199,173
ICLE*v2*	6,085	3,753,030

Each national corpus has been categorized into several batches, themselves consisted of a variable number of essays. As in the study, Turkish and Japanese subcorpora are examined, following are Tables 16 and 17 illustrating batches, essays, Text Types (A: Argumentative, L: Literary and O: Other), Number of Essays and Number of Words.

Table 16

Turkish subcorpus of ICLE. (Granger et al., 2009, pp. 37)

Batch	Essay Codes	TEXT TYPES			Number	Number
		A	L	O	of Essays	of Words
TRCU1	TRCU1001-1177	177	0	0	177	128,297
TRKE2	TRKE2001-2072	72	0	0	72	50,228
TRME3	TRME3001-3031	31	0	0	31	21,007
	Total	280	0	0	280	199,532

Note. A =Argumentative, L =Literary, O =Other. "TRCU1" and "TRME3" have been preferred as the selection of TICLE segments.

As illustrated in Table 16, Across the institutional subcorpora, TRCU1, TRKE2 and TRME3, forming the TICLE; TRCU1 and TRME3 have been selected to analyze data. These two institutions (also bold and italic cells in Table 16) comprise of 208 essays for a total number of 149,304 words. The second corpus of the study was selected; Japanese subcorpus with a very close number of words to Turkish subcorpus illustrated in Table 17. The institutions codes manipulated in the study are depicted at the Apendix 1.

Table 17.

Japanese subcorpus of ICLE. (Granger et al., 2009, pp. 33, 34)

Batch	Essay Codes	TEXT TYPES			Number	Number
		A	L	O	of Essays	of Words
JPAI1	JPAI1001-1002	2	0	0	2	1,451
JPDO1	JPD01001	1	0	0	1	679
JPFJ1	JPFJ1001	1	0	0	1	622
JPHI1	JPHI1001-1020	5	0	0	5	2,887
JPKO1	JPKO1001-2031	20	0	0	20	10,762
JPKO2	JPKO2001-2031	31	0	0	31	18,871
JPKW1	JPKW1001-1002	2	0	0	2	1,160

(Table 17. Continued)

JPKY1	JPKY1001-1002	2	0	0	2	1,206
JPMI1	JPMI1001-1002	2	0	0	2	1,123
JPMJ1	JPMJ1001-1002	2	0	0	2	1,057
JPMU1	JPMU1001-1002	2	0	0	2	972
JPNH1	JPNH1001	1	0	0	1	538
JPOK1	JPOK1001	1	0	0	1	837
JPRI1	JPRI1001-1002	2	0	0	2	1,109
JPSE1	JPSE1001	1	0	0	1	834
JPSH1	JPSH1001-1004	4	0	0	4	2,422
JPST1	JPST1001-1002	2	0	0	2	1,248
JPSW1	JPSW1001-1039	39	0	0	39	18,338
JPSW2	JPSW2001-2021	21	0	0	21	11,691
JPSW3	JPSW3001-3031	31	0	0	31	16,532
JPSW4	JPSW4001-4032	31	0	0	31	16,735
JPTF1	JPTF1001-1043	43	0	0	43	23,011
JPTK1	JPTK1001-1002	2	0	0	2	987
JPTM1	JPTM1001-1028	28	0	0	28	16,793
JPWA1	JPWA1001-1019	19	0	0	19	9,433
JPWA2	JPWA2001-2009	9	0	0	9	4,291
JPWA3	JPWA3001-3020	20	0	0	20	10,097
JPWA4	JPWA4001-4012	12	0	0	12	7,257
JPWA5	JPWA5001-5029	29	0	0	29	14,649
JPWA6	JPWA6001	1	0	0	1	649
	Total	366	0	0	366	198,241

Note. A =Argumentative, L =Literary, O =Other.

In Table 17 it is illustrated that the underlined institutes; “JPAI1”, “JPDO1”, “JPFJ1”, “JPHI1”, “JPKO1”, “JPKO2”, “JPKW1”, “JPKY1”, “JPMI1”, “JPMJ1”, “JPOK1”, “JPRI1”, “JPSE1”, “JPSH1”, “JPST1”, “JPSW1”, “JPSW2”, “JPSW3”, “JPSW4”, “JPTF1” and “JPTM1”, were selected to analyse data. These institutions (the underlined cells in Table 17) comprise of 271 essays for a total number of 149,368 words. The institutions codes manipulated in the study are depicted at the Appendix 2.

3.2.2. LOCNESS

As TICLE and JPICLE are NNs corpora, we needed to compare the data extracted from the learner corpora by a Ns corpus. In corpus research studies, LOCNESS has been preferred as a native corpus, i.e. control corpus or reference corpus, in comparison with learner corpora, especially with ICLE. Granger et al. (2009, p. 42) state that "to ensure comparability with the ICLE data, the Louvain team has collected a corpus of essays written by native English students, the Lovain Corpus of English Essays (LOCNESS), which is the mirror of the ICLE". Since LOCNESS is the mirror of the ICLE, it has very similar criteria applied in the design of the ICLE that provides the opportunity to compare the data within each other.

There are two reasons lying under the preference of LOCNESS as a control corpus. The first reason is the comparability between ICLE (TICLE and JPICLE) and LOCNESS. The second reason is that LOCNESS is the NS corpus most commonly used and suggested for comparison so far as in the studies of Ringbom (1998b), Lorenz (1998) Virtanen (1998), Granger and Petch-Tyson (1996), Aarts and Granger (1998), Aijmer (2002), Lin (2002), Narita, Sato, & Sugiura (2004), Altenberg & Tapper (1998), and Tapper (2005). It may be uttered that LOCNESS has gained a reliability being a reference corpus in the field of learner English study.

LOCNESS is a corpus of native English essays made up of:

- British pupils' A level essays: 60,209 words
- British university students essays: 95,695 words
- American university students' essays: 168,400 words

Total number of words: 324,304 words

http://www.uclouvain.be/en-cecl-locness.html

LOCNESS was built by the Centre for English Corpus Linguistics at the Catholic University of Louvain, Belgium and made available for public use in 1998. It can be summarised by the statements of Guo (2006, p. 53) as in the following:

The texts of the corpus are essays produced by British and American native speakers from 1991 to 1995. The corpus comprises four components, i.e., essays of British A-Level students, essays of British university students, argumentative essays of

American students and literary-mixed essays of American students. The texts of the corpus include examination papers, timed essays and free essays which increases the comparability across ICLE. No reference tools were used in examination papers whereas in some timed essays and free essays reference tools were used. The length of essays is around 500 words; it is similar to the basic design criteria of ICLE. The age of students is mostly between 17 and 23 though there are a very small number of students much older. Although the NS profile for the essays of British A-Level students and of British university students is not available, it can be assumed that most of the students are NSs of English. The texts cover a very wide range of topics from social problems such as water pollution, nuclear power, sex, violence, and gender roles to campus-related issues such as cheating in college, controversy in the classroom, and prayer in schools. Both parts of the corpus, the British essays and the American essays, have country-specific topics. In contrast to this, the major interest in American essays is found in quite different areas such as the Confederate Flag, the US government, book banning in America, gun control and the legalisation of marijuana. The topics are also very similar presented in the ICLE corpus specifically being argumentative essays. From the topics covered by the texts, the overall feature of the writing style can be interpreted as argumentative.

In the current study, argumentative essays of American students have been preferred as a selection providing an approximate number of words selected in ICLE. WordSmith Tools Analyzing Software was used to retrieve the data in LOCNESS.. Following is the institutions with codes, number of essays and words for the selected segment (American) of LOCNESS.

Table 18

Selected proportion of LOCNESS data distribution.

Institution	Codes	Number of Essays	Number of Words
Marquette University	ICLE-US-MRQ	46	54,285
Indiana University at Indianapolis	ICLE-US-IND	27	13,382
Presbyterian College, South Carolina	ICLE-US-PRB	6	12,447
University of South Carolina	ICLE-US-SCU	53	52,885
University of Michigan	ICLE-US-MICH	43	16,502
	Total	175	149,501

3.2.3. TUC

Turkish University Corpus (TUC) has been compiled for this current and more studies in corpus research. It is a native corpus, i.e. reference corpus or control corpus, which is to be made use of in comparison with especially TICLE data and of course other corpora selected in the study. TUC has been designed in University of Çukurova and compiled by the data collected from mainly two Universities, namely; University of Çukurova and Kahramanmaraş Sütçü İmam University. TUC has not been made available for public research yet. Table 19 illustrates the distribution of the institutions, codes, number of essays and words utilized in TUC.

Table 19

Selected proportion of TUC data distribution.

Institution	Codes	Number of Essays	Number of Words
University of Çukurova	TUC-CU	75	45,119
Kahramanmaraş Sütçü İmam University	TUC-KSU	108	62,757
	Total	183	107,876

As Table 19 illustartes, the total number of words is 107,876 compiled out of the written performances of 183 participants. From University of Çukurova, 75 essays have been collected which is equal to 45,119 in this corpus. The rest 108 essays have been collected from Kahramanmaraş Sütçü İmam University with a number of 62,757 words.

TUC has the same criteria in line with ICLE to be highly comparable including argumentative essays from Turkish University students in their L1 Turkish. There are two points that TUC differs from ICLE; the first one is; the topic asked to the participants in TUC are all in Turkish (Turkish translations of the topics presented in ICLE (see Appendix 3) that the participants submitted their argumentative essays in their mother tongue. The other point is that the data collected from the participants have been analysed by WordSmith Tools analyses software programme, while in ICLE, CLAWS7 was used for retrieving, sorting and concordancing the data. As TUC is in the same design criteria with ICLE, it is very close to LOCNESS in comparability that ICLE is.

3.3. Wordsmith Tools Version 5.

Granger et al. (2009, p. 41) state that among the learner corpus researches, the most popular, sufficiently versatile and powerful tool is WordSmith Tools allowing for sophisticated linguistic manipulations of the data. In the current study, the fifth version of Wordsmith Tools (Scott, 2008) has been used (Version 1; Scott, 1996, Version 2; Scott, 1997, Version 3; Scott, 1999, Version 4; Scott, 2004). The user of WordSmith Tools can use the tools to find out how words are used in the selected texts and has integrated suite of programs in order to look at the uses of the words in selected texts. Wordsmith Tools has three components; *The Wordlist* providing a list of all the words or word-clusters in a text of alphabetical or frequency order. The second is *Concord* which is a concordancer that lets the user to see any word or phrase in context which represents a sort of company that the word keeps. The third component is *Keywords* that the user can find the key words in a text. The tools have been used by many language teachers or students, and by researchers studying language patterns in many different languages all over the world that have been also used by Oxford University Press for their own lexicographic work in preparing dictionaries (Scott, 2010, p.2).

In the current study, *Concord* has been used to extract the selected language patterns by the texts of LOCNESS-US and TUC materials. In many learner corpora

studies, Wordsmith Tools have been used and/ or suggested to the corpus researchers by the material of LOCNESS or any other corpus; Campoy, Bellés-Fortuno and Gea-Valor (2010), Cobb (2003), Gabrielatos (2007), Ghadessy, Henry, & Roseberry (2001), Granger (2004, 2008), Guo (2006), Kaszubski (1998), Ke (2004), McEnery (2005), Narita, Sato & Sugiura (2004), Oakes (1998) and many other journal articles have used WordSmith Tools (Over than 150 articles have used thise software programme, visit the web-link to see a list of those articles http://www.lexically.net/wordsmith/corpus_linguistics_links/articles_using_wordsmith.htm).

3.4. Data Analysis

Ke (2004, p.10) suggests main hypothetical modals as "would", "could", "might" and Kerslake (2003, p. 219) suggests strong hypothetical modals as "would past participle", "could past participle", "might past participle", by bearing this in mind those modals we have analyzed these six variables across four (TICLE; JPICLE; LOCNESS and TUC) corpora. The variables in concern have been searched over TUC by their Turkish counterparts; however, they have been retrieved to observe the use in "if clauses." Subsequently, the meanings were appointed to the above-mentioned variables by their Turkish counterparts, which have already been illustrated in Table 7. In addition to the variants shown in Table 7 *"Eğer, mademki, koşuluyla, takdirde"* were also searched over TUC to observe any related hypothetical conditionals. We used Concord tool to extract the data searched. As an initial step, we have screened out the entire sentences by all the selected variables in English and their counterparts in Turkish. This step was applied one by one to each corpus (LOCNESS and TUC). Then we began to survey the occurrence of the patterns in concern, however, we eliminated some very ambiguous sentences, and otherwise, it could result in misinterpretation on the process of judging the searched variables.

12 other variables, which have been categorized for errors made, have been derived across the same four corpora. These 12 variables have been derived from each main variable in two subcategories which illustrate the errors made; if clauses and main clauses. For example, for the main variable *would*; we have illustrated two subcategories; (1) would errors in if clause and (2) would errors in main clause for error analysis. Totally, 18 variables have been analyzed in this study as detailed in Table 20

below.

Table 20

All variables analyzed across four corpora- TICLE, JPICLE, LOCNESS, TUC- with their abbreviations processed in SPSS.

