T.C.

FIRAT UNIVERSITY

INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF ENGLISH LANGUAGE AND LITERATURE

A CORPUS-INFORMED STUDY BASED ON A CONTRASTIVE ANALYSIS OF STANCE MARKERS IN LEARNER ENGLISH: FROM CORPUS TO CLASSROOM

MASTER THESIS

SUPERVISOR
Asst. Prof. Dr. Aysel ŞAHİN KIZIL

PREPARED BY
Zehra SAVRAN

ELAZIĞ–2017

T.C

FIRAT ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

BATI DİLLERİ VE EDEBİYATI ANABİLİM DALI

İNGİLİZ DİLİ VE EDEBİYATI BİLİM DALI

A CORPUS-INFORMED STUDY BASED ON A CONTRASTIVE ANALYSIS OF STANCE MARKERS IN LEARNER ENGLISH: FROM CORPUS TO CLASSROOM

YÜKSEK LİSANS TEZİ

DANIŞMAN	HAZIRLAYAN
Yrd. Doç. Dr. Aysel ŞAHİN KIZIL	**Zehra SAVRAN**

Jürimiz, .../...../..... tarihinde yapılan tez savunma sınavı sonunda bu yüksek lisans tezini oy birliği / oy çokluğu ile başarılı saymıştır.

Jüri Üyeleri:

1. Yrd. Doç. Dr. Aysel ŞAHİN KIZIL (Danışman)

2. Yrd. Doç. Dr. Abdurrahman KİLİMCİ

3. Yrd. Doç. Dr. F. Gül KOÇSOY

F.Ü. Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun …… tarih ve …… sayılı kararıyla bu tezin kabulü onaylanmıştır.

Prof. Dr. Ömer Osman UMAR

Sosyal Bilimler Enstitüsü Müdürü

ÖZET

Yüksek Lisans Tezi

Öğrenici İngilizcesinde Tutum Belirteçlerinin Karşılaştırmalı Analizine Dayalı Derlem Temelli bir Çalışma: Derlemden Sınıfa

Zehra SAVRAN

Fırat Üniversitesi
Sosyal Bilimler Enstitüsü
Batı Dilleri ve Edebiyatları Anabilim Dalı
İngiliz Dili ve Edebiyatı Bilim Dalı
Elazığ-2017, Sayfa: XIV + 159

Herhangi bir dilsel etkileşim içerisinde tutum belirteçlerinin büyük rolünün tanınması üzerine yapılan çalışmalar giderek artmaktadır. Bu çalışmada, Karşılaştırmalı Aradil Analizi (CIA) çerçevesinde, öğrenici İngilizcesinde kullanılan bilgisel tutum belirteçlerini araştırmak ve ortaya çıkarmak amacıyla derlem temelli yaklaşımdan yararlanılmıştır. Araştırmanın iki temel hedefi: 1) Türk öğrenicilerin sözlü İngilizcede kullandıkları tutum belirteçlerinin ortaya çıkarılarak bu belirteçleri anadili İngilizce olan kullanıcıların sözlü İngilizcede kullandığı belirteçlerle karşılaştırmak; 2) Derlem temelli öğretim metotlarının öğrenici İngilizcesinde bilgisel tutum belirteçlerinin kullanımı açısından bir fark oluşturup oluşturmadığını ölçmektir. Bu hedefler doğrultusunda, bu çalışma birbirini tamamlayan iki bölümden oluşmaktadır. İlk bölümde öğrenici İngilizcesi veri kaynağını Louvain Uluslararası Aradil Konuşma İngilizcesi Veritabanı (LINDSEI) Projesi'nin Türkçe alt bileşeni ve anadil konuşuru veri kaynağını The Louvain Corpus of Native English Conversation (LOCNEC) veri tabanı oluşturmaktadır. İkinci bölüm ise, Fırat Üniversitesi'nde İngiliz Dili ve Edebiyatı bölümünde öğrenim gören 39 lisans öğrencisi ile yürütülmüştür.

Çalışmanın ilk bölümünden elde edilen bulgulara göre, konuşma dilinde sıklıkla kullanılan tutum belirteçleri açısından Türk öğrenciler ile anadili İngilizce olanların İngilizcesi arasındaki benzerlik ve farklılıklar incelendiğinde, oldukça karmaşık bir tablo ortaya çıkmıştır. Çalışmanın ikinci bölümünde ise, derlem temelli öğretim metotlarının konuşma dilinde yararlanılan bilgisel tutum belirteçlerini öğretmede kısa vadede etkili olabileceği; fakat öğrenmede uzun vadede kalıcılığı sağlamak amacıyla öğrencilerin hedef belirteçler üzerine devamlılığı olan bir eğitime ihtiyaç duyabilecekleri tartışılmıştır.

Anahtar Kelimeler: Tutum, Bilgisel Tutum Belirteçleri, Söylem İşaretçileri, Sözlü Öğrenici Derlemi, İkinci Dil Edinimi (SLA), Yabancı Dil Olarak İngilizce (EFL), Derlem Temelli Öğretim, Karşılaştırmalı Aradil Analizi (CIA).

ABSTRACT

Master Thesis

A Corpus-Informed Study Based On A Contrastive Analysis Of Stance Markers In Learner English: From Corpus To Classroom

Zehra SAVRAN

Fırat University
Institute of Social Sciences
Department of Western Languages and Literatures
Division of English Language and Literature
Elazig-2017, Page: XIV + 159

A growing body of research has been conducted upon the recognition of the vital role of stance markers in any linguistic interaction and negotiating knowledge. Within the framework of Contrastive Interlanguage Analysis (CIA), this thesis adopted a corpus-informed approach to describe and uncover the role of epistemic stance markers in spoken interlanguage. The main goals of the present study are two-fold: 1) to investigate the use of epistemic stance markers in Turkish EFL learners' spoken production, comparing and contrasting them with the target structures in native speech 2) to measure the change, if any, occurring in epistemic stance marker use in L2 learners' oral performance as a result of corpus-informed treatments. In achieving these goals, the present study consists of two interrelated studies. Study 1 exploits the Turkish subcomponent of Louvain International Database of Spoken Interlanguage (LINDSEI) for non-native data, and as the native counterpart, The Louvain Corpus of Native English Conversation (LOCNEC). Study 2 was conducted in Fırat University, Elazığ with 39 students majoring in English Language and Literature.

The findings of Study 1 revealed a rather distinctive picture of the similarities and differences between native and learner language regarding the commonly used epistemic stance markers in spoken language. As for Study 2, it was found out that a

corpus-informed teaching methodology could be effective in teaching EFL learners how to express their epistemic stance appropriately in speech in the short term, however; the learners may need to revisit the target forms regularly in the long term to increase the degree of retention of epistemic stance markers.

Key words: Stance, Epistemic Stance, Discourse Markers, Spoken Corpora, Second Language Acquisition (SLA), English as a Foreign Language (EFL), corpus-informed, contrastive interlanguage analysis (CIA), learner corpora.