Abbreviation	Definition
If Would	Would used in If Conditional Sentence
Would EIC	Would used as an Error in If Clause
Would EMC	Would used as an Error in Main Clause
If Could	Could used in If Conditional Sentence
Could EIC	Could used as an Error in If Clause
Could EMC	Could used as an Error in Main Clause
If Might	Might used as in If Conditional Sentence
Might EIC	Might used as an Error in If Clause
Might EMC	Might used as an Error in Main Clause
If Would PP	Would Past Participle used in If Conditional Sentence
W PP EIC	Would Past Participle used as an Error in If Clause
W PP EMC	Would Past Participle used as an Error in Main Clause
If Could PP	Could Past Participle used in If Conditional Sentence
Could PP EIC	Could Past Participle used as an Error in If Clause
C PP EMC	Could Past Participle used as an Error in Main Clause
If Might PP	Might Past Participle used in If Conditional Sentence
M PP EIC	Might Past Participle used as an Error in If Clause
M PP EMC	Might Past Participle used as an Error in Main Clause

Note. PP= Past Participle; EIC= Error in If Clause; EMC= Error in Main Clause

For each of the preferred items we have applied two statistical analyses; the first one reports overall frequency and ratio of the variables and the second one has been processed by Statistical Package for Social Sciences Version 17 for Windows – Pearson Chi-square Test (SPSS V. 17.0 for Win.). Aarts and Granger (1998), Altenberg and Tapper (1998), Biber (2006), De Cock and Granger (2004), Freddie (2005), Gabrielatos (2007), Granger (1996a, 1998, 1999, 2002, 2004, 2008), Granger and Tyson (1996), Oakes (1998), Ringbom (1999), Scott (1996, 1997, 1999, 2004, 2008) and many others have suggested and/ or used frequency statistics by SPSS and / or Word Smith Tools.

Granger (2002, p.4) suggests this type of analyses in corpus-based studies especially to be conscious of the existing knowledge in language by regarding frequency as a perspective of language " ... of which we have very little intuitive awareness but one that plays a major part in many linguistic applications that require knowledge not only of what is possible in language but what is likely to occur."

There have been many techniques for testing statistical significance. McEnery, Xiao, Tono (2006, p.55) states that "The most commonly used statistical test in corpus linguistics is probably the chi-square test (also called the Pearson chi-square test)". Stefanowitsch (2005) states the need for chi-square in line with frequency analyses by the following statement;

> ... we have to know what distribution we would have expected based on chance. Once we know these expected frequencies, we can compare them to our observed frequencies to see whether the latter are different enough for us to say that they cannot have come about by accident.
>
> http://www-user.uni-bremen.de/~anatol/qnt/qnt_dist.html

In addition, chi-square test serves for testing whether the observed difference is occurring not just by chance i.e. statistically significant. Biber, Conrad, Reppen (1998, p.9) state the significance of using chi-square test as "significance test show how likely it is that quantitative results could have occurred by chance, and thus they always be reported in research articles describing a corpus-based study".

As three of the corpora are very close to each other in number (TICLE= 149,304; JPICLE= 149,368; LOCNESS= 149,501), and one of the corpora (TUC= 107,876) differs in size with others, we have decided to run the analyses on chi-square which has been run on many similar corpus-based studies as Aijmer (2002), Connor, Precht, & Upton (2002), Leńko-Szymańska (2004), Narita, Sato, & Sugiura (2004), Tapper (2005), Tono (2004), and many others.

The chi-square test has been used to determine whether categorical data has the same proportion of use across participants in four corpora. The tests in this study show that the overuse and underuse by indicating significance of difference across corpora in comparison to each other. Separate chi-square tests have been cross run for each of two groups compared in order to observe the significant difference among the four groups.

CHAPTER IV

DATA ANALYSIS AND RESULTS

4.1. Introduction

In this study, we aim to explore the use of hypothetical conditionals across four groups of corpora; namely, TICLE, JPICLE, LOCNESS and TUC. While Wordsmith Concord Tool has been used to represent the data obtained from LOCNESS and TUC. CLAWS7, an integrated tool of ICLE, has been employed for the data of TICLE and JPICLE. Having provided the raw data results by means of these two tools, the related frequency of occurrence and correct use of hypothetical conditionals have been analysed on SPSS 17 by running frequency analysis and Pearson Chi-Square Test to make comparisons across groups. Results of the data analysis reveal the target variables by allowing to make sense towards learners' Interlanguage Grammar of the participants. Moreover, this analysis represents the use in quantitative analysis that we could compare across the groups. Each section presents the frequency & ratio and chi-square test analyses of the target variable. Each of the target variables has been indicated under the title of related use in If-Clause Types; Type 2 and Type 3.

4.2. Analyses and Results

Each of the examined target variables has been examined and categorised under the two sub-titles of If-Clauses; Type 2 and Type 3. The first category; Type 2, caters for the examined variables; "would", "could" and "might." The second one; Type 3, caters for the examined variables; "would + have + V*en*", "could + have + V*en*" and "might + have + V*en.*" Table 21 depicts, the overall frequency and ratio analysis of the target variables having been preferred in Type 2 and 3.

Table 21

Overall Frequency and Ratio Analysis of the Target Variables in Type 2 and 3 If-Clauses among four Corpora.

Groups	TICLE		JPICLE		LOCNESS		TUC	
	f	*%*	*f*	*%*	*f*	*%*	*f*	*%*
t 2	67	0.044	91	0.060	150	0.100	92	0.085
t 3	4	0.002	8	0.005	11	0.007	39	0.036
∑	149,304		149,368		149,501		107,876	

Note. *TICLE* =*Turkish International Corpus of Learner English*
JPICLE = *Japanese International Corpus of Learner English*
LOCNESS = *Louvain Corpus of Native English Essays*
TUC = *Turkish University Corpus*
∑ = *Total Tokens*
t = *Type*

As illustrated in Table 21; the participants in TICLE; have used the variables "would", "could" and "might" 67 times out of the examined total tokens; 149,304. This result is not similar to any other groups in Table 21. Turkish EFL learners have used the variables in Type 2 category much more frequently in their mother tongue, statistically, 92 times out of 107,876 total tokens. It is impressing that Turkish EFL learners underuse those variable in Type category of If-Clause Sentences in target language, English, when compared to their performances in their mother tongue, Turkish. That is to say, Turkish EFL learners use the variables "would", "could" and "might" in Type 2 with the percentage of 0.044 in English and the use of these variables in mother tongue is rated as 0.085%. However, more interestingly, Japanese EFL learners' written performances reflected a more stable use of "would", "could" and "might" value in Type 2 category; 91 times, when compared to the results of Turkish EFL learners'. The use of these variables is observed by the native English speakers in LOCNESS data and the participants have used "would", "could" and "might" with the frequency of 150. That is to say, Japanese EFL learners may probably underuse "would", "could" and "might" as they have used Type 2 with the percentage of 0.060 while these variables have been used 0.100 % in the target language. As Table 21 illustrates, the results obtained from TICLE and JPICLE do not match to each other in terms of having similar

numerical proof, and the use has been observed as a sort of mother tongue effect; it may be wise to claim that there would not be any possible interlanguage property on the use of "would", "could" and "might" in Type 2 If-Clause structures examined in the written performances of JPICLE and TICLE interlanguage data.

The examined variables "would + have + V*en*", "could + have + V*en*" and "might + have + V*en*" in Type 3 If-Clause structures among four corpora reflect a little difference to the use examined with the variables "would", "could" and "might" in Type 2 If-Clause structures. Table 21 presents the frequency of the variables "would + have + V*en*", "could + have + V*en*" and "might + have + V*en*" which have been used in Type 3 with a ratio of 0.002% by Turkish EFL learners. Taking into consideration the frequency and ratio in the TUC data (*f* 39-0.036%), the statistical results indicate that Turkish EFL learners use these variables much less in the target language. However, the results obtained from JPICLE display that "would + have + V*en*", "could + have + V*en*" and "might + have + V*en*" have been used eight times with the percentage of 0.005% over 149,308 total tokens. These results allow to make a comparison of the observed use between JPICLE and TICLE. It is seen that Japanese EFL leaners have used the examined variables in Type 3 twice more than the Turkish EFL learners and Japanese EFL learners have preferred Type 3 much more close to the use of the native use (JPICLE: *f* 8, 0.005% - LOCNESS: 11, 0.007% - TICLE: *f* 4, 0.002%) while Turkish EFL learners have not displayed such a close use to the target language in their written argumentative essays. The statistical results of JPICLE and TICLE share a common point in that both corpora results, as demonstrated in Table 21, are not close to the use marked in TUC (*f* 39, 0.036%).

Analyses have been divided into subcategories in order to depict a clearer picture of each separate variable and its past participle form. For example, in one column, the quantitative results of the variable "would" have been illustrated on If-Clause Type 2; in the other column, the quantitative results of the variable "would + have + V*en* + " have been illustrated in If-Clause Type 3.

4.2.1. Analysis and Results of "would" and "would + have + V*en*" across Type 2 and Type 3

Table 22 demonstrates that there have been 67 tokens in Type 2 If-Clause sentences recorded in TICLE data. The frequency of 67 was counted of the total use of

"would", "could" and "might." "would" has been preferred in TICLE 58 times with a percentage of 86.57. The remaining frequency over total frequency of 67 stands with the frequency of 9 that have been observed in the uses of "could" and "might" in If-Clause Type 2 conditionals shown in Tables 25 and 28. The second group, JPICLE, has 91 tokens in the total sum frequency of Type-2 If-Clause patterning. "would" has been used 48 times regarding a ratio of 52.74%, the rest of the total tokens refers to the total use of "could" and "might" with the total frequency of 43. The native essay corpus (LOCNESS), selected as the control group, is found to have 118 frequency of "would" with a percentage of 78.66% out of the total use of "would", "could" and "might" (*f* :150).

Table 22

Overall Frequency and Ratio Analysis of "would" and "would + have + V*en*" *in Type 2 and 3 If-Clauses among four Corpora.*

Groups	TICLE			JPICLE			LOCNESS			TUC		
	∑t	*f*	*%*	∑t	*f*	*%*	∑t	*f*	*%*	∑t	*f*	*%*
t 2 "would"	67	58	86.57	91	48	52.74	150	118	78.66	92	76	82.60
t 3 "would + have + V*en*"	4	4	100	8	3	37.50	11	10	90.90	39	30	76.92

Note. *TICLE* =*Turkish International Corpus of Learner English*
JPICLE = *Japanese International Corpus of Learner English*
LOCNESS = *Louvain Corpus of Native English Essays*
TUC = *Turkish University Corpus*
∑*t* = *Total Type Tokens*
t = *Type*

As it can be clearly seen in Table 22, the use of "would" by Turkish EFL learners depicts that they use the variable in closer manner to the examined native norm in English and Turkish than the examined use in JPICLE, because the frequency in TUC is 76 with a percentage of 82.60 out of the total use in Type 2 (*f*: 92). Table 22 clearly illustrates that Turkish EFL learners use "would" in Type 2 If-Clause sentences more frequently than Japanese EFL learners do in their written argumentative essays. On one hand, taking into consideration the frequency analysis, it is clear that American native speakers of English use "would" in Type 2 If-Clause structures much more frequently

than any other observed group in the study. This is followed by the use in Turkish mother tongue, thirdly; TICLE, finally the last and the least frequent use comes from JPICLE's examined result. On the other hand, taking into consideration the ratio analysis, the use of "would" changes in terms of the order among the groups. Turkish EFL learners use "would" in Type 2 If-Clauses much more than any other group in the study. The second most use is remarked by the American native speakers' examined results. The third most use has been observed in the examined Turkish mother tongue productions. Finally the Japanese EFL learners possess the least use, as it is displayed with the frequency analysis result. In order to reach a conclusion on the use of "would" in Type 2 If-Clauses language pattern, we have attributed our comments to the results obtained via ratio analysis. Turkish EFL learners may overuse the item in concern; however, it would be easier to say that Japanese EFL learners may underuse "would" when compared to the use in the written performances of American native speakers'. The overuse by Turkish EFL learners may be particularly related to the use observed in Turkish mother tongue, because the use of proportion extracted from TICLE is 86.57% and TUC is 82.60%. This indicates that Turkish EFL learners may use "would" by the effect of their mother tongue. However, it needs to be confirmed by the results obtained via chi-square test in Table 23.

Table 22 displays that the variable "would + have + V*en*" use has been preferred mostly by the Turkish EFL learners referring 100% of use in Type 3 If-Clause sentences. Turkish EFL learners have not preferred other forms of variables; namely, "could + have + V*en*" and "might + have + V*en*" as a selection in their written productions in target language. Secondly, "would + have + V*en*" has been mostly used in Type 3 If-Clause sentences by the American native speakers referring 90.90%. Thirdly, the use in Turkish mother tongue has revealed percentage of 76.92. The last and the least use is hold by the Japanese EFL learners' referring 37.50%. The same conclusion reached by the use of "would" can be retrieved again for "would + have + V*en*"; that is, Turkish EFL learners may overuse "would + have + V*en*" in Type 3 If-Clause sentences. But most probably, it can be attributed as a possible mother tongue effect, which is observed in their written performances in target language. Again, this conclusion should be confirmed by Table 23 which represents the chi-square results.

As illustrated in Table 23, Pearson chi-square test result of "would" use between the pairs; TICLE-JPICLE and TICLE-TUC, does not indicate any significant difference. On the other hand, the p value has been measured lower than 0,01 between LOCNESS

and TUC. Additionally, the p value is lower than 0,001 between TICLE and LOCNESS, and JPICLE and LOCNESS.

Table 23

Overall Pearson Chi-Square Results of "would" across four Corpora

	would	
Groups	c^2	**(p<.05)**
TICLE-JPICLE	1,017	0,313
TICLE-LOCNESS	21.200	0,000***
TICLE-TUC	3,597	0,058
JPICLE-LOCNESS	34,905	0,000***
LOCNESS-TUC	7,120	0,008**

*Note. Sig. = significance of difference *p<, 05; **p<, 01; ***p<, 001.*

Productions made by Turkish-speaking EFL learners in English and those made by the Native Group in English (LOCNESS) have been compared in order to see whether there is a significant difference across these groups with respect to the use of "would." It has been revealed that the former has a more tendency in using the variable in concern. The same usage has been examined for L2 and mother tongue and the test results have indicated that the Turkish EFL learners and Turkish native speakers do not significantly differ significantly in the use of "would." In other words, Turkish EFL learners employ "would" more frequently in the essays they write in English than those they write in Turkish. Finally, LOCNESS and TUC have been compared to find out whether there is a significant difference in regard to the use of "would." No significant difference has been found in this concern (.008= p< 0.05).

The same procedure has been applied to the productions made by Japanese EFL learners in English. The test results have shown that Japanese EFL learners use "would" less frequently than the American Native speakers of English (.000= p< 0.05). Consequently, the two-interlanguage productions of JPICLE and TICLE have not shown any significant difference values to one each other in using "would" in hypothetical conditionals, but each group has revealed statistically significant difference in comparison to that of the use in the control group, LOCNESS.

Table 24

Overall Pearson Chi-Square Results of "would + have + V*en*" *across four Corpora*

	would + have + V*en*	
Groups	c^2	**(p<.05)**
TICLE-JPICLE	0,001a	0,976
TICLE-LOCNESS	4,288	0,038*
TICLE-TUC	17,009	0,000***
JPICLE-LOCNESS	0,647b	0,421
LOCNESS-TUC	4,672	0,031*

*Note. Sig. = significance of difference *p< ,05; **p< ,01; ***p< ,001*

a= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 3,04;

b= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 0,39.

As Table 24 illustrates, overall chi-square results of "would + have + V*en*" between the pairs TICLE – JPICLE, and JPICLE – LOCNESS cannot be calculated as the (f) value is observed 5 and less than 5 in 20% and more than 20% of the cells, chi-square significance test could not be examined. In this situation, as the variable in concern could not be examined by means of chi-square analysis, the descriptive analyses results have shown that there seems to be a difference between TICLE – JPICLE, and JPICLE – LOCNESS with respect to the frequency results as illustrated in Table 22. However, the results have showed that there is a significant difference between Turkish EFL learners and American native speakers of English in regard to the use of "would + have + V*en*" form in hypothetical conditionals. That is, Turkish EFL learners use this variable more frequently than Japanese EFL learners. American native speakers and Turkish speakers use this variable much more often than Japanese EFL learners. The obtained results of Turkish EFL learners and Turkish mother tongue use represent no significant difference (0,058= p>0,001). Hence, it has been confirmed that this variable is underused in both the English interlanguages of Japanese EFL learners' and Turkish EFL learners'. It may be concluded that the underuse of this item in this study may have revealed a possible property of English interlanguage as both learner productions reveal no significant difference. A significant difference has been found between American Native use and Turkish mother tongue use in this regard. This supports the view that American native participants use this variable most frequently of

all the groups. In conclusion, chi-square results have confirmed the results presented and commented on Table 22 that no mother tongue effect has been found since the observed property of interlanguage has revealed statistically significant difference over the use in Turkish mother tongue use.

4.2.2. Analysis and Results of "could" and "could + have + V*en*" across Type 2 and Type 3

The total use of "would", "could" and "might" in TICLE data refers with the frequency of 67 as displayed in Tables 22. As illustrated in Table 22, the use of "would" has a frequency of 58 and the rest was used nine times out of 67. Table 25 demonstrates the frequency of "could" with the frequency of seven. So, the use of "could" in Type 2 If-Clause sentences has frequency of seven and percentage of 10.45. TUC has revealed the frequency of 14 and percentage of 15.21, which is the closest ratio to that of TICLE's. This can be interpreted as that Turkish EFL learners' use of "could" in Type 2 If-Clause sentences in the target language displays similarity to their use of the same structure in their mother tongue. The use of "could" by the American participants has indicated that they use "could" slightly more often than they used it in TUC; in other words, the use of "could" in TUC is close to that of the examined use in LOCNESS (17.33%).

Table 25

Overall Frequency and Ratio analysis of "could" and "could + have + V*en*" *in Type 2 and 3 If-Clauses among four Corpora.*

Groups	TICLE			JPICLE			LOCNESS			TUC		
	∑t	*f*	*%*	∑t	*f*	*%*	∑t	*f*	*%*	∑t	*f*	*%*
t 2 "could"	67	7	10.45	91	32	35.16	150	26	17.33	92	14	15.21
t 3 "could + have + V*en*"	4	0	0	8	2	25.00	11	1	9.09	39	9	23.07

Note. *TICLE* = *Turkish International Corpus of Learner English*
JPICLE = *Japanese International Corpus of Learner English*
LOCNESS = *Louvain Corpus of Native English Essays*
TUC = *Turkish University Corpus*
∑*t* = *Total Type Tokens*
t = *Type*

Out of the four corpora examined in the study, the tokens extracted from the JPICLE data have revealed that JPICLE has had the most frequent use of “could” in Type 2 If-Clause sentences among the other groups with a frequency of 32 and percentage of 35.16. That is to say, in this study it has been revealed that Japanese EFL learners have overused “could” in Type 2 If-Clause sentences when compared to the use of corresponding native language structure in LOCNESS (*f*: 26, 17.33%). Considering the use of “could”, Japanese and Turkish students have not displayed similar occurrence in terms of their interlanguage productions in their written essays, because Turkish EFL learners may have underused “could” in Type 2 If-Clause structures when compared to that of native use in LOCNESS. However, Turkish EFL learners have used the same variable in concern in a similar rate with that of TUC’s results. This may be accepted as a result indicating that Turkish EFL learners may have a potential to get any possible effect of Turkish mother tongue in using “could” in Type 2 If-Clauses.

Turkish EFL learners have not preferred any language patterning of “could + have + V*en*” in Type 3 in If-Clause sentences. However, patterning of “could + have + V*en*” in Type 3 in If-Clause sentences has been preferred mostly over the other three groups by the data extracted from Turkish mother tongue corpus (TUC) with the frequency of nine out of 39 and 23.07 %. The closest ratio to that of TUC’s belongs to the analysis conducted on Japanese EFL learners’ written performance in which JPICLE has revealed frequency of two and percentage of 25.00. Another native corpus in this study, LOCNESS, has not indicated such a large percentage of what JPICLE and TUC have employed. The frequency has been observed as one out of 11 token frequency, which refers to 9.09%. Considering the use of “could + have + V*en*” in Type 3 in If-Clause sentences, it may be uttered that the two interlanguage productions of JPICLE and TICLE have not indicated any close relationship in terms of frequency. However, JPICLE has revealed very close result with that of TUC statistically.

As seen in Table 26, the difference between Turkish and Japanese subcorpora of ICLE has been found statistically significant with respect to the use of “could” (.041= $p<0.05$).

Table 26

Overall Pearson Chi-Square Results of "could" across four Corpora.

	could	
Groups	c^2	**(p<.05)**
TICLE-JPICLE	4,187	0,041*
TICLE-LOCNESS	9,029	0,003**
TICLE-TUC	0,269	0,604
JPICLE-LOCNESS	1,476	0,224
LOCNESS-TUC	5,844	0,016*

*Note. Sig. = significance of difference *p< ,05; **p< ,01; ***p< ,001;*

More statistically significant difference has been measured between TICLE and LOCNESS regarding the use of "could" (.003= p<.05) which means that the use of "could" has been observed more often in native use by American participants. Hence, Turkish-speaking EFL learners underuse "could" in comparison with the American Native users of English. On the contrary, the difference between Turkish-speaking EFL learners and Turkish native users with respect to the use of "could" is not statistically significant. In other words, the use of "could" in hypothetical conditionals has been observed in the similar sense within mother tongue usage and L2 usage, which might be attributed to mother tongue effect on L2 since the item in concern has occurred slightly more often in L1 than in L2.

In contrast, its use in English native language has been found to be relatively higher in L2. It leads us to reach the conclusion that the overuse may be attributed to a sort of interlanguage property considering the results between TICLE and LOCNESS. On the other hand, the data extracted from JPICLE have not confirmed such a conclusion, as there is not a statistically significant difference between Japanese L2 English use and American Native use.

Nonetheless, the usage proportion of "could" by Japanese EFL learners has a lower proportion than the use within American Native use as it is observed in the same sense of TICLE and LOCNESS. All in all, native American speakers of English and Turkish EFL learners significantly differ in the use of "could" (….= p<0.05).

As Table 27 illustrates, overall chi-square results of "could + have + V*en*" among all the groups can not be calculated as the (f) value is observed as 5 and less than

5 in 20% and as more than 20% of the cells, chi-square significance test could not be examined.

Table 27

Overall Pearson Chi-Square Results of "could + have + Ven" across four Corpora

	could + have + V*en*	
Groups	c^2	**(p<.05)**
TICLE-JPICLE	0,769a	0,380
TICLE-LOCNESS	1,192c	0,275
TICLE-TUC	6,926d	0,008
JPICLE-LOCNESS	0,098e	0,755
LOCNESS-TUC	3,420b	0,064

Note. a:2 cells (50,0%) have expected count less than 5. The minimum expected count is 0,43.
b: 2 cells (50,0%) have expected count less than 5. The minimum expected count is 0,46.
c: 2 cells (50,0%) have expected count less than 5. The minimum expected count is 2,81.
d: 2 cells (50,0%) have expected count less than 5. The minimum expected count is 0,78.
e: 2 cells (50,0%) have expected count less than 5. The minimum expected count is 2,82.

The descriptive comments made on the overall use in Table 25 have been preferred as the basis of the study for this item.

4.2.3. Analysis and Results of "might" and "might + have + V*en*" across Type 2 and Type 3

The results obtained from TICLE and TUC have showed the same frequency of two out of different number of total tokens (TICLE: 67, TUC: 92) as illustrated in Table 28. However, the percentages of the use of "might" in Type 2 If-Clauses have demonstrated a similar result (TICLE: 2.98%, TUC: 2.17). This similar use of proportion may be accepted as a possible mother tongue effect observed in the target language and mother tongue productions of different Turkish informants in the study. Another leaner corpus group, JPICLE, has revealed the largest proportion in using "might" in Type 2 in If-Clauses having a frequency of 11 out of 91 total tokens and 12.08% of overall use within Type 2. The data extracted from LOCNESS corpus has exhibited conspicuously

much more use than that of TICLE and TUC's; however, less use than that of JPICLE's. That is to say, Japanese EFL learners' use of "might" Type 2 If-Clauses has propounded predisposition to overuse the variable in concern when compared to that of American native speakers' written performances.

Table 28

Overall Frequency and Ratio Analysis of "might" and "might + have + V*en*" *in Type 2 and 3 If-Clauses among four Corpora.*

Groups	TICLE			JPICLE			LOCNESS			TUC		
	∑t	*f*	*%*	∑t	*f*	*%*	∑t	*f*	*%*	∑t	*f*	*%*
t 2 "might"	67	2	2.98	91	11	12.08	150	6	4.0	92	2	2.17
t 3 "might + have + V*en*"	4	0	0	8	3	37.50	11	0	0	39	0	0

Note. *TICLE* =*Turkish International Corpus of Learner English*
JPICLE = *Japanese International Corpus of Learner English*
LOCNESS = *Louvain Corpus of Native English Essays*
TUC = *Turkish University Corpus*
∑*t* = *Total Type Tokens*
t = *Type*

Table 28 has set forth the fact that the groups other than JPICLE have not revealed any observed usage of "might + have + V*en*" in Type 3 If-Clauses in this study. In JPICLE the use of "might + have + V*en*" in Type 3 If-Clause sentences has served with the frequency of three out of eight tokens and 37.50% proportionally. It is possible to comment that overuse of the variable in concern has been committed by Japanese EFL learners, whereas, the examined variable has not been preferred by any other groups, specifically, the reference or control corpus, LOCNESS.

According to the statistical results displayed in Table 29, overall chi-square result of "might" can not be calculated as the (f) value is observed as 5 and less than 5 in 20% and as more than 20% of the cells, Pearson chi-square significance test could not be administered.

Table 29

Overall Pearson Chi-Square results of '"might" across four Corpora

	might	
Groups	c^2	**(p<.05)**
TICLE-JPICLE	2,920a	0,087
TICLE-LOCNESS	1,903c	0,168
TICLE-TUC	0,017d	0,898
JPICLE-LOCNESS	0,075	0,784
LOCNESS-TUC	1,452b	0,228

*Note. Sig. = significance of difference *p< ,05; **p< ,01; ***p< ,001.*

a= 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,78;

b= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 3,42;

c= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 3,20;

d= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 1,87;

The frequency Table 28 has been considered to comment on the variable "might" descriptively. Considering the statistical results demonstrated in Table 30, overall chi-square result of "might + have + V*en*" among all the groups cannot be calculated as the (f) value is observed as 5 and less than 5 in 20% and more than 20% of the cells, chi-square significance test could not be examined.