TABLE OF CONTENTS

ÖZET ..... **II**

ABSTRACT ..... **IV**

TABLE OF CONTENTS ..... **VI**

LIST OF TABLES ..... **X**

LIST OF FIGURES ..... **XII**

PREFACE ..... Hata! Yer işareti tanımlanmamış.

LIST OF ABBREVIATIONS ..... **XIV**

CHAPTER I

1. INTRODUCTION **1**

1.1 Introduction 1

1.2. Background of the study 2

1.3 Purpose of the study 7

1.4 Research questions 8

1.5 Significance of the study 9

1.6 Definition of terms 10

1.7 Thesis Overview 11

1.8 Chapter Summary 12

CHAPTER II

2. LITERATURE REVIEW **13**

2.1 Introduction 13

2.2 Corpus Linguistics 13

2.2.1 History of Corpus Linguistics 15

2.3 Learner corpora 19

2.3.1 Major Learner Corpora around the World 22

2.3.1.1. LINDSEI: Louvain International Database of Spoken English Interlanguage 25

2.3.1.2 LOCNEC: Louvain Corpus of Native English Conversation 26
2.4 Stance: An overview.....26
2.4.1 Stance Markers in Spoken Language.....31
2.4.2 Epistemic Stance 33
2.4.3 Epistemic Stance Markers in L2 spoken Corpora.....37
2.5 Corpus and Language Teaching 39
2.5.1 Teaching Stance Markers.....43
2.5.2 Teaching Stance Markers through Learner Corpora: A Meeting Point 45
2.6 Chapter Summary.....46

CHAPTER III

3. RESEARCH METHODOLOGY 47
3.1 Introduction 47
3.2 Research Design 47
3.3 Study 1: Corpus-based Linguistic Analysis 50
3.3.1. Materials: LINDSEI & LINDSEI-TR.....50
3.3.1.1 Learner variables.....50
3.3.1.2 Task Variables 52
3.3.1.3 Reference Corpus: LOCNEC.....54
3.3.2 Procedures.....54
3.3.3 Data Analysis 55
3.3.3.1 Quantitative Analysis.....55
3.4 Study 2: Corpus-Informed Intervention 58
3.4.1 Participants.....58
3.4.2 Data Collection: 60
3.4.3 Procedures.....62
3.4.4 Data Analysis 64
3.5 Chapter Summary.....66

CHAPTER IV

4. FINDINGS AND DISCUSSION.....67

4.1 Introduction 67

4.2 Study 1: Corpus-based Linguistic Analysis 67

4.2.1 Epistemic stance markers in LINDSEI and LOCNEC: Quantitative Analysis.....68

4.2.2 Epistemic stance markers in LOCNEC: Quantitative Analysis.....76

4.2.3 Epistemic stance markers in LINDSEI vs LOCNEC.....81

4.3 Study 2: Corpus-Informed Intervention 89

4.3.2 Overall Learning Effects of Concordance Use on Learning Epistemic Stance Markers.....89

4.3.3 Effects of Intervention on the Frequency Counts and Range of Epistemic Stance Markers.....92

4.3.4 Individual Epistemic Stance Markers in Pre-test vs. Post-test.....96

4.3.5 Learners' Epistemic Marker Use According to the Task Types 98

4.3.6 Learner Attitudes Towards the Use of Concordancing in Learning Epistemic Stance Markers in Spoken Language 100

4.4 Chapter summary.....106

CHAPTER V

5. CONCLUSION 108

5.1 Introduction 108

5.2 Overview of the Current Study.....108

5.3 General Findings: Study 1 Corpus-Based Contrastive Analysis 109

5.4. General Conclusions: Study 2 Corpus-Informed Intervention 112

5.5 Pedagogical Implications and Suggestions for ELT 115

5.6 Limitations of the Study and Suggestions for Further Research 117

REFERENCES.....119

APPENDICES.....136

CURRICULUM VITAE.....1589

LIST OF TABLES

Table 1. Three phases of corpus evolution 18

Table 2. Design criteria for building learner corpora 21

Table 3. Types of processing of learner data 22

Table 4. Major learner corpora around the world 23

Table 5. English background knowledge: LINDSEI-TR 51

Table 6. Distribution of other foreign languages in LINDSEI-TR 51

Table 7. LINDSEI task variables summarized 53

Table 8. Epistemic stance markers investigated 55

Table 9. Characteristics of participants summarized 59

Table 10. Grammatical classes of epistemic stance markers in LINDSEI-TR vs LOCNEC 68

Table 11. The most common epistemic markers in LINDSEI-TR 70

Table 12. The most frequent epistemic markers in LOCNEC 76

Table 13. Epistemic stance markers in LINDSEI vs LOCNEC 81

Table 14. Overused and underused epistemic stance markers in LINDSEI-TR 88

Table 15. Pre-test vs. Post-test: overall scores 89

Table 16. Analysis of covariance for overall epistemic marker use by proficiency level 90

Table 17. Retention of epistemic stance marker knowledge 91

Table 18. Comparison of data in terms of frequency counts 93

Table 19. Comparisons of pre-test and post-test results on epistemic range scores 95

Table 20. Residual effects of the intervention on epistemic range scores 95

Table 21. Frequency information of epistemic devices in three task types 98

Table 22. Overall range of epistemic devices in three task types 99

Table 23. Perceptions towards the use of BNC for learning epistemic stance markers 102

Table 24. Perceptions on the difficulties on using BNC 103

Table 25. Perceptions towards the general use of BNC for learning epistemic stance markers 105

LIST OF FIGURES

Figure 1. LINDSEI Design Criteria 26

Figure 2. Stance Triangle. 28

Figure 3. Epistemic Stance Markers. 35

Figure 4. Grammatical Classes of Epistemic Stance in Order of Frequency 36

Figure 5. Applications of Corpora in Language Teaching 41

Figure 6. Critical Values in Log-likelihood Statistics. 56

Figure 7. Grammatical Distribution of Epistemic Stance Markers in LINDSEI-TR vs LOCNEC 69

Figure 8. Individual Epistemic Stance Markers in Pre-test 96

Figure 9. Individual Epistemic Stance Markers in Post-test. 97

PREFACE

I would like to thank a few people who have made it possible for me to complete this thesis.

First and foremost, I would like to express my greatest appreciation to my supervisor Asst. Prof. Dr. Aysel ŞAHİN-KIZIL for her excellent guidance and teaching. This study has become possible with her never-ending support, patience, motivation, immense knowledge of linguistics and love for her job. I must state that I have felt very privileged to be trained by an instructor like her as beyond being an excellent instructor, she has always been like one of the family. I could not have imagined having a better advisor and mentor for my MA study.