Table 30

Overall Pearson Chi-Square Results of "might + have + Ven" across four Corpora

	might + have + Ven	
Groups	c^2	**(p<.05)**
TICLE-JPICLE	2,317a	0,128
TICLE-LOCNESS	,b	NA
TICLE-TUC	,c	NA
JPICLE-LOCNESS	1,950d	0,163
LOCNESS-TUC	,f	NA

Note. a:2 cells (50,0%) have expected count less than 5. The minimum expected count is 1,30.

b: No statistics are computed because if might past participle is a constant.

c: No statistics are computed because if might past participle is a constant.

d: 2 cells (50,0%) have expected count less than 5. The minimum expected count is 1,18.

e: No statistics are computed because if might past participle is a constant.

The descriptive comments made on the overall use by means of Table 28 have been preferred as the basis of the study for this item.

4.3. Error Analysis and Results

We have derived and analysed the errors in two corpora; TICLE and JPICLE as those groups represent the use of the target variables in L2. We have not derived the errors in the other two corpora; LOCNESS and TUC since they belong to native use and we cannot judge the errors of native use. Nonetheless, the tables of each variable report the statistical values of LOCNESS and TUC. Following is Table 31 which illustrates an overall analyses obtained from each modal verb analysed within the hypothetical conditionals.

Table 31

Overall View of the Modal Verb Uses Across If and Main Clauses

Modal Verbs	If Clause	Main Clause
would	-	-
would + have + V*en*	+	+
could	+	+
could + have + V*en*	-	+
might	-	+
might + have + V*en*	+	+

Note. + = *The correct use;*

- = *Error committed.*

As Table 31 demonstrates, all the uses of "would + have + V*en*", "could", "might + have + V*en*" in both if clauses (IC) and main clauses (MC) of hypothetical conditional sentences have been used correct in syntactic form by Japanese and Turkish EFL learners. In addition, the modal verbs; "could + have + V*en*" and "might", in main clauses of hypothetical conditional structures have been used correct in syntactic form. As the symbol "-" refers error, the use of "would" in both if and main clause; "could + have + V*en* " and "might" in if clauses have revealed several errors committed syntactically in the observed hypothetical conditional sentences. All in all, most of the

variables used in hypothetical conditionals have been used correct in syntactic form. We have illustrated the quantitative results showing the errors committed by Japanese and Turkish EFL learners. Following are error-showing tables which illustrate the quantitative results in both frequency and ratio analyses and chi-square analyses across the two learner corpora examined by SPSS.

In the process of determining the errors; we did not analyse just the sentences incorrect in syntactic form, but also the immediately former and latter sentences. This procedure was applied in order to be sure whether the participant intended to form a hypothetical conditional sentence or any other type of conditionals since these three sentences displayed the context more clearly.

Errors observed in this study have been categorized in two groups; if clauses and main clauses of hypothetical conditionals. For each variable, the errors have been observed on the use of "would", "could" and "might" in Type 2 if-conditional sentences. Also, "would + have + V*en*", "could + have + V*en*" and "might + have + V*en*" in Type 3 if-conditional statements have been analysed in terms of errors. For each category of the variables, descriptive results and chi-square results have been calculated (both in if clause and main clause). Following is Table 32, which illustrates the total errors displayed in a division of overall Type 2 and 3 error categories with the observed errors in sub-categories; if and main clauses.

Table 32

Overall Frequency and Ratio Analysis of Target Variable Errors in Type 2 and 3 If and Main Clauses between two Learner Corpora.

Groups	Type 2 would, could, might				Type 3 would-, could-, might + have + V*en*				Total Error Tokens
	f IC	% IC	*f* MC	% MC	*f* IC	% IC	*f* MC	% MC	
TICLE	11	78.57	3	21.43	0	0	0	0	14
JPICLE	13	81.25	2	12.50	1	6.25	0	0	16

Note. IC= If Clause

MC: Main Clause

The analyses conducted on the errors committed by Japanese and Turkish EFL learners depict a similar picture of results to each other as clearly illustrated in Table 32. The errors committed have been extracted from the written productions of the informants in TICLE and JPICLE and these errors have been categorised under two types: Type 2 and 3. Type 2 includes the errors made by using "would", "could" and "might" both in if and main clauses of hypothetical conditional sentences composed of using if. Type 3 consists of the errors committed by using the following variables examined in this study; "would + have + V*en*", "could + have + V*en*" and "might + have + V*en*." The total number of the errors committed in Type 2 and Type 3 if-conditional sentences is 30; 14 of which committed by Turkish EFL learners and the rest 16 committed by Japanese EFL learners. Turkish EFL learner have made 14 errors in using Type 2 structures, the informants have not committed any error on using Type 3 if-conditional sentences. The Turkish participants have made 11 of 14 errors on using the variables "would", "could" and "might" in the if-clause segment of conditional sentences, which is equivalent to 78.57% proportionally. In TICLE data, it has been observed that three of the other identified errors were committed in the use of the variables; "would", "could" and "might", in the main-clause segment of hypothetical if-conditional sentences with a percentage of 21.43.

Japanese EFL learners have committed similar number of errors in using "would", "could" and "might" in the if-clause segment of if-conditional sentences analysed within the TICLE data, observed as 13 at stake; however, the frequency has been measured with a different result from JPICLE, JPICLE: 81.25%, TICLE: 78.57%. It is because of the reason that those Japanese EFL learners have made different errors than the Turkish EFL learners have made. In the main clause of if-conditional sentences, it has been observed that Japanese EFL learners have committed two errors referring to the percentage of 12.50. While Turkish EFL learners have not committed any errors in using "would + have + V*en*", "could + have + V*en*" and "might + have + V*en*" in Type 3 sentences, Japanese EFL learners have committed errors in using "could + have + V*en*" in if clause segment of if-conditional sentences. The frequency and the ratio observed are in if clause parts of if-conditional sentences are; *f*: one, 6.25%. It may be concluded that Japanese EFL learners have shown a tendency in committing errors in using Type 3 If-Clause structures, while the same has not been observed within the TICLE data. Overall analysis in Table 32, has demonstrated that Japanese EFL learners have a slightly more tendency in committing errors by composing Type 2 If-Clause

sentences than Turkish EFL learners have. All in all, the results are very similar to each other in both the JPICLE and TICLE data. Considering the error commitments extracted from the two interlanguage corpora productions, it may be attributed a sort of possible interlanguage property in having such close statistically close rate of committing errors.

4.3.1. "would" and "would + have + V*en*" Error Analyses and Results in Type 2 and 3 If and Main Clauses

In this subcategory, the errors made by using "would" in Type 2 and "would + have + V*en*" in Type 3 if and main clauses of hypothetical conditionals have been reported with the descriptive results and statistically chi-square results in addition to the error types with provided examples.

Table 33

Overall Frequency and Ratio Analysis of "would" and "would + have + V*en*" *Errors in Type 2 and 3 If and Main Clauses between two Learner Corpora.*

Groups	Type 2 would				Total Type 2 Errors	Type 3 would + have + V*en*				Total Type 3 Errors
	f IC	% IC	*f* MC	% MC		*f* IC	% IC	*f* MC	% MC	
TICLE	10	71.43	3	21.43	14	0	NA	0	NA	0
JPICLE	13	86.67	2	13.33	15	0	0	0	0	1

Note. *IC= If Clause*

MC: Main Clause

Table 33 displays at first stake that Turkish EFL learners have committed errors with "would" in Type 2 If-Clause sentences less than Japanese EFL learners have used the variable in concern. In the TICLE data it has been observed that "would" has been committed as error in the if clause part of the full Type 2 conditional sentences with the frequency of ten and regarding 71.43%. In addition, errors of "would" usage has been made in the main clause of full Type 2 conditional sentences with the frequency of three referring to the percentage of 21.43. Japanese EFL learners have made more errors in using "would" in if clause part of the sentences that the frequency has been calculated

as 13 referring to 86.67%. Japanese EFL learners have committed two errors in using "would" in the main clause part of the sentences referring to 13.33%. This result has paved us the way to make a clear comparison between the two groups; Japanese EFL learners conspicuously have made more errors in using "would" in the if clause part of the full Type 2 conditional sentences than Turkish EFL learners have made. However, the difference is slight in frequency but not so slight in the examined ratio values. Nonetheless, we may articulate the result that both the Japanese and Turkish interlanguage written productions have shown similar rate of values in committing errors by using "would" in if clause part of Type 2 conditional sentences. It is impressing that both the Japanese and Turkish EFL learners have not committed any errors on using "would + have + V*en*" in both if and main clause parts of the Type 3 conditional sentences. All in all, Table 33 has represented an example of the similar rate of committing errors observed in the written interlanguage productions of both JPICLE and TICLE data.

4.3.1.1. "would" Errors in Type 2 If Clauses

According to the chi-square test results as displayed in Table 34, Turkish and Japanese EFL learners have committed errors using "would" in if clauses of hypothetical conditionals in a similar sense that there is no statistically significance of difference between the two groups. That is to say, the interlanguages of Turkish and Japanese EFL learners' observed in the study demonstrate a similar proportion of error log, as indicated by Tables 32 and 33. As there are no errors in American Native and Turkish Native Corpora, it may be concluded that interlanguague development shows similar properties in terms of the rate of errors made by Japanese and Turkish EFL learners in their L2 English by using "would" in if clauses of hypothetical conditionals.

Table 34

Overall Pearson Chi-Square Results of "would" Errors in If clauses across four Corpora

	would errors in if clause	
Groups	c^2	**(p <.05)**
TICLE-JPICLE	0,039	0,844
TICLE-LOCNESS	8,639a	0,003
TICLE-TUC	9,029b	0,003
JPICLE-LOCNESS	7,963c	0,005
LOCNESS-TUC	d	NA

Note. *a= 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,57;*
b= 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,68;
c= 1 cells (25,0%) have expected count less than 5. The minimum expected count is 4,71;
d= No statistics are computed because weic is a constant.

In this procedure, "would" has been selected as the basis of categorization. There are several error types listed below extracted from both Turkish and Japanese subcorpora of ICLE (several of the extracts have been preferred as examples below). Following are the classified error types on which the sentences in correct in syntactic form have been illustrated by under two learner groups.

Error Type 1: Cases in which the participant cannot use the correct form of the main verb in if clause in terms of tense agreement of hypothetical conditionals.

Error Type 2: Cases in which the participant cannot provide the correct form of the auxiliary verb in if clause in terms of tense agreement of hypothetical conditionals.

Error Type 3: Cases in which the participants cannot use both the correct form of pronoun and auxiliary verb in if clause in terms of tense agreement of hypothetical conditionals.

Error Type 4: Cases in which the participants cannot provide the correct form of modal verb in if clause in terms of tense agreement of hypothetical conditionals.

Error Type 5: Cases in which the participant cannot prefer the correct voice (active/passive) of the structure.

Examples show the error-logged sentences in which the errors have been identified in **Bold** and *Italic* fonts. The codes of each participant have been provided

with the examples as well. For example, ICLE-TR-CUK-0078.1 stands for the explanation detailed as below and Appendices 1 and 2 provide the names of the codes of each institution in TICLE and JPICLE observed in this study;

ICLE : International Corpus of Learner English

TR : Turkey

CUK : Çukurova University

0078 : The participant whose essay was ordered as 78 in CUK.

1 : The category in CUK.

Error Type 1: Cases in which the participant cannot use the correct form of the main verb in if clause in terms of tense agreement of hypothetical conditionals.

a. TICLE

Example 23.

If I ***learn*** the language, but couldn't use it, would it be beneficial for my life?

(Retrieved from ICLE-TR-CUK-OO62.1)

In Example 23, the shown bold and italic verb '***learn***' should be corrected by using '***learnt***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 24.

If you ***look*** at the personal, intellectual and physical features of your candidate for marriage that would be better.

(Retrieved from ICLE-TR-CUK-O169.1)

In Example 24, the shown bold and italic verb '***look***' should be corrected by using ***'looked'***, because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure

Example 25.

> And I think that after the final decision, there will be remaining questions in the relatives' minds such as "did we do the right, were there any other opportunities, what would happen if ***say*** "No? "etc.
>
> (Retrieved from ICLE-TR-CUK-O159.1)

In Example 25, the shown bold and italic verb '***say***' should be corrected by using '***said***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 26.

> If we ***will learn*** only usefull or necessary things are used in our life, information would be more permanent in students' mind.
>
> (Retrieved from ICLE-TR-MER-0011.3)

Regarding Example 26, we can suggest the following structure as a correct form;

> If we learnt only useful or necessary things used in our lives, information would be more permanent in students' mind.

As it can be seen clearly from the suggested correct version above, there are more errors to be corrected, however, in this study, we have not focused on the other types of errors and just identified the errors relevant to the form in hypothetical conditionals.

Example 27.

> What would be happen if we **glass** all forests in?
>
> (Retrieved from ICLE-TR-CUK-OO67.1)

In Example 27, the participant coded as ICLE-TR-CUK-OO67.1 made errors both in if clause and main clause. However, '***glass***' should be corrected by '***glassed***' in this piece of example. This excerpted example has been processed in the category of "errors in main clause" as well.

Example 28.

> Japan is getting internationalization, so if Japanese **stady** English, Japan would become a cosmopolitan city.
>
> (Retrieved from ICLE-JP-SWU-0011.1.)

As illustrated in Example 28, the participant coded as ICLE-JP-SWU-0011.1. made an additional error other than preferring main verb in correct form in terms of tense, which is a spelling error. Instead of '***stady***', '***studied***' would be of correct preference of form.

Example 29.

> Secondly, we should consider the pains and suffering of the victims' family and how would they feel if a criminal who killed their loved ones **come** out from the jail eventually
>
> (Retrieved from ICLE-JP-TF-0001.1)

In Example 29, the shown bold and italic verb '***come***' should be corrected by using '***came***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 30.

> We would only have more people who are sad if we ***have*** the death penalty.
>
> (Retrieved from ICLE-JP-TF-0038.1.)

In Example 30, the shown bold and italic verb '***have***' should be corrected by using '***had***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 31.

> If you ***have*** a friend from Korea, and the other friend from France, what language would you use to communicate with them? Japanese? No way. They don't know Japanese. Body language? - Parhaps. The final possibility is English.
>
> (Retrieved from ICLE-JP-SWU-0026.3)

In Example 31, the shown bold and italic verb '***have'*** should be corrected by using '***had***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 32.

> However, I would get the English ability like native speakers if I ***live*** in surrounding that I must catch and speak English in person.
>
> (Retrieved from ICLE-JP-SWU-0031.4.)

In Example 32, the shown bold and italic verb '***live'*** should be corrected by using '***lived***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Error Type 2: Cases in which the participant cannot provide the correct form of the auxiliary verb in if clause in terms of tense agreement of hypothetical conditionals.

Example 33.

> To sum up; nobody would of course prefer the death if there ***is*** another opportunity and it is the most difficult decision for everybody because our religion, cultural structure and feelings are not appropriate for this situation.
>
> (Retrieved from ICLE-TR-CUK-O174.1)

In Example 33, the shown bold and italic auxiliary verb '***is'*** should be corrected by using '***was/were***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 34.

> If both men and women ***are***thought as people without considering sex difference, the inequality between them would not occur.
>
> (Retrieved from ICLE-TR-CUK-O120.1)

In Example 34, the shown bold and italic auxiliary verb '***are'*** should be corrected by using '***were***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 35.

If the woman in such situation ***is not*** assisted , she would have a terrible death, being maintained by machines.

(Retrieved from ICLE-TR-CUK-O164.1)

In Example 35, the shown bold and italic auxiliary verb '***is not***' should be corrected by using '***was/were not***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

b. JPICLE

Example 36.

Third, would less people commit a crime if there **is** the death penalty? The answer is, no. There is a data showing that the criminal acts didn't increase when they stopped having the death penalty in Sweden.

(Retrieved from ICLE-JP-TF-0038.1)

In Example 36, the shown bold and italic auxiliary verb '***is***' should be corrected by using '***was/were***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 37.

Moreover, if a different name ***is*** usual, such an idea would never occur.

(Retrieved from ICLE-JP-TM-0003.1)

In Example 37, the shown bold and italic auxiliary verb '***was***' should be corrected by using '***was/were***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Error Type 3: Cases in which the participants cannot use both the correct form of pronoun and auxiliary verb in if clause in terms of tense agreement of hypothetical conditionals.

Example 38.

> However, if ***you are*** asked where you'd like to live, in big cities or in suburban area, I would immediately answer that I would like to live in Suburban area like Chiba.
>
> (Retrieved from ICLE-JP-AI-002.1)

In Example 38, the shown bold and italic auxiliary verb '***you are***' should be corrected by using '***I was/were***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure. In the main clause of the sentence, the participant used the pronoun; I. This sheds light on the main pronoun choice of the participant's.

Error Type 4: Cases in which the participants cannot provide the correct form of modal verb in if clause in terms of tense agreement of hypothetical conditionals.

Example 39.

> If you ***can*** choose where you want to live, which would you choose? The city or the country?
>
> (Retrieved from IC-JP-SWU-0016.3.)

In Example 39, the shown bold and italic modal verb '***can***' should be corrected by using '***could***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 40.

> But, if we **can** communicate equally with them, we wouldn't assume ourselves inferior to them.
>
> (Retrieved from ICLE-JP-TF-0003.1.)

In Example 40, the situation is the same with that of the Example 39's. The shown bold and italic modal verb '***can***' should be corrected by using '***could***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

4.3.1.2. "would" Errors in Type 2 Main Clauses

With the light of Table 35, it has been observed that overall chi-square result of would errors committed in main clause among all the groups cannot be calculated as the (f) value is observed 5 and less than 5 in 20% and more than 20% of the cells, chi-square significance test could not be examined. Hence, the descriptive comments made on the overall use Table 33 have been preferred as the basis of the study for this item.

Table 35

Overall Pearson Chi-Square Results of "would" Errors in Main Clauses across four Corpora

	"would" errors in main clause	
Groups	c^2	**(p <.05)**
TICLE-JPICLE	0,565a	0,452
TICLE-LOCNESS	2,544b	0,111
TICLE-TUC	2,660c	0,103
JPICLE-LOCNESS	1,297d	0,255
LOCNESS-TUC	e	NA

Note. a= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 2,17;
b= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 1,37;
c= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 1,40;
d= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 0,78;
e= No statistics are computed because wemc is a constant.

Errors in this category have been categorized made in main clauses of hypothetical conditionals. In this procedure, "would" was selected as the basis of categorization. There are several error types listed below extracted from the both Turkish and Japanese subcorpora of ICLE. Some of the extracts have been preferred as examples below; the errors have been identified in **Bold** and *Italic* fonts.

Error Type 1: Cases in which the participant cannot use the correct form of the main verb in if clause in terms of tense agreement of hypothetical conditionals.

a. TICLE

Example 41.

Some babies are born out of wedlock and if the women's parents do not accept theywould***left*** their own babies.

(Retrieved from ICLE-TR-CUK-OO99.1)

In Example 41, the shown bold and italic main verb ***'left'*** should be corrected by using '***leave***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

Example 42.

What ***would be happen*** if we glass all forests in?

(Retrieved from ICLE-TR-CUK-OO67.1)

In Example 42, the shown bold and italic main verb '***would be happen***' should be corrected by using '***would happen***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

b. JPICLE

Example 43.

If we were in trouble, it would ***influenced*** them

(Retrieved from ICLE-JP-SWU-0015.4.)

In Example 43, the shown bold and italic main verb '***influenced***' should be corrected by using '***influence***', because the informant is thought to be intended to form a Type 2 If-Conditional statement referring to hypothetical conditional structure.

4.3.2. "could" and "could + have + V*en*" Error Analyses and Results in Type 2 and 3 If and Main Clauses

This category serves with two sub-categories; the errors made by using "could" in if and main clauses of Type 2 and the errors made by using "could + have V*en*" in if and main clauses of Type 3 hypothetical conditional statements.

It has been demonstrated in Table 36 that both Japanese and Turkish EFL learners have not committed any errors in using "could" in Type 2 if and main clauses of the examined conditional sentences. In addition, Turkish EFL learners have not made any errors on using "could + have + V*en*" in Type 3 if and main clauses of full conditional sentences analysed in this study.

Table 36

Overall Frequency and Ratio Analysis of "could" and "could + have + V*en*" *Errors in Type 2 and 3 If and Main clauses between two Learner Corpora.*

Groups	Type 2 could				Total Type 2 Errors	Type 3 could + have + V*en*				Total Type 3 Errors
	f IC	% IC	*f* MC	% MC		*f* IC	% IC	*f* MC	% MC	
TICLE	0	0	0	0	14	0	NA	0	NA	0
JPICLE	0	0	0	0	15	1	100	0	0	1

Note. *IC= If Clause*

MC: Main Clause

Table 36 has shown that Japanese EFL learners have committed one error in frequency. The frequency has been statistically calculated as equal to 100% in the total proportion of the errors in Type 3 clauses. That is to say, Japanese EFL learners have not committed any other errors in composing Type 3 conditional sentences by using the following structures; "would + have + V*en*" and "might + have + V*en*." Considering the overall frequency analysis results, it may be concluded that both Japanese and Turkish EFL learners have demonstrated a stable use of the variable in concern in their written argumentative essays in English. It may not be attributed as a possible interlanguage property as the calculated value is very low. However, to rank the order, it can be said

that Japanese and Turkish EFL learners have shared in common on using particularly "could" and then "could + have + V*en*."

4.3.2.1. "could + have + V*en*" Errors in Type 3 If Clauses

In this subcategory the errors made by using "could + have + V*en*" in if clauses of hypothetical conditionals have been reported with the chi-square results in addition to the error types with provided examples.

As it is illustrated in Table 37, the overall chi-square result has shown that if could past participle errors in if clause cannot be calculated as the (f) value is observed 5 and less than 5 in 20% and more than 20% of the cells, chi-square significance test could not be examined. Hence, frequency Table 36 has been considered to comment on descriptively.

Table 37

Overall Pearson Chi-Square results of '"could + have + Ven" Errors in If Clauses across four Corpora

	"could + have + Ven" error in if clause	
Groups	c^2	**(p <.05)**
TICLE-JPICLE	0,769a	0,380
TICLE-LOCNESS	b	NA
TICLE-TUC	c	NA
JPICLE-LOCNESS	d	NA
LOCNESS-TUC	e	NA

Note. a= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 0,43;
b= No statistics are computed because cpp eic is a constant;
c= No statistics are computed because cpp eic is a constant.;
d= No statistics are computed because cpp eic is a constant;
e=No statistics are computed because cpp eic is a constant.

Below is an illustrated example of the error committed by the Japanese EFL learner participant. According to the chi-square results illustrated in Table 37, the comparisons between the pairs TICLE-LOCNESS, TICLE-TUC, JPICLE-LOCNESS

and LOCNESS-TUC have revealed no statistics because could + have + V*en* is a constant value.

Error Type 1: Cases in which the participant cannot use the correct form of the main verb in if clause in terms of tense agreement of hypothetical conditionals.

Example 44.

> Her mother regretted that if she ***could know*** that she had a problem with her heart earlier then her daughter could have recovered earlier.
>
> (Retrieved from **ICLE-JP-HI-0004.1.**

Following is the suggested correct structure of the illustrated Example 44;

> Her mother regretted that if she *could have known* that she *had had* a problem with her *heart,* then her daughter could have recovered.

We have suggested the correct form of the error-logged sentence by taking into consideration the former and latter sentences that the participant (ICLE-JP-HI-0004.1) produced as shown in Example 44. These former and latter sentences have provided us the judgement that the informant mentioned about a hypothetical event referring to past time. In addition, the last words (underlined at the example) of the participant's have empowered our judgement on categorising the sentence and correcting it under the category of Type 3 conditional sentence.

4.3.3. "might" and "might + have + V*en*" Error Analyses and Results in Type 2 and 3 If and Main Clauses

This category serves with two sub-categories; the errors made by using "might" in if and main clauses of Type 2 conditional statements and using "might + have + V*en*" in if and main clauses of Type 3 conditional statements.

Table 38

Overall Frequency and Ratio Analysis of "might" and "might + have + V*en*" *Errors in Type 2 and 3 If and Main Clauses between two Learner Corpora*

Groups	Type 2 might				Total Type 2	Type 3 might + have + V*en*				Total Type 3
	f IC	% IC	*f* MC	% MC	Errors	*f* IC	% IC	*f* MC	% MC	Errors
TICLE	1	7.14	0	0	14	0	NA	0	NA	0
JPICLE	0	0	0	0	15	0	0	0	0	1

Note. *IC= If Clause*

MC: Main Clause

Considering the statistical results illustrated in Table 38, Turkish EFL learners have committed one error in frequency referring a percentage of 7.14% in Type 2 in using "might." However, no other category has been fulfilled by the errors made by the other group. Both groups have indicated that there have been no errors observed in any other category illustrated at Table 38. It may be attributed as a possible interlanguage property that both learner corpora have shared a comment point on using "might and "might + have + V*en.*" On the other hand, the shared similar use by Japanese and Turkish EFL learners may be regarded as a sort of evidence that both groups have used these variables correct in syntactic form to the extent that native speakers use.

4.3.3.1. "might" Errors in Type 2 If Clauses

In this subcategory the errors made by using 'might' in if clauses of hypothetical conditionals have been reported with the chi-square results in addition to the error types provided with examples.

As Table 39 displays, the overall chi-square result of "might" errors in if clause cannot be calculated as the (f) value is observed 5 and less than 5 in 20% and more than 20% of the cells, chi-square significance test could not be examined. For the other comparisons no statistics have been provided, as this variable is a constant. Hence, frequency Table 38 has been considered to comment descriptively.

Table 39

Overall Pearson Chi-Square Results of "might" Errors in If Clauses across four Corpora

	"might" errors in if clause	
Groups	c^2	**(p <.05)**
TICLE-JPICLE	1,306a	0,253
TICLE-LOCNESS	b	NA
TICLE-TUC	c	NA
JPICLE-LOCNESS	d	NA
LOCNESS-TUC	e	NA

Note. *a= 2 cells (50,0%) have expected count less than 5. The minimum expected count is 0,43.*

b= No statistics are computed because memc is a constant.

c= No statistics are computed because memc is a constant.

d= No statistics are computed because memc is a constant.

e= No statistics are computed because memc is a constant.

Below is an illustrated example of the error committed by the Turkish EFL learner. The error has been categorised under the title of 'Error Type 6', which employs the errors stemmed from using wrong voice of the statement.