Besides my advisor, I would like to extend my special thanks to Asst. Prof. Dr. Abdurrahman KİLİMCİ and Asst. Prof. Dr. F. Gül KOÇSOY for being in my thesis committee, for their precious teaching, comments and feedback.

My heartfelt thanks are extended to my beloved parents, Ahmet SAVRAN and Zeynep SAVRAN, who are my first teachers in life. I am extremely grateful to them that they have always believed in me, loved me, and supported me in all my decisions. I have always felt proud of being their daughter.

I also would like to thank my precious sister, İlay KAYA for being the best part of my life. She has always been a source of inspiration for me. I'm truly indebted to her for everything I have in life. I have always felt much stronger and happier with her. Also, I must thank her for giving me a great brother, Deniz KAYA.

I owe special thanks to my beloved ones, to my future family, Mete ÇELİK, and Güneş, for their endless love and support, and for being there for me whenever I needed them. They have always been a source of happiness for me.

Last but not least, I would like to thank, my little niece, Naz KAYA, who is the brightest part of my life. Life is worth living with her.

LIST OF ABBREVIATIONS

BNC : British National Corpus

CIA : Contrastive Interlanguage Analysis

EFL : English as a Foreign Language

ESL : English as a Second Language

NNS : Non-Native Speakers

NS : Native Speakers

L1 : First (native) Language

L2 : Second Language

LINDSEI : Louvain International Database of Spoken English Interlanguage

LINDSEI-TR: Louvain International Database of Spoken English Interlanguage, Turkish Component

LOCNEC : The Louvain Corpus of Native English Conversation

LL : Log-likelihood

SLA : Second Language Acquisition

CL : Corpus Linguistics

ELT : English Language Teaching

CHAPTER I

INTRODUCTION

1.1 Introduction

The idea of collecting texts or parts of texts to conduct linguistic research on is not a newly emerged one. However, it was the advent of technology which has led to a new branch of linguistics *corpus linguistics*. Even though the origins of the methods used in corpus linguistics date back to late eighteenth and nineteenth century, the real breakthrough of the discipline came out with the development of technology and corpus analytical tools in 1970s and 1980s, when it became possible to store, compile and analyse different types of corpora (Johansson, 2008). Although there have been various definitions to the function of corpus linguistics, it is generally defined as a systematic collection of naturally occurring texts. Since the corpus linguistics has proven itself to offer highly efficient and concrete evidence to the nature, structure and function of language over the years, it has become a widely-used method in linguistic research (Nesselhauf, 2005).

Since the emergence of corpus linguistics, the scope of this branch of linguistics has widened. Different types of corpora such as written, spoken, learner and specialized corpora were compiled for different purposes. Krieger (2003) states that by conducting analysis on these different kinds of corpora, one can examine almost any language patterns and linguistic mechanisms in terms of lexis, structure, lexico-grammar, discourse, phonology, morphology and pragmatics along with the contextual features of these language patterns.

In addition to the patterns investigated through the use of corpora, over the last decades or so, there has been an increasing interest in the linguistic mechanisms that the speakers and writers employ to express their feelings, judgements and assessments. Being one of the linguistic mechanisms, stance has been an area of continuous interest to linguistics. Stance refers to the lexical and grammatical expression of personal feelings, judgements, attitudes or commitment that a writer or speaker has towards the information given in a proposition (Biber & Finegan, 1989).

Since it is well established in the literature that the contextual and cultural function of the stance markers plays a more crucial role in interaction rather than literal semantic meanings, there has been a great number of studies to explain the stance markers that the native speakers (NS) use in written language (Silver, 2003; Adams & Quintana-Toledo, 2013; Aull & Lancaster, 2014; Lancaster, 2014). In addition to a good number of investigations of stance in native written language, stance has also been examined in native spoken language (Brezina (2013; Precht, 2003; Keisanen; 2007; Kärkkäinen, 2006).

Besides the studies performed to explore the phenomenon of stance in native language, there has also been attempts to examine stance markers in learner corpora through contrastive interlanguage analysis in written medium (Efstathialdi, 2010; Jiajin & Manying, 2008; Chang & Schleppegrell, 2011) and in spoken medium (Gablasova, Brezina, Mcenery, & Boyd, 2015; Letica 2009).

A closer look at the relevant literature, however, implies that most of the research carried out focuses on written language rather than spoken language and the extant literature calls for the need to explore the use of stance markers in spoken language and in different second/foreign language (henceforth L2) contexts (Gablasova et al. 2015) and also investigate the effectiveness of teaching stance markers to non-native speakers (NNS) of English.

For that purpose, this study seeks to identify frequent stance markers in spoken language through a contrastive interlanguage analysis and then test the efficiency of applying the results obtained from the corpus based analysis to the classroom.

1.2. Background of the study

When corpus linguistics first came in sight in the late 1950's, it was an enterprise, which drew little or no attention from linguistics or computer science. However, it has been an interest to a rapidly increasing number of researchers in most language learning related disciplines (Granger, 1993).

In the 1960s and 1970s, corpus linguistics lost its popularity due to the criticisms of Chomsky regarding the competence – performance theory. Chomsky stressed the unfruitfulness of corpora for linguistic studies based on the idea that a linguist should focus on language competence rather than performance. Another criticism the corpus

linguists face was related to data processing. However, as stated above, the developments in technology around 1980s has turned corpus linguistics into a feasible methodology (McEnery & Wilson, 2001). From that time on, it was made possible to compile, analyse and store large bodies of naturally occurring data, which has widened the scope in study of language and, there has been various purposes for compilation of corpora such as answering questions and solving problems on the linguistic levels of lexis, grammar, pragmatics or discourse. Stubbs and Halbe (2013) lists the major areas in which the corpus methodology used: a) different languages from English, German to sign language; b) varieties of English (English as a Lingua Franca, English for Academic purposes, etc.); c) written text types; d) spoken and written language in social situations such as workplace and classroom; e) vocabulary and phraseology to develop multilingual and bilingual dictionaries; and f) work which has pedagogical implications such as corpus analysis of child language. Since the scope of the corpus is a vague one, it can be claimed that thanks to the developments in technology, corpus linguistics has revealed a lot about the nature of native and learner languages and been applied to some practical problems including preparing dictionaries for advanced learners, helping to make documents easier for average readers, and comparing quantitative features of texts and so on (Stubbs & Halbe, 2013).

Being a linguistic mechanism that reveals substantial facts about the language use, stance markers are another major area to investigate through corpus linguistics thanks to newly developed corpus software.

Since the notion of stance is rather vague, it has been covered under different labels in various linguistic research such as *evidentiality* (Chafe, 1986), *evaluation* (Hunston & Thompson, 2000), *hedging* (Hyland, 1996) and *affect* (Ochs, 1989), *appraisal* (White, 2001). However, there are no clear-cut dividing lines between the explanations of these terms. As Du Bois (2007) states:

> One of the most important things we do with words is take a stance. Stance has the power to assign value to objects of interest, to position social actors with respect to those objects, to calibrate alignment between stance takers, and to invoke presupposed systems of cultural value (p. 139)

Analysis of stance has gained new dimensions with the advantages of corpus analytical tools. Since the corpus aims to describe naturally occurring data in a

particular language, it could be considered to be a very useful tool for a linguist to investigate the typical patterns of spoken or written language utilizing machine-readable data on the corpus. Hunston (2007) claims that using corpora in language studies not only enables the linguists to quantify forms but also to explore the different uses of a word in context. Investigating the phenomenon of stance, therefore, poses a challenge for the corpus linguists as stance is regarded more as a meaning, not form, which requires a more profound understanding of discourse but at the same time corpus methods can make a useful contribution to the investigation of stance (Hunston, 2007).

Biber, Johansson, Leech, Conrad & Finegan, (1999) asserted that there are three main categories of stance: 1) epistemic stance which is related to certainty, doubt, actuality, source of knowledge, imprecision, viewpoint, and limitation, 2) attitudinal stance which is related to states, evaluations, emotions and attitudes), and 3) style stance (- which is related to style of speaking). As stance plays an important role in conveying many various kinds of personal feelings and assessments that reveals how certain writer or speaker is about the proposed content and what stance they are taking towards the message in context, all these three categories of stance have been explored in different contexts and in different mediums to be able to give a broader description of stance taking.

Most of the research conducted on stance taking through corpus-based analysis focused on academic writing. Looking at the adverbial as a hedge and booster on the adverbial *evidently*, Silver (2003) sets out to study how evaluation is expressed in academic discourse, and stresses the adverbial's role in expression and evaluative proposition as well as constructing writer and reader identity. Studying a different register, Adams and Quintana-Toledo (2013) explore the occurrence of adverbial stance markers in legal research articles and concludes that the expression of doubt is favoured and the frequency of the markers is much higher in conclusion parts. Aull and Lancaster (2014) compare the stance markers in early and advanced academic writing and highlight construction of academic stance. Lancaster (2014) conducts a comparative analysis on high and low performers' stance taking styles in the learners' argumentative essays and revealed that while high performers take a more novice academic stance, low performers are more inclined to take a 'student' stance, which is less committed. Based on the results of the study, Lancaster (2014) highlights the importance of raising

awareness in instructors' feedback on student work to make their explanations more explicit to the learners.

Along with the studies carried out in written native language, the investigations in spoken native language have provided a deeper insight in order to understand stance better. Brezina (2013) conducts a corpus-based analysis on the use of epistemic stance markers in spoken language and finds out that in expressing certainty or uncertainty, speakers have a tendency to be repetitive in expressing similar epistemic markers. Comparing British and American English conversation through corpora, Precht (2003) argues that British and American speakers tend to design their socialization patterns differently. Keisanen (2007) examines stance taking patterns in negative yes/no interrogatives and tag questions in spoken American English conversation drawing the data from corpus, and establishes that the patterns used in negative yes/no interrogatives and tag questions show a highly-restricted pattern in linguistic terms and also that the speakers can employ one or the same set of stance markers to perform different actions. Similarly, Kärkkäinen (2006) explores the epistemic stance markers in American English conversation and concludes that the markers function to enrich the interaction between the interlocutors, and at the same time they have a highly-routinized pattern in spoken native language.

Inspired by the research in native language, a number of studies recently have focused on non-native speakers through corpus-based analysis to give a broader description to the nature of stance.

Gablasova et al. (2015) report that most of the stance studies on L2 have concentrated upon written language and stresses the need to find out more about how L2 speakers take their stance in spoken language. Efstathialdi (2010) conducts a comparative study on the use of epistemic markers as a means of hedging and boosting in first language (L1) and L2 speakers of Modern Greek. Compiling a non-native speaker corpus, and the native speaker corpus, the study reveals that the compilation of electronic corpora can contribute greatly to understanding of semantic nuances in the expression of epistemic stance markers. She adds that the findings of the study favour the use of exercises based on corpora in the English as a Foreign Language (EFL) classroom. Finally, since the study explores the written language, she notes that further research on spoken production may help reach a wider variety of epistemic markers

with different patterns in terms of pragmatic use. In parallel with this study, Jiajin and Manying (2008) carried out a comparative study using a corpus consisting of 122 pieces of Chinese EFL learners' English and Chinese argumentative essays. The findings of the study suggest that in terms of stance marking, Chinese learners' English essays share the same tendency with their Chinese writings and they tend to use similar stance types in two languages. Due to L1 effect that has been found to take place in writings of English language learners, the writers state that the strategies to be pursued to teach stance markers should be redesigned according to the customs peculiar to the learners' mother tongue. In the same line, Chang and Schleppegrell (2011) search for the authorial stance markers in academic writing and add that making these linguistic resources explicit to the L2 learners can help the students develop an authorial stance in their products. The study asserts that corpus may have great potential to assist L2 learners to present interpersonal meanings in diverse contexts. Chandrasegaran and Kong (2007) analyse the stance related behaviours of non-native students in an online forum and report that the learners not only project a stance but also their stance behaviour differs. They claim that further research might provide a more precise description to the types of stance, and stance-taking strategies of learners with different backgrounds. Chen (2010) compares data regarding epistemic modality in academic writing through a native and a non-native corpus. Examining also the pragmatic development process of L2 learners, the study concludes that the epistemic markers native and non-native speakers employ in their writings differ from each other and the awareness of using epistemic markers increase with the level of proficiency in L2. What's more, she emphasizes the need for explicit pragmatic instruction in writing and that underused epistemic devices should be stressed more in textbooks. Fordyce (2014) investigates the effect of implicit and explicit instruction on L2 learners' use of epistemic stance collecting written data from the students and favours explicit instruction to teach stance markers. The study implies the need for intervention focusing on form-function mapping, structural complexity, and processing demands of the target forms.

Differently from the previous studies on written language, exploring stance in L2 speech, Gablasova et al. (2015) discuss epistemic stance in spoken L2 via

Trinity Lancaster Corpus. Covering three epistemic categories; adverbs, adjectives and verbal expressions through different task types, the study highlights that there is a systematic change in L2 speakers' preference of stance markers across different tasks. They also claim that advanced L2 speakers' expression of epistemic stance was related to task type and individual speaker style. Lastly, they argue that rather than focusing on verbal, adverbial and adjectival epistemic devices, future research should be conducted on other major epistemic forms. Letica (2009) used two corpora to investigate epistemic modality; the first one compiled from picture description tasks in English by 33 Croatian students, and the second one compiled from the same task type in Croatian by the same group of students. The study points out that learners use epistemic markers less frequently in their L2 than in their L1 and that this result doesn't correlate with the L2 learners' proficiency level. Last but not least, the study suggests that as the task type used in the research (i.e. picture description) may have influenced the bundling of the markers used, other tasks requiring conversational exchange would be more informative of the range of epistemic use. In the same vein, Baumgarten and House (2010) compares the stance markers *I think* and *I don't know* in native and non-native speech and they find out that even if they are the mostly used markers in both languages, they only partially overlap in functional terms. They concluded that L1 and L2 speakers use stance expressions *I think* and *I don't know* with different range of functions.