Error Type 5: Cases in which the participant cannot prefer the correct voice (active/passive) of the structure.

Example 45.

> Hence, the couple, having preconceived that their baby is liable to live through such a awe-striking distortion of physical disturbation must take into consideration probable troubles the baby ***might be faced*** with if born.
>
> (Retrieved from ICLE-TR-MER-0155.1

Example 45 illustrates a compound sentence consisting of if clause (Type 2) and a defining sentence. The defining sentence has been remarked as the following;

> Hence, the couple, having preconceived that their baby is liable to live through such a awe-striking distortion of physical disturbation must take into consideration probable troubles...

As named defining sentence illustrated above, the sentence has two errors; one of which is committed on preferring the correct form of the article (…. such a awe-striking). The participant has preferred "a" instead of using "an" before the adjective clause '… awe striking distortion…' Another error has been logged as a spelling error which has been underlined at the example shown above. The participant would write "distribution" instead of writing "disturbation." However, these errors have not been a criteria in processing the errors, because we have indicated the errors committed in the if and main clauses in this study. Another compound sentence, named as if clause, has been displayed below;

> … the baby ***might be faced*** with if born.

We have assumed that the participant has reduced linguistic construction in Example 45. That is, the participant has not written 'she / he was born' just after 'if.' The reduced construction is used in hypothetical conditional statements and the concern here is a proof of judging this sentence as a hypothetical conditional statement. Another proof is that the writer is assuming a situation in the defining sentence of the whole example. The extracted example has been regarded at the category of Type 2 since the first compound of the whole sentence has been composed of using Present Simple Tense and there is the reduction of subject pronoun and auxiliary verb. The main clause of the conditional statement has been preferred as the selection. In that point, we have decided that the error indicated in bold and italic fonts needs to be corrected as; … might face …, since this patterning needs active voice in the statement.

CHAPTER V

CONCLUSION AND DISCUSSION

5.1. Introduction

This study attempts to examine the use of hypothetical conditionals across Japanese EFL learners, Turkish EFL learners, American and Turkish Native speakers in order to represent the range of using the target language pattern by comparing each selected group. This range of using hypothetical conditionals has represented the common tendencies employed in English Interlangauge productions of Turkish and Japanese EFL learners' with a comparison to the use in native norms of English and Turkish.

5.2. Conclusions

1. What is the distribution of the hypothetical conditional sentences from most frequently use to the least within native speakers of English (American university students) and Turkish (Turkish university students) and non-native speakers of English (Japanese &Turkish university students)?

The distribution of the hypothetical conditionals has been referred as the modals used in Type 2 if clauses; "would", "could", "might" and in Type 3 if clauses; "would+ have + V*en*", "could+ have + V*en*" and "might+ have + V*en*." In the light of the results obtained from the analyses, it is clear to put in an order within the Type 2 and 3 If-conditional statements from the most frequently use to the least in groups as;

Within the use analysed in Type 2 if-conditional statements, it has been observed that American university students use hypothetical conditionals in their native language more than any other group observed in the study. Turkish university students use hypothetical conditionals in their mother tongue less than American university students; however, more than Japanese and Turkish EFL learners. The use of hypothetical conditionals in English and Turkish has indicated the common tendency in native usage. However, it has been observed that hypothetical conditionals have been

used in English native language more than in Turkish mother tongue. Nonetheless, the comparison within the native uses has represented a shared usage-taking place in native norms.

The third group on the use of Type 2 if-conditional statements has been observed as Japanese EFL learners. This group may have represented a character of English Interlanguage by not having similar use to that of American native use. Also, hypothetical conditionals have been preferred more common than the other group representing English Interlanguage, namely Turkish EFL learners. That is to say, Type 2 If-conditional statements have been used in the order mostly to the least as; American native, Turkish mother tongue, Japanese Interlanguage and Turkish Interlanguage use.

Within the use analysed in Type 3 If-conditional statements of hypothetical conditionals, it has been observed that Turkish university students have employed the most frequent use of all the groups. The second most usage has been examined in the written essays of American university students in their native language productions. In the investigated Type 3 If-Conditional statements, the third and fourth groups have been employed by the examined two-interlanguage productions in an order of Japanese EFL learners and Turkish EFL learners. All in all, in both Types (2 and 3) of If-conditional statements referring to hypothetical conditionals, the native use is ranked as the most employed use, but the order within it has shown a difference. In Type 2 if-conditionals, American university students have employed the most usage while Turkish university students have employed the most frequent use in Type 3 if-conditional statements. In the learner groups; JPICLE and TICLE, it has been observed that Japanese EFL learners use both types more commonly than the Turkish EFL learners.

2. Do advanced Japanese and Turkish EFL learners use Type 2 and 3 if-conditional statements referring to hypothetical conditionals in the same extent as native speakers of American university students?

In both Type 2 and 3 if-conditional statements, it has been observed that Japanese and Turkish EFL learners have not shown a significant similarity. However, Japanese EFL learners have used the variables examined in Type 2 and 3 If-conditional statements more than the Turkish ELF learners have used. As the observed use is higher than the use analysed in two interlanguage productions, Japanese EFL learners have

more tendency than Turkish EFL learners in using Type 2 and 3 if-conditionals to that of investigated American use. That is to say, Turkish EFL learners have been observed as the less close group in using Type 2 and 3 If-Conditional statements to the use examined by American university students.

3. Which of the modal verb forms investigated in hypothetical conditionals are most commonly marked among the argumentative written essays in L2 English by Japanese and Turkish non-native speakers of English, in L1 English by American university students and in L1 Turkish by Turkish university students?

The investigated use of modal verbs; "would", "could" and "might" in Type 2 If-conditional statements referring the use in hypothetical if conditionals, have shown a higher significant use to that of the investigated use of modal verbs; "would + have + V*en*", "could + have + V*en*" and "might + have + V*en*", in Type 3 if-conditional statements referring the use in hypothetical if conditionals. Each modal verb use under the title of Type 2 and 3 have been presented below by the light of the results quantitative in character illustrated in Chapter IV.

All in the observed groups and Types, the most frequent use has been discovered in the use of "would" by American native speakers in Type 2 if-conditional statements. Japanese EFL learners have employed the next most frequent uses in the use of "could" and "might" in Type 2 if-conditional statements. In Type 3 if-conditional statements, Turkish university students have employed the most frequent use of the modal verb "would + have + V*en*" and "could + have + V*en*" in their L1 written performances. The last and the least frequently observed modal verb form, "might + have + V*en*" has been employed by Japanese EFL learners.

4. If there is a significant difference in the use of hypothetical conditionals used by Japanese & Turkish EFL learners across the American Native use and Turkish L1 use, is this due to

a) mother tongue effect?
b) the property of interlanguage?

It has been observed that several of the items represent significant difference across groups. We may not conclude with reference to hypothetical conditionals at once; however, we may explain the reasons by taking separate hypothetical conditional modal verbs into consideration and may come to a conclusion by the light of the examined use of separate language patterns. There is significant difference between the pair of groups; TICLE-LOCNESS; JPICLE-LOCNESS; JPICLE-TUC; LOCNESS-TUC for the observed use of "would." Since there is no significant difference observed in the pair of two English Interlanguages (TICLE-JPICLE), it might show property of interlanguage especially compared to the native norms that both groups display significant difference across the use both in American native. As there is no significant difference between the use by Turkish-speaking EFL learners and their mother tongue use, Turkish-speaking EFL learners may use "would" in the target language with representing mother tongue effect.

There has been observed another significant difference between the pairs; TICLE-JPICLE, TICLE-LOCNESS and LOCNESS-TUC. As there is significant difference within the groups of Turkish- and Japanese-speaking EFL learners, there may not be representing any property of English interlanguage. This inference may be attributed to the native use in English and mother tongue use in Turkish. On one hand, Both English interlanguage groups tend to use "could" without any significant difference when compared to the mother tongue use in Turkish. On the other hand, as there is significant difference between the use in Turkish-speaking EFL learners and American native participants, Turkish-speaking EFL learners tend to be effected by their mother tongue use, which does not indicate any significant difference between TICLE-TUC. Japanese-speaking EFL learners show no significant difference when compared to the American native use and Turkish mother tongue use. We may not attribute any property of interlanguage or mother tongue effect for this group.

Chi-square test results have not indicated any calculation of significant difference on the uses of "might, could + have + V*en*" and "might + have + V*en*", but would + have + V*en*. There is not any significant difference in the use of "would + have + Ven" between the groups TICLE-JPICLE. The same has been observed for the variable "would" and it has been attributed to the property of interlanguage since the two English interlanguage groups represent similarity. For the current variable, "would + have + V*en*", we may infer that Turkish-and Japanese-speaking EFL learners tend to show property of interlanguge, however, Turkish-speaking EFL learners do not use this

hypothetical modal in the same extent that Japanese-speaking EFL learners use it without any significant difference compared to the American native use. To observe the reason behind it, we have examined the Turkish-speaking EFL learners' use with that of Turkish mother tongue use, but there was not any similarity that it has not illustrated any mother tongue effect. This has strengthen the view that there may be the property of interlanguage observed for the related use of "would + have + V*en*."

Consequently, we may reach a conclusion by the light of the observed significant differences that the two English interlanguages, namely English interlanguages of Turkish-and Japanese-speaking EFL learners, may tend to represent property of interlanguage in the use of hypothetical conditionals.

5. To what extent do Japanese & Turkish non-native speakers of English use hypothetical conditionals correct in syntactic form?

The survey of the learner subcorpora –TICLE and JPICLE- has yielded some important findings in identifying the use of hypothetical conditional forms correct in form. As Ke (2004, p. 22) points out "… Chinese learners have a tendency to complicate simple conditionals; however, simplify the complex conditionals" Japanese and Turkish EFL learners have the similar tendency in using hypothetical conditionals. To illustrate the tendency in separate forms of if clause and main clause; we may compare Japanese and Turkish EFL learners in a division of tendency towards committing errors. As we have observed the if- and main-clauses of each hypothetical conditional modal (would, could, might, would + have + V*en* , could + have + V*en* and might + have + V*en*), the comparison is presented below by each hypothetical conditional modal.

Japanese and Turkish speaking EFL learners have used three variables correct in form both in if- and main clauses of "could", "would + have + V*en*" and "might + have + V*en*." On the other hand, the learners in concern have used "might" and "could + have + V*en*" in main clauses correct in syntactic form. There remains three differences in using the hypothetical conditional modals correct in form; (1) Japanese speaking EFL learners have had a slightly more tendency using "might" in if clauses correct in form than Turkish speaking EFL learners; on the other hand, (2) Turkish speaking EFL learners have been observed to be more prone to use "could + have + V*en*" in if clauses

correct in syntactic form than Japanese speaking EFL learners. Final but the most important remark has been put on “would” which has been presented with the highest usage result (3), has shown that Japanese speaking EFL learners are slightly more prone to use “would” both in if- and main clauses of hypothetical conditionals correct in syntactic form than Turkish speaking EFL learners.

Consequently, it may be concluded that Japanese and Turkish EFL learners are prone to use most of the hypothetical conditional modals correct in syntactic form, but Japanese speaking EFL learners differ from Turkish speaking EFL learners on the aspect of using the structure in concern slightly closer to the native norm. We have observed this difference as a slight difference according to the frequency result; however, it could be observed in a wider corpus in size in order to have chi-square results.

5.3. Implications for English Language Teaching

In this body of research, we have examined the use of hypothetical conditionals. There have occurred two main reasons to apply in the English Language Teaching methodology that one of them is the difficulty of learning and producing hypothetical conditionals in the target language as suggested by Ke (2004, p. 9). The second reason the current study has revealed is that EFL learners coming from Japanese and Turkish mother tongue backgrounds underuse the most commonly used hypothetical conditional modals when compared to the native usage in English. There may not be enough evidence to prove any mother tongue interference on learners’ use of hypothetical conditionals, but may be sufficiently represented by a possible interlanguage property on learners’ use of hypothetical conditionals.

Corpus has been a useful tool over decades in identifying the productions of different language patterns and it has been useful in this study on representing the use of hypothetical conditionals by English L2 learners. Corpus determines the way for language teachers that they can inspire teacher trainers to develop methods on specific language patterns. Thompson (2001) refers corpus as a useful tool in identifying the necessary language patterns to be modified in classroom teaching. The main pedagogical conclusion for ELT specialists to be drawn in the process of preparing teaching materials or methodologies is that Japanese and Turkish EFL learners need to be exposed to a wider range of registers and a more extensive training in producing the

hypothetical conditionals. In this process, it would be worthy of investigating all the learning process that the Japanese and Turkish learners are exposed to on learning hypothetical conditionals. It would be put into practice by examining the classroom materials such as textbooks which language teachers use to teach the related pattern. For Turkish learners, there is an important point to take into consideration. This is the semantic difficulty they experience in composing hypothetical conditional sentences in English syntactically. This stems from the reason that a conditional sentence can mean both real and unreal because of the suffixation system of Turkish. Hence, it would be a very strong claim to judge a sentence whether it is real or unreal by just following the syntactic rules in Turkish. There remains one solution to classify a conditional sentence into real or unreal, the reader must be aware of the context in which the semantic attempt stands for the purpose of constructing a hypothetical sentence (Hengirmen, 1998).