The literature sketched above implies a need to further analyse stance taking in L2 spoken discourse. In addition, as Fordyce (2014) stresses the key role of the pragmatic competence in knowledge of an L2: when the results of the analysis of stance markers with corpus analytical tools are put into practice in educational settings, it has potential to enable practitioners to take more informed steps in teaching stance markers. To this end, the present study adopting a-two-phase research agenda aims at first identifying stance markers employed by Turkish learners of English and comparing them with those in native language and thereafter investigating the effect of explicit teaching of stance markers to L2 learners through corpus informed treatments.

1.3 Purpose of the study

Gablasova et al. (2015) argues that effective communication requires the urge for the ability to differentiate between various interactional contexts and to adjust one's

speech according to these contexts. They add that L2 speakers utilize the same set of stance markers without taking the type of genre into account and add that it is a real struggle for even advanced L2 learners to employ and make correct use of linguistic items to be able to take a stance in their speech. Stance taking, Kiesling (2009) reports that, is one of the fundamental properties of interaction. In addition, epistemic markers are regarded as the crucial part of communication (Conrad & Biber, 1999; Kärkkäinen, 2003). However, it is reported in the relevant literature that epistemic stance use in L2 spoken interaction still remains a relatively unexplored area and most of the previous research focused on written language mainly exploring academic texts (Letica, 2009). In addition, it is claimed that epistemic stance in terms of instructional effects has been given little attention (Fordyce, 2014).

Addressing the identified gaps in the relevant literature, the purpose of the study is twofold. The study firstly sets out to compare and contrast the epistemic stance markers employed by native and non-native speakers through a corpus based analysis, then to examine the effect of explicit teaching of stance markers through corpus informed steps to L2 learners in Turkish EFL context.

1.4 Research questions

The present study aims to investigate the epistemic stance markers in a spoken corpus compiled from Turkish EFL learners and compare the findings with a native speaker spoken corpus. Conveying the findings gathered from corpus-based analysis to the instructional settings, it also intends to test the efficiency of explicit teaching of stance markers to Turkish EFL learners. Therefore, the study will seek to answer the following research questions:

1. What is the distributional pattern of epistemic stance markers according to their grammatical classes in native and non-native speech?

2. What are the most frequent epistemic stance markers native English speakers tend to use in their spoken discourse?

3. What are the most frequent epistemic stance markers Turkish learners of English tend to use in their spoken L2 production?

4. What are the instances of NNS epistemic stance markers underused or overused in their speech as compared to NS speech?

5. In what degree is explicit teaching through concordancing effective in developing EFL learners' use of epistemic stance markers in their spoken discourse?

6. What are the Turkish EFL learners' perceptions of the use of concordancing in learning epistemic stance markers?

1.5 Significance of the study

As aforementioned, a growing body of research has been conducted upon the recognition of the vital role of stance markers in any linguistic interaction and negotiating knowledge (Du Bois, 2007; Brezina, 2009; Kiesling, 2009; Biber & Finegan, 1988). However, in spite of the established importance of stance taking in interaction, it has been given relatively little attention in L2 context, particularly in Turkish L1 background (Sahin-Kızıl, 2013; Şahin-Kızıl, & Kilimci, 2014). Although insights from these studies have contributed considerably to the understanding of stance in Turkish L1 background, most of the relevant studies have centred upon written language (Kilimci, 2003; Kilimci, 2009; Can, 2009; Can, 2012).

Furthermore, Kärkkäinen (1992) stresses the widely-held view that it is difficult for L2 learners to acquire epistemic devices due to their being implicit markers of speaker attitude and puts forward that problem has resulted from the lack of explicit teaching of these markers, and from the lack of research on their functions. In the same line, Şahin-Kızıl and Kilimci, (2014) draw attention to the need to increase the learners' awareness of linguistic properties of spoken English through authentic materials and maintain that explicit teaching of the linguistic items would be helpful for Turkish learners' pragmatic development in terms of fluent speech and getting closer to a native-like competency in L2 production.

Taking into consideration the previous research in the relevant literature, the study aims to contribute to the previous literature in a few various ways by:

1) investigating the stance markers in spoken language rather than written language which can be claimed to be an under-researched area in second language acquisition (SLA),

2) identifying the different stance markers assigned by native and non-native speakers through a corpus analysis,

3) implementing an explicit teaching of stance markers to L2 learners based on the idea that acquiring and performing stance markers in speech is difficult even for the advanced learners (Gibbs, 1990; Fordyce, 2014).

In sum, the significance of this study stems from being one of the first attempts to analyse oral production data in Turkish EFL context through considerably large spoken corpora of interlanguage. Firstly, the present study might provide insights into the Turkish learners' spoken interlanguage specifically with respect to epistemic stance marker use through the findings obtained from contrastive analysis of corpora. This could, in turn, inform the practitioners about the educational materials designed for the learners. In this respect, the corpus-informed teaching method adopted in this study can be guiding for ELT teachers to create more effective teaching environments especially for the purpose of teaching speaking skills to the students. Finally, the results of this study may contribute to the relevant literature by leading other researchers to conduct studies with learners from different L1 backgrounds.

1.6 Definition of terms

Corpus is a large, systematic collection of pieces of language texts in electronic form, selected according to criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research (Sinclair, 2005).

Corpus linguistics is a field, which focuses upon a set of procedures, or methods, for studying language (McEnery & Hardie, 2012).

Computer Learner Corpora are electronic and standardized collections of authentic foreign or second language textual data compiled according to explicit design criteria for a particular Second Language Acquisition and Foreign Language Teaching purpose (Granger, 2002).

Contrastive Interlanguage Analysis (CIA) refers to the method of carrying out either a comparison of learner data with native data or a comparison between different kinds of learner data (Granger, 2003).

Interlanguage is defined as "a separate linguistic system based on the observable output which results from a learner's attempted production of a target language norm." (Selinker, 1972, p. 214)

Stance refers to the lexical and grammatical expression of personal feelings, judgements, attitudes or commitment that a writer or speaker has towards the information given in a proposition (Biber & Finegan, 1989).

Epistemic stance is "marking the degree of commitment to what one is saying, or marking attitudes towards knowledge." (Kärkkäinen, 2006, p. 19).

First Language (L1) is the language acquired during early childhood, normally starting before the age of three years (Saville - Troike, 2005).

Second language (L2) or Foreign Language (FL) is typically an official or societally dominant language other than native language needed for education, employment, and other basic purposes (Saville & Troike, 2005).

English as a Foreign Language (EFL) is the language learned mostly in classroom setting without being considerably exposed to the language being learned**.**

1.7 Thesis Overview

The present study comprises of five chapters designed under the titles of introduction, literature review, methodology, findings and discussion, and conclusion.

The first chapter presents brief information on corpus linguistics and its language learning related uses, the phenomenon of stance and related studies on it both in native and learner corpora. The motive to the study is explained in the background in the study. Then, the purpose and the significance of the study is provided. Definitions of the key terms used in the study are given and finally, the last section summarizes the whole chapter.

The second chapter of this dissertation provides detailed information about the relevant literature that this study is based on. This chapter mainly has six sections. The first section opens with a brief introduction to the chapter. The second section provides an overview of corpus linguistics along with its historical development and its use in different fields of language studies. In the third section, main learner corpora around the world, historical evolution of learner corpora together with the studies conducted by means of learner corpora are presented. Then, the corpora exploited in the study are introduced in detail. The fourth section provides detailed information on stance and epistemic stance as a sub-category of stance with a detailed account of studies carried

out with an emphasis on corpus-based analyses of learner language in spoken medium. Fifth section explains the link with corpus and language teaching, elaborating on the methods to teach stance markers, then narrowing down the focus of the section to the teaching stance markers through learner corpora. The last section concludes with a summary of the chapter.

The third chapter provides information about the methodology of the present study. The chapter consists of five main sections. The first section gives a brief introduction to the chapter. The second chapter explains the research design followed for Study 1 & 2. The corpora under investigation, learner and task variables, quantitative and function analysis for the Study 1 are presented in the third chapter. The fourth section introduces the participants, instruments, procedures and data analysis process of Study 2. The chapter concludes with a summary.

The fourth chapter presents the findings reached through corpus-based analysis, which is referred to as Study 1 and the findings from Study 2, which aims to explore the efficiency of explicit teaching of stance markers identified in Study 1. The chapter closes with a summary.

The last chapter of the study is the conclusion chapter of this dissertation. It presents a summary of the results and findings of the study and provides implications and suggestions for language teaching and future studies.

1.8 Chapter Summary

This chapter briefly summarizes the corpus linguistics and the part it takes in language teaching. The background section introduces the phenomenon of stance and stresses the need for the research on epistemic stance in spoken language. The chapter gives information about the significance and purpose of the study together with the research questions. Finally, the definition of the key terms and overview of the thesis are provided.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

This chapter provides the literature about the corpus linguistics stressing its role in language teaching. Firstly, it presents an overview on corpus linguistics and its historical development. Then, learner corpora with its basic features and major learner corpora projects around the world are presented. The next section introduces stance taking, with a focus on its use on spoken language, and research on epistemic stance through learner corpora. The last section intends to explain the connection between corpus and language teaching, emphasizing the importance of teaching stance markers to L2 learners through learner corpora.

2.2 Corpus Linguistics

Since corpus linguistics came into sight in the 1950s, an era of American structuralists such as Harris, Fries, and Hill who asserted what the linguist must study is a corpus of naturally occurring discourse, (Leech, 1992) this branch of linguistics has grown considerably. In early corpus linguistics, under the influence of the positivist and behaviourist approaches, the post-Bloomfieldian American linguists became interested in exploring observable data (Tognini-Bonelli, 2010). Since the debut of the term corpus linguistics; however, there has been disagreement on how to define corpus linguistics. Different definitions of the term have been offered to answer whether it is a methodology, a discipline, a theory, or an approach. Leech (1992) regarded corpus linguistics both as a methodology for language studies and a new research enterprise emphasizing its incomparableness to other branches of linguistics. Leech (1992) favoured the view that corpus linguistics is not a "domain of study" (p. 106), but rather is a methodology providing a baseline for linguistic research. Similarly, Kennedy (1998) referred to corpus linguistics as a scholarly enterprise led by compilation and analysis of computerized databases of written text or transcribed speech. Gries (2009) put forward that corpus linguistics is a method(ology) adding to Leech's (1992) view of *computer corpus linguistics,* as he named it, as a 'new philosophical approach' (p. 106). McEnery and Wilson (1996) defined corpus linguistics as a methodology differentiating

it from other branches of linguistics, such as syntax and semantics requiring explanation and description. They added that while corpus linguistics doesn't constitute an area of linguistics in itself, it could be used in almost every area of linguistics. Hunston (2006) regarded corpus linguistics as a sophisticated method that seeks to answer linguistics questions and can be used to test the linguistics hypotheses and, can feed the quantitative dimension of linguistic studies. On the other hand, the pioneering researchers like Aarts (2002), Teubert (2005) and Williams (2006) suggested that the corpus linguistics is a discipline.

As can be seen from the definitions, there is no agreement among different pioneers of the field. However, Gries (2009) notes that the difference between the definitions would not have a negative impact or any difference with regard to the practical issues. Presenting a comprehensive overview on the definition of corpus linguistics, Taylor (2008) claims that the corpus linguist could be someone who designs, compiles or analyses corpus or who does all three and various definitions given to corpus linguistics "are to be welcomed".

This huge range of different definitions of corpus linguistics raises the question of what a corpus linguist does. Biber, Conrad and Reppen (1998) suggested that the Corpus Approach mainly holds four essential characteristics:

- It is empirical and analyses the actual language use patterns in natural texts
- It makes use of large and principled collection of natural texts, namely corpus for linguistic research
- It extensively utilizes computers for analysis
- It depends on both quantitative and qualitative analytical techniques.

Emphasizing the crucial role of computers in corpus linguistics, which made it possible to store large bodies of natural language and to analyse complex patterns in language use providing consistent and reliable analyses, Biber, Conrad and Reppen (1998) asserts that what corpus-based analysis do is much more than simple counts of linguistic features. They add that if one exploits a representative corpus properly, the corpus can present additional data on the *systematic* ways the linguistic items are used. (p. 4) Kennedy (1998) also lists the challenges in corpus linguistics as follows:

- How can we best exploit the data in machine-readable form?

- What linguistic theories will foster the structure of corpus based research?
- What linguistic features should we seek for?
- What applications can make use of description of languages based on corpus research? (p. 3)

Bennet (2010) focuses on two fundamental questions that the corpus linguistics attempts to answer:

1. What particular patterns are associated with lexical or grammatical features?

2. How do these patterns differ within varieties and registers?

Offering answers to the question of what corpus linguistics is, Bennet (2010) also explains what it is not summarizing her three points in that it is not able to provide negative evidence, explain why, and provide all possible language at one time. Although these three issues in corpus linguistics resulted in some criticism regarding its infeasibility, almost all areas of linguistics exploit the data extracted from corpus linguistics to gain more insight into understanding the nature of patterns of language use.