5.4. Suggestions for Further Studies

In this study, the use of hypothetical conditional markers “would”, “could”, “might”, “would + have + V*en*”, “could + have + V*en*” and “might + have + V*en*” has been merely discussed. We did not further analyse other subordinators suggested by Quirk (1985) “should”, “was”, “were”, “inversions”, “unless”, “in case”, “just so”, “if only”, “.. wish”, “on/ providing/ assuming . . . (that)”, and many other semantic variants. All the other forms of variants arranging a hypothetical conditional should be investigated. This may benefit of our understanding hypothetical conditionals in English and the counterparts of these variants in Turkish. Additionally, as we did not employ a Japanese mother tongue corpus, we could not have compared the use of these hypothetical conditional markers across the other corpora. It would be very beneficial if we could have compared the use specifically with the use in Turkish mother tongue. Another requirement was in regard to the researcher of the study. If the researcher of the study were bilingual in both Japanese and Turkish or the study were employed by two researchers of which mother tongues were Turkish and Japanese, it would be a great wise to analyse the selected language pattern. By bearing these limitations in mind, we could propose solutions to the difficulties that EFL learners might have in the course of being proficient in EFL by following Altenberg and Tapper’s (1998, p. 93) suggestion

“The ICLE corpus offers a very promising research for future research” in second language acquisition studies.

REFERENCES

Aarts, J. & Granger, S. (1998). Tag sequences in learner corpora: a key to interlanguage grammar and discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 132-142). London & New York: Addison Wesley Longman.

Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hong & S. Petch- Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55-76). Amsterdam/Philadelphia: John Benjamins.

Aijmer, K. (2008). Parallel and comparable corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook,* Volume 1 (pp. 275-291). Berlin and New York: Walter de Gruyter.

Allen, S., Ozyurek, A., Kita, S., Brown, A., Furman, R. & Ishizuka, T. (2007). Language- specific and universal influences in children's syntactic packaging of Manner and Path: A Comparison of English, Japanese, and Turkish. *Cognition,* 102, 16-48.

Altenberg, B. & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80-93). London & New York: Addison Wesley Longman.

Atkins, B. T. S., Clear, J. & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing.* 7 (1), 1-16.

Baker, P., Hardie, A. & McEnery, T. (2006). *A Glossary of corpus linguistics.* Edinburgh: Edinburgh University Press Ltd.

Banguoğlu, T. (1990). *Türkçenin grameri*. Ankara: Türk Tarih Kurumu Basımevi.

Baobao, C. Danielsson, P. & Teubert, W. (2005). Chinese- English translation database: extracting units of translation from parallel texts, In G. Barnbrook, P. Danielsson, M. Mahlberg (Eds.), *Meaningful texts: The extraction of semantic information from monolingual and multilingual corpora,* (pp. 103-132). London: Continuum.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics. investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D. Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English.* London: Longman.

Bikeliene, L. (2008). Resultive connectors in advanced Lithauan learner's English writing. *Kalbotyra,* 59 (3), 30-37.

Bondi, M. & Scott, M. (Eds.) (2010). *Keyness in texts.* Amsterdam: Benjamins

Campoy, M. C., Bellés-Fortuno B., & Gea-Valor, M. B. (Eds.) (2010). *Corpus-based approaches to English language teaching,* London & New York: Continuum, Corpus and Discourse.

Can, C. (2010). A Turkish learner corpus of English in second language acquisition studies: TICLE as a subcorpus of ICLE. *Language Journal,* Volume. 144, 16-34.

Carter, R. A. & McCarthy, M. J. (2006). *Cambridge grammar of English*: *A comprehensive guide to spoken and written English grammar and usage*. Cambridge: Cambridge University Press.

Chambers, A. (2010). What is data-driven learning? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 345-358). London: Routledge.

Cheng, M. (2002). The standoff-what is unsaid? A pragmatic analysis of the conditional marker. *"If" discourse studies,* 13(3), 312.

Chomsky, N. (2006). *Language and mind. (3rd Ed.)* Cambridge, United Kingdom: Cambridge University Press.

Cılız, M. (2010). *Building Up A learner corpus through creative nonfiction prose: An experimental research*. Unpublished master thesis, University of Gaziantep, Gaziantep.

Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review,* 59(3), 393-423. Retrieved from January 19, 2011 http://www.lextutor.ca/cv/pdf/learner_corpus.pdf

Connor, U., Precht, K. & Upton, T. (2002). Business English: Learner data from Belgium, Finland and the U.S. In Granger, S. Hung, J. & Petch-Tyson, S. (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 175-194). Amsterdam: John Benjamins.

Dancygier, B. (1998). *Conditionals and prediction*. Cambridge: Cambridge University Press.

Dancygier, B. & Sweetser, E. (2005). *Mental spaces in grammar: conditional constructions.* Cambridge: Cambridge University Press.

De Cock, S. & Granger, S. (2004). High frequency words: the bête noire of lexicographers and learners alike. A close look at the verb 'make' in five monolingual learners dictionaries of English. *Paper Presented at the XIth EURALEX International Congress, Lorient.*

Dias, G. M. & Pereira Lopes, J. G. S. (2005). Extracting concepts from dynamic legislative text collections. In G. Barnbrook, P. Danielsson, M. Mahlberg (Eds.), *Meaningful Texts. The extraction of semantic information from monolingual and multilingual corpora* (pp. 5-16). London: Continuum.

Diessel, H. (2001). The ordering distribution of main and adverbial clauses: a typological study. *Language* 7(2), 433-455.

Dodd, B. (Ed.) (2000). *Working with German corpora*. Birmingham: The University of Birmingham Press.

Eckkrammer, E. M. (2005). A diachronic genre corpus: problems and finding from the DIALAYMED-Corpus (diachronic multilingual corpus of layman-oriented medical texts). In G. Barnbrook, P. Danielsson, M. Mahlberg (Eds.), *Meaningful Texts. The extraction of semantic information from monolingual and multilingual corpora* (pp. 17-30). London: Continuum.

Fillmore, C. (1992). Corpus linguistics or computer-aided armchair linguistics. In J. Svartvik (Ed.), Proceedings of Nobel Symposium: 82. 35-60. *Directions in Corpus linguistics*. Berlin - New York: Mouton de Gruyter.

Freddie, M. (2005). Arguing linguistics: corpus investigation of one functional variety of academic discourse. *Elsevier: Journal of English for Academic Purposes,* 4. 5-26.

Gabrielatos, C. (2007). If-conditionals as modal colligations: A corpus-based investigation. In M. Davies, P. Rayson, S. Hunston & P. Danielsson (Eds.), *Proceedings of the corpus linguistics conference: Corpus linguistics.* Birmingham: University of Birmingham. Retrieved February 18, 2011 from [http://corpus.bham.ac.uk/corplingproceedings07/paper/256_Paper.pdf]

Gass, S. M. & Selinker, L. (2008). *Second language acquisition, an introductory course* (3rd ed.). New York: Routledge Press, Taylor & Francis.

Ghadessy, M., Henry, A. & Roseberry, R. L. (Eds.) (2001). *Small corpus studies and ELT theory and practice*. Amsterdam: John Benjamins Publishing Company.

Gilquin, G. & Granger, S. (2010). How can data-driven learning be used in language teaching?. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 359-70). London: Routledge.

Göknel, Y. (2010). *Turkish grammar, transformational and generative contrastive*, Istanbul: Vesta Press.

Granger, S. (1996a). *Romance words in English: from history to pedagogy*. Paper presented at the Words. *Proceedings of an International Symposium*, Stockholm. Retrieved December 28, 2010 from http://sites.uclouvain.be/cecl/archives/Granger_1996_Romance_words.pdf

Granger, S. (1996b). From CA to CIA and back: an integrated approach to comuterized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.

Granger, S. (1997). On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared. In J. Aarts, I. de Mönnink & H. Wekker (Eds.), *Studies in English language and teaching* (pp. 185-98). Amsterdam & Atlanta: Rodopi.

Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). London & New York: Addison Wesley Longman.

Granger, S. (1999). Use of tenses by advanced EFL learners: evidence from an error-tagged computer corpus. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora. Studies in honour of Stig Johansson* (pp. 191-202). Amsterdam & Atlanta: Rodopi.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hong & S. Petch- Tyson (Eds.) *Computer learner corpora, second language acquisition and foreign language teaching.* (pp. 3-33), Amsterdam/Philadelphia: John Benjamins.

Granger, S. (2004). Computer learner corpus research: Current status and future prospects. In U. Connor & T.A. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective.* (pp. 123-145) Amsterdam: Rodopi.

Granger, S. (2008). Learner corpora in foreign language education. In N. Van Deusen-Scholl & N. Hornberger (Eds.), *Encyclopedia of language and education. Second and foreign language education, Springer,* Vol.4, 337-351.

Granger, S. (2009). The Contribution of Learner Corpora to Second Language Acquisition and Foreign language Teaching. In K. Aijmer, (Ed.), *Corpora and language Teaching*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Granger, S., Dagneux, E. & Meunier, F. (Eds.) (2002). *International corpus of learner english.* Version 1 (CD-ROM & Handbook). Presses Universitaires de Louvain

Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. (2009). *The International Corpus of Learner English. Version 2.* Handbook and CD-ROM. Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S. & Meunier, F. (2008). Phraseology in language learning and teaching. Where to from here? In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247-52). Amsterdam & Philadelphia: John Benjamins.

Granger, S. & Paquot, M. (2009). Lexical verbs in academic discourse: a corpus-driven study of learner use. In M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic writing: at the interface of corpus and discourse* (pp. 193-214). London & New York: Continuum.

Granger, S. & Petch-Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes: journal of English as an international and intranational language*, 15(1), 17-27.

Guo, X. (2006). *Verbs in the written english of chinese learners: A Corpus-based comparison between Non-native speakers and native speakers.* Unpublished doctoral dissertation, University of Birmingham, Birmingham.

Halliday, M.A.K. (2007). *Language and education.* London: Continuum.

Harris, R. A. (1993). *The linguistics wars.* Oxford: Oxford University Press.

Hengirmen, M. (1998). *Türkçe temel dilbilgisi*. Ankara: Engin Yayınları.

Hunston, S. & Francis, G. (2000). *Pattern grammar a corpus-driven approach to the lexical grammar of English.* Amsterdam: John Benjamins.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Kabak, B. & Schiering, R. (2004). A corpus study on the distribution of Turkish ise and its clitic form. *Turkic Languages*, 8, 232-244.

Kaiser, S., Ichikawa, Y., Kobayashi, N. & Yamamoto, H. (2001). *Japanese: A*

Comprehensive Grammar. London: Routledge.

Kaszubski, P. (1998). Enchancing a writing textbook: a national perspective. In S. Granger (Ed.), *Learner English on computer* (pp. 80-93). London & New York: Addison Wesley Longman.

Ke, Y. (2004*). Form- function mapping in the acquisition of if-conditionals: A corpus based study*. Retrieved on January 10, 2011 from http://www.bartleby.com/64/C001/061.htm

Kerslake, C. (2003). A new look at conditional constructions in Turkish. In A.S. Özsoy (Ed.), *Studies in Turkish linguistics* (pp. 215–226). İstanbul: Boğaziçi University Press.

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

Kuno, S. (1978). Japanese: A characteristic OV language. In W.P. Lehmann (Ed.) *Syntactic typology: Studies in the phenomenology of language* (pp. 57–138). Austin: University of Texas Press.

Kurzon, D. (2008). ʻAgoʼ and its grammatical status in English and other languages .In D. Kurzon and S. Adler (Eds.), *Adpositions: Pragmatic, semantic and syntactic perspectives* (pp. 209-227). Amsterdam: John Benjamins

Lakshmanan, U. & Selinker, L. (2001). Analyzing interlanguage: how do we know what learners know? *Second Language Research,* 17, 4: 395.

Lawson, A. (2001). Collecting, aligning, and analyzing parallel corpora. In Ghadessy, M, A. Henry & R.L. Roseberry (Eds), *Small corpus studies and ELT theory and practice* (pp. 279-310). Amsterdam: John Benjamins Publishing Company.

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer and B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 8-29). London: Longman

Lehmann, W.P. (1973). A structural principle of language and its implications. *Language*. 49, 47-66

Leńko-Szymańska, A. (2004). Demonstratives as anaphora markers in advanced learnersʼ English. In G. Aston, S. Bernardini & D. Stewart, (Eds.), *Corpora and language learners* (pp. 89-108). Amsterdam / Philadelphia: Benjamins.

Lin, L. H. F. (2002). Overuse, underuse and misuse: using concordancing to analyze the use of it in the writing of Chinese learners of English. In M.Tan (Ed.), *Corpus*

Studies in Language Education (pp. 63-80). Bangkok: IELS Press.

Lorenz G. (1998). Overstatement in advanced learners' writing: stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on computer* (pp. 53-66). London & New York: Addison Wesley Longman.

Mahlberg, M. (2005). *English general nouns: A corpus theoretical approach.* Amsterdam/ Philadelphia: John Benjamins Publishing Company.

McEnery, T. (2005). *Swearing in English: Bad language, purity and power from 1586 to the Present.* London: Routledge.

McEnery, T. & A. Wilson (2001). *Corpus linguistics* (2nd ed.), Edinburgh: Edinburgh University Press.

McEnery, T. Xiao, R. & Tonio, Y. (2006). *Corpus-based language studies: An advanced resource book.* New York: Routledge.