Addressing the definitions and function of corpus linguistics, it is quite clear that it can be applied to many fields of research such as language teaching and learning, lexicography, applied linguistics, discourse analysis, pragmatics, sociolinguistics, literary and translation studies, and so on. As McCarthy and O'Keeffe (2010) states, "Corpus linguistics has had much to offer other areas by providing a better *means* of doing things." (p. 7)

2.2.1 History of Corpus Linguistics

Since the emergence of corpus linguistics, it has gained ground despite the criticisms it has faced and a huge variety of definitions it has been offered as mentioned above. The criticism it has received mainly originates from the distinction between the principals of traditional linguistics or generative grammar and corpus linguistics. Biber, Conrad and Reppen (1998) state that language studies fall into two main areas being studies of structure and studies of use. While traditional linguistics emphasizes structure, a different perspective to language, which is also the main focus of corpus linguistics, emphasizes language use exploring how speakers and writers use language

patterns rather than looking at possible theories of language. Gries (2010) puts forward that the problematic relation between traditional linguistics and corpus linguistics result from the fact that there have been different views on the definition of corpus linguistics so as to answer whether it is a method, methodology, theory or discipline. He adds that corpus linguistics has some things that are less attractive to the theoretical linguists. Meyer (2002) highlights that in spite of the different perspectives the generative grammarians and corpus linguists take, corpora still has a lot to contribute to theories of language as it provides valuable resource to test the linguistic theories.

In 1940s and 1950s, corpus based approach was rising among American structuralists based on the idea that the linguist should study authentic language (Leon, 2005). However, the time when one of the first computer corpus (Brown Corpus) was being compiled in 1960s, was a period dominated by generative grammarians. And other approaches to language which weren't based on the principals of generative grammar were not tolerated. Corpus linguistics therefore was on the decline mainly because of Chomsky's criticisms to that new approach (Leech, 1992).

One of the earliest criticisms by Chomsky roots in the distinction between the rationalist and empiricist approach to language. The fundamental division between the two approaches is based on the decision whether to rely on introspective judgements on artificial data or on naturally occurring data through the use of corpus (McEnery & Wilson, 2001). According to Chomsky (1965), there are three levels of adequacy that the linguistic theories may be evaluated: observational, descriptive and explanatory adequacy. Corpus linguistics' aim is to achieve descriptive adequacy (a lower level of adequacy) while generative grammar aims to achieve the highest level of adequacy, which is explanatory, to reach the abstract principles which can be regarded as a part of Universal Grammar. Unlike generative grammar, corpus linguistics value descriptive adequacy more than explanatory adequacy as they seek for natural language use (Meyer, 2002).

Another influential criticism Chomsky made against corpus linguistics was related to the competence and performance, which have been replaced by I (Internal) and E (External) Language. Competence refers to our internalised knowledge of a language while performance is external evidence of language competence (McEnery & Wilson, 2001). Chomsky (1962) argued that the main work of a linguist must be to

study linguistic competence, but corpus doesn't provide an efficient start line for this goal. In Chomsky's (1962) words:

> Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so wildly skewed that the description would be no more than a mere list. (p. 159)

Despite the fact that generative grammarians did not favour corpus linguistics, there were some attempts to exploit corpus-based data for advancing generative theory. Aarts (1992) made use of Survey of English Usage (SEU) corpus to analyse small clauses that can be used in a clause. Nevertheless, corpora didn't take a substantial role in generative grammar; therefore, the evolution of corpus linguistics had slowed down because of Chomsky's influence.

Although the development of corpus was slow in 1970s, that was the period some major corpus work began with the idea of annotated corpora in excess of one million words, which Leech (1991) calls it as second generation of corpus. In the early 1960s, Randolph Quirk, one of Firth's pupils, started to plan and execute SEU corpus on spoken and written British English. However, it took a long time for it to be computerized. Henry Kučera and Nelson Francis developed the Brown Corpus containing a million words of American English, which is accepted as the first computerized corpus. In 1975, Jan Svartvik began to build London-Lund Corpus of spoken English. Following these corpora, a good deal of famous corpus work such as the British National Corpus (BNC) and the Lancaster-Oslo-Bergen corpus (LOB), and Collins Birmingham University International Language Database (COBUILD) projects were developed.

It was the advent of technology that accelerated the development of these major corpora. A large collection of texts and transcribed speech has been available with the development of corpus analytical tools and corpus software. Although first computerized concordances emerged in the late 1950s, corpus linguistics experienced a revolution in 1980s and 1990s with the concordancing programs widely used nowadays like Wordsmith Tools and Monoconc (1996) (Tognini-Bonelli, 2010). Sinclair (1991) highlights the essential role of computers in corpus-based work in following words:

> Thirty years ago when this research started it was considered impossible to process texts of several million words in length. Twenty years ago it was considered marginally possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today it is very popular (p. 1).

In the same vein, McCarthy and O'Keeffe (2010) put forward that technology 'has been the major enabling factor in the growth of corpus linguistics but has both shaped and been shaped by it'. To make a clear description of the progress that electronic corpus has made, Tognini-Bonelli, (2010) notes three stages of corpus linguistics as shown in Table 1.

Table 1. Three phases of corpus evolution

	Time-Span	**Major Drives and Gains**	**Corpus Characteristics**
Stage 1	The first twenty years 1960-1980	- learning how to build and maintain corpora of up to a million words - no material available in electronic form - everything has to be transliterated on a keyboard	-Small Corpus -Standard -Sampled
Stage 2	a) The decade of scanners 1980s b) the First Serendipidity 1990s	-a target of twenty million words becomes realistic - text becomes available as the by- product of computer typesetting -another order of magnitude to the target size of corpora	-Standard -Sampled -Multi-modal -Modern Diachronic Corpus -Dynamic -Open-ended data flow
Stage 3	The new millennium, and the Second Serendipidity 2000s	- text that never had existence as hard copy becomes available -unlimited quantities of data from the internet	-The Web as Corpus -Web texts as source of information

As can clearly be inferred from the information above, technology has had a huge effect on the evolution of corpora; it has also given way to development of different types of corpora as well.

Bennet (2010) distinguishes among three basic types of corpora indicating that it is important to know which one to analyse. Generalized corpora is the broadest type of corpus in more than 10 million words containing varieties of language to be able to reach generalized assumptions. This type of corpora can comprise of spoken or written language from a great variety of genres, one example being the BNC. Specialized corpora, differently from generalized corpora, contains data drawn from certain type of texts developed to provide an answer representative of that certain type. One example is the Michigan Corpus of Academic Spoken English (MICASE) containing spoken data drawn from individuals in an academic setting. Meyer (2002) stresses that focus can be drawn towards a fully representative collection of American academic language by narrowing down the scope of corpus. Another type of corpora is learner corpora that is defined as an electronic collection of authentic texts produced by foreign and second language learners (Granger, 2003). Granger (2003) suggests that given the diverse nature of corpora has enabled linguists to compare language varieties in spoken and written medium and in generalized and specialized corpora. One example is the International Corpus of Learner English, which is comprised by essays of learners with 14 different native language backgrounds.