Meyer, C.F., (2002). *English corpus linguistics.* Cambridge: Cambridge University Press.

Narita, M., Sato, C. & Sugiura, M. (2004). Connector usage in the English essay writing of Japanese EFL learners. *Proceedings of 4th International Conference on Language Resources and Evaluation. LREC*, 1171–1174.

Nesselhauf, N. (2005). *Collocations in a learner corpus. Studies in corpus linguistics.* Amsterdam: John Benjamins.

Oakes, M. P. (1998). *Statistics for corpus linguistics. (Edinburgh textbooks in empirical linguistics.)* Edinburgh: Edinburgh University Press.

O'Keeffe, A. & McCarthy, M. (2010). Historical perspective: What are corpora and how have they evolved? . In A. O'Keeffe & M. McCarthy (Eds.), *The routledge handbook of corpus linguistics*. London: Routledge.

O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge: Cambridge University Press.

Özsoy, A. S. (1999). *Türkçe-Turkish*. Istanbul: Boğaziçi University Press.

Pravec, N. (2002). Survey of Learner Corpora. *ICAME Journal* 26: 82-114. Retrieved March 8, 2011 from http://nora.hd.uib.no/icame/ij26/pravec.pdf

Powell, G. (1998). What is the role of transfer in interlanguage? *Centre for Research in Language Education CRILE Working Papers, 33*. 1-4 Retrieved from February 16, 2011 http://www.ling.lancs.ac.uk/groups/crile/docs/crile33Powell.pdf

Quirk, R., Greenbaum, S. Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.

Reguzzoni, M. (2008). An investigation into a pedagogic corpus of maritime English (ME). *Aston corpus postgraduate conference*, May. 22-23. Birmingham, UK.

Ringbom H. (1998b). Vocabulary frequencies in advanced learner English: a cross-linguistic approach. In S. Granger. (Ed.), *Learner English on computer* (pp. 41-52). London & New York: Addison Wesley Longman.

Ringbom H. (1999). High-frequency verbs in the ICLE corpus. In A. Renouf (Ed.), *Explorations in Corpus Linguistics* (pp. 191-200). Amsterdam and Atlanta: Rodopi, 191-200.

Römer, U. (2007*).* Learner language and the norms in native corpora and EFL teaching materials: A case study of English conditionals. In S. Volk-Birke & J. Lippert (Eds.), *Anglistentag 2006 Halle. Proceedings. Trier:* 355-363 WissenschaftlicherVerlag Trier.

Scott, M. (1996). *WordSmith Tools*. Oxford: Oxford University Press. ISBN 0-19-458984-6.

Scott, M. (1997). *WordSmith tools, version 2*. Oxford: Oxford University Press. ISBN 0-19-459283-9.

Scott, M. (1999). *WordSmith tools, version 3*. Oxford: Oxford University Press. ISBN 0-19-459289-8.

Scott, M. (2004). *WordSmith tools, version 4*. Oxford: Oxford University Press. ISBN: 0-19-459400-9.

Scott, M. (2008). *WordSmith tools, version 5*. Liverpool: Lexical Analysis Software.

Scott, M. (2010). *WordSmith Tools, Version 5 Manual*. Liverpool: Lexical Analysis Software. 2.

Selinker, L. (1972). Interlanguage. *IRAL,* 10, 209-31.

Sinclair, J. (Ed.) (1990). *Collins COBUILD English grammar*. London: Harper Collins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair (2000). Foreword. In B. Dodd (Ed.), *Working with German corpora* (pp. vii-xi). Birmingham: The University of Birmingham Press.

Stefanowitsch, A. (2005). Quantitative thinking for corpus linguistics. Retrieved February 18, 2011 from http://www-user.uni-bremen.de/~anatol/qnt/qnt_dist.html

Svartvik, J. (2007). *Corpus linguistics 25+ years on.* Amsterdam: Rodopi.

Taylor, J.R. (1997). Conditionals and polarity. In A. Athanasiadou & R. Dirven (Eds.), *On conditionals again* (pp. 289–306). Amsterdam: Benjamins.

Tapper, M. (2005). Connectives in advanced EFL learners' written English – preliminary results. In F. Heinat & E. Klingvall (Eds.), *The department of English in Lund: Working papers in linguistics 5.* Lund: Department of English, Lund University.

Teubert, W. (2005). Language as an economic factor: the importance of terminology. In G. Barnbrook, P. Danielsson, M. Mahlberg (Eds.), *Meaningful texts. The extraction of semantic information from monolingual and multilingual corpora* (pp. 96-106). London: Continuum.

Thompson, G. (2001). Corpus, comparison and culture: doing the same things differently in different cultures. *In small corpus studies and ELT theory and practice* (pp. 311-334). Amsterdam: John Benjamins Publishing Company.

Tognini Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam and Philadelphia: John Benjamins.

Tono, Y. (2003). Learner corpora: Design, development and applications. In D. Archer, P. Rayson, A. Wilson & T. McEnery (Eds.), *Proceedings of the Corpus Linguistics 2003 conference. UCREL technical paper number 16* (pp. 800-809). UCREL, Lancaster University.

Tono Y. (2004). Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. In G. Aston, S. Bernardini & D. Stewart (Eds), *Corpora and language learners* (pp. 45-66). Amsterdam / Philadelphia: Benjamins.

Underhill, R. (1976). *Turkish grammar.* Cambridge Mass: MIT Press.

Virtanen T. (1998). Direct questions in argumentative student writing. In S. Granger. (Ed.). *Learner English on computer* (pp. 94-106). London & New York: Addison Wesley Longman.

Ziegeler, D. (2000). *Hypothetical modality: Grammaticalisation in an L2 dialect.* Philadelphia: John Benjamins.

Zülkadiroğlu, G. (2009). *Japonca koşul anlatımları —"-tara", "-ba", "-to", "-nara"—.* Unpublished master thesis, University of Ankara, Ankara.

APPENDICES

APPENDIX 1

The institutions codes manipulated in the study for TICLE.

Nation	Essay Codes	Code
TURKEY	University of Çukurova	CU
TURKEY	Mersin University	ME

APPENDIX 2

The institutions codes manipulated in the study for JPICLE.

Nation	Essay Codes	Code
JAPAN	Aichi Shukutoku University	AI
JAPAN	Dokkyo University	DO
JAPAN	Fuji University	FJ
JAPAN	Hiroshima International University	HI
JAPAN	Keio University	KO
JAPAN	Kooriyama Women's University	KW
JAPAN	Kyoto University	KY
JAPAN	Meiji University	KJ
JAPAN	Miyagi University of Education	MI
JAPAN	Okayama University	OK
JAPAN	Rikkyo University	RI
JAPAN	Seijyo University	SE
JAPAN	Shinshu University	SH
JAPAN	Shonan Institution of Technology	ST
JAPAN	Showa Women's University	SW
JAPAN	Tamagawa University	TM
JAPAN	Tokyo University of Foreign Studies	TF

APPENDIX 3

Corpus collection guideline for compiling TUC.

Yönergeler

1. Aşağıdaki anketi doldurunuz.
2. Verilen kompozisyon konularından bir tanesini seçiniz.
3. Seçtiğiniz konu üzerine Türkçe olarak **en az 500 kelimeden oluşan** kurallara uygun bir komposizyon yazınız.
4. Yazdığınız kompozisyonu ve anket bilgilerinizi derlembilim@gmail.com adresine "Word" dökümanı olarak yollayınız.
5. Yazılan kompozisyonda sadece kendi görüşlerinizi dile getiriniz, herhangi bir kaynaktan(internet, kitap, gazete, dergi ...vb.) yararlanmayınız.

Anket

Adınız-soyadınız:

Öğrenci no ve şubeniz:

Yaşınız :

Cinsiyetiniz :

Ana diliniz :

Üniversite Adı:

Yaşadığınız Ülke:

Bildiğiniz diğer diller:

Komposizyon Konuları

1. Suç hiç kimseye yarar sağlamaz. Çünkü eninde sonunda yakalanırsınız.
2. Cezaevi sistemi eskidi. Uygar toplumlar suçlularını cezalandırmaktansa onları rehabilite etmelidir.
3. Çoğu üniversite diploması teoriktir ve öğrencilerini gerçek hayata hazırlamaz. Bu yüzden hiçbir değeri yoktur.
4. Bir erkeğin yada kadının mali geliri (kazancı), yaşadığı topluma yaptığı katkılarla uyumlu olmalıdır.
5. Batı toplumunda sansürün yeri.
6. Marx dinin kitlelerin afyonu olduğunu söylemiştir. 20 yy. sonlarında yaşıyor olsaydı dini televizyonla değiştirirdi.
7. Tüm ordular tamamen profesyonel askerlerden oluşmalıdır. Askerlik sisteminde değer kavramı yoktur.

8. Körfez savaşı sonrası bize insanın ülkesi için savaşmasının hala önemli (kutsal) bir şey olduğunu gösterdi.
9. Feministler, kadınların davasına yarardan çok zarar vermişlerdir.
10. Hayvan Çiftliği romanında George Orwell şöyle yazmıştı "Bütün insanlar eşittir; fakat bazıları diğerlerinden daha eşittir." Bu görüş günümüzde ne kadar doğrudur?
11. Eski bir şarkı sözüne göre para bütün kötülüklerin anasıdır.
12. Avrupa; Egemenliğin yitirilmesi mi yoksa yeni bir toplumun doğuşu mu?
13. 19 yüzyılda Victor Hugo "Doğanın insanlara seslendiğini ancak insanların onu önemsemediğini düşünmek ne acı" demiştir. Bu tümcenin bugün hala geçerli olduğunu düşünüyor musunuz?
14. Bazı insanlara göre bilim, teknoloji ve sanayileşme ile yönetilen günümüz dünyası da düş ve hayal gücüne artık yer yok. Bu konuda düşünceniz nedir?

APPENDIX 4

The verb formation of the word "-sev" and conditional marker "–sA" (Banguoğlu, 1990, p. 444 and Hengirmen, 1998, p.237)

Past	**Indirect Past**	**Present**	**Continuous**
Sevdiysem	Sevmişsem	Seversem	Seviyorsam
Sevdiysen	Sevmişsen	Seversen	Seviyorsan
Sevdiyse	Sevmişse	Severse	Seviyorsa
Sevdiysek	Sevmişsek	Seversek	Seviyorsak
Sevdiyseniz	Sevmişseniz	Severseniz	Seviyorsanız
Sevdiyseler	Sevmişseler	Severseler	Seviyorsalar

Future	**Necessitative**	**Optative-Conditional**
Seveceksem	Sevmeliysem	Sevseymişim
Seveceksen	Sevmeliysen	Sevseymişsin
Sevecekse	Sevmeliyse	Sevseymiş
Seveceksek	Sevmeliysek	Sevseymişiz
Sevecekseniz	Sevmeliyseniz	Sevseymişsiniz
Sevecekseler	Sevmeliyseler	Sevseymişler

CURRICULUM VITAE

PERSONAL DETAILS

Name : Yunus Emre

Surname : AKBANA

Place and Date of Birth : Bingöl, Turkey – 20 June 1986

Work Adress : Kahramanmaraş Sütçü İmam University – Department of Foreign Languages

E-mail : yemreakbana@hotmail.com
y.emre@ksu.edu.tr

EDUCATIONAL BACKGORUND

Date	Institute	Degree
2009 – 2011	University of Çukurova - The Institute of Social Sciences, Department of English Language Teaching, Adana, Turkey	MA
2004 – 2008	Anadolu University - Department of English Language Teaching, Eskişehir, Turkey	BA
2000 – 2004	Mithatpaşa Şükrü Ayna Foreign Language Intensive High School	High School

WORK EXPERIENCE

Date	Institute	Position
2009 – …	Kahramanmaraş Sütçü İmam University - Kahramanmaraş, Turkey	English Lecturer
2008 – 2009 (Fall Semester)	Bahçelievler Gazi Elementary School – Sakarya, Turkey	English Teacher

CONFERENCES ATTENDED

1-3 June, 2010	Zirve University 1st International Foreign Language Teaching Conference: Independent Learning, Gaziantep, Turkey
19-22 May, 2010	3rd ELT Symposium, Learning from our Learners, Tekirova, Antalya, Turkey
1 April, 2010	Kahramanmaraş Sütçü İmam University, First Local ELT Training Program, Kahramanmaraş, Turkey
27,28 March, 2010	Yeditepe University. Istek Schools, International ELT Conference, İstanbul, Turkey
22 June – 03 July, 2009	Sabancı University, 2nd School of Languages Trainer Education Program (SLTEP), İstanbul, Turkey
08, 09 May, 2009	TOBB 'Reaching Our Horizons', Ankara, Turkey

PAPERS PRESENTED

28-28 May, 2011	A Corpus Based Study on L2 English Conditional Adverbial Clauses**,** 23rd International Conference on Foreign/ Second Language Acquisition, Szcyrk, Poland
05-07 May, 2011	25. Ulusal Dilbilim Kurultayı, Çukurova University, Adana, Turkey
12-15 June, 2010	Second Language Acquisition (SLA) of English Reflexive and Referring Pronouns by Turkish University Prep-Class Students, 3rd Annual International Conference on Philology, Literature & Linguistics, Athens, Greece
28 April, 2010	What makes a good teacher?, Kahramanmaraş Sütçü İmam University, Second Local ELT Training Program, Kahramanmaraş, Turkey