Briefly stated, the relevant literature shows that corpus linguistics has a relatively long history and it has been a gradually growing discipline employed in a plenty of research fields over the decades. As McCarthy and O'Keeffe (2010) state, corpus linguistics has proven itself to be a healthy and vigorous discipline in language studies, with its revolution mainly owing to technological advancements in the late twentieth century. Despite the criticisms corpus linguistics has faced, it has been a very widely used approach expanding its scope with the development of different type of corpora. Being one major type, learner corpora can be claimed to provide a very influential basis for language learning and language acquisition studies.

2.3 Learner corpora

While corpus linguistics has a long story in a wide range of areas of research, it was only since the late 1980s and the early 1990s that the learner corpora (LC) have begun to be compiled which led to the recognition of the theoretical and practical value of computer learner corpora (CLC) by SLA researchers, corpus linguists, and applied linguists (Granger, 1998).

CLC can be defined as electronic collections of language data in general terms. Granger (2002) provides a more detailed definition:

Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardized and homogeneous way and documented as to their origin and provenance (p.7).

This definition involves a few key terms that need to be touched upon: authenticity, textuality, and explicit design criteria. Regarding authenticity, Sinclair (1996) points out that the data in corpora is authentic as it is compiled from the genuine interactions between people as they *go about their normal business* differently from the data elicited from experimental and introspective studies based on various artificial settings. As Granger (2012) points out, fully natural learner language data is very difficult to gather especially in foreign language learning classrooms. Experimental data types like fill-in-the-blanks activities are reported to fall outside the scope of learner corpora. Therefore, it should be noted that there are different degrees of naturalness or authenticity of language between the natural and experimental language data continuum, which Nesselhauf (2004) categorizes this continuum as *peripheral learner corpora* within the CLC framework. With regard to learner corpora, whereas the data gathered from free essay writing, for example, is somewhat controlled due to the time and topic limits that the learners get exposed to, they can be considered as authentic at the same time as it results from authentic classroom activity (Granger, 2002).

Textual data, as the second key word in Sinclair's (1996) definition, is considered to be a distinguishing feature of LC. Learner corpora are comprised of continuous stretches of discourse, not separate sentences or words. Therefore, to qualify as corpus, LC should involve both the correct and incorrect use of language by the learners (Granger, 2002). In addition, the roots of LC can be traced back to error analysis studies, which focus on decontextualized errors of learners and ignore the rest of their language production. As a result, Larsen- Freeman and Long (1991) note that error analysis researchers weren't able to access to the whole system of interlanguage. However, with the advent of recent learner corpora, researchers are now able to identify the 'deviation from the standard, i.e. the language of the native speakers of a particular

language' (Pravec, 2002 p. 81), so the researchers can investigate what's and what's not in a corpus.

Furthermore, when the variation in EFL and ESL is taken into consideration, a learner corpus should be compiled according to strict and clear design criteria due to many different types of learners and learning situations. As Tono (2003) asserts, 'if data is gathered in an opportunistic way without proper control and documentation of learner and task variables, the resulting corpus will be unlikely to be of much use" (p. 801). Additionally, concerning the current SLA studies, which traditionally rely on empirical data, Gass and Selinker (2001) draw attention to the absence of detailed information about the learners and the linguistic environment that the production was obtained. LC, therefore, stands as a very rich type of resource for SLA research if they are gathered according to strict design criteria. Table 2 presents the design considerations in building a learner corpus.

Table 2. Design criteria for building learner corpora

Types of feature		
Language-related	**Task-related**	**Learner-related**
Mode (spoken/written)	Data collection (cross-sectional/longitudinal)	Internal-cognitive (age/cognitive style)
Genre (letter/diary/fiction/essay)	Elicitation (spontaneous/prepared)	Internal-affective (motivation/attitude)
Style (narration/argumentation)	Use of references (dictionary/source text)	L1 background
Topic (general/leisure,etc.)	Time limitations (fixed/free/homework)	L2 environment (ESL/EFL – level of school
		L2 proficiency (standard test score)

Adopted from "Learner corpora: design, development and applications" by Tono 2003 p.800, Copyright 2003 by Ucrel

As seen in Table 2, there are three major categories in compiling learner corpora: 1) language-related, 2) task-related, 3) learner-related. Tono (2003) stresses that the researchers to compile a learner corpus for the first time should be guided very carefully since if the corpora are compiled following strict design criteria can be shared by others so as to reach sound findings.

Lastly, another important feature of LC is that they are computerized. It has been possible now, therefore, to collect large bodies of data, store it on the computer, and analyse it using the software tools that help better describe learner language (Granger, 1998). In addition, many early learner corpora targeted English and it took the researcher great time and effort to analyse the data. However, the inclusion of computers to LC research has made it possible to compile corpora in a more efficient and quicker way, and has expanded its impact area by including a number of different languages into learner language analysis (Granger, 2012).

2.3.1 Major Learner Corpora around the World

Although learner corpus is a quite young field in language research, its great potential in revealing substantial facts about learner language, testing the established SLA theories, and designing instructional tools and materials for the learners has now been acknowledged. However, the design of a learner corpus can change according to the purpose of compiling learner language data. Table 3 provides the types of processing data.

Table 3. Types of processing of learner data

Extra-textual information	Header information (learner/ language/ task variables)
Level of transcription	Orthographic (+ phonemic/ phonetic for spoken corpora)
Level of annotation	Sentence-boundary disambiguation
	Tokenisation
	POS tagging
	Lemmatisation
	Parsing (Treebanking)
	Semantic tagging (word senses/ semantic relationships and categories)
	Discourse tagging(apologies/greetings/politeness/?? moves/acts??/etc.)
	Error tagging
	Prosody annotation
	Anaphoric annotation

Adopted from “Learner corpora: design, development and applications” by Y. Tono, 2003 Proceedings of Corpus Linguistics p.801, Copyright 2003 by Ucrel

Granger (2002) highlights that the format that the learner corpora comes in must be standardized using software tools to make the comparison with the native corpora possible. In addition, learner and task variables should be documented clearly in learner corpora so as to enable researchers to gather sub-corpora.

So far, a good number of learner corpora projects have been compiled with some of them still in the process of development. Table 4 illustrates the major learner corpora projects around the world.

Table 4. Major learner corpora around the world

Learner corpus	Subjects/Task/Size	Annotation	Comparison
International Corpus of Learner English (ICLE)	-University EFL 3/4 year students -16 nationalities -Written essays -4,5 million	Error tagged Pos tagged	NNS vs NNS (different L1s) NS vs NNS
The Hong Kong University of Science and Technology Learner Corpus (HKUST)	-Chinese undergraduate students -Written academic texts -25 million words	Error tagged	NS vs NNS
Louvain International Database of Spoken English Interlanguage (LINDSEI)	-11 nationalities -3/4 year students -50 interviews + -100,000	orthographic	NNS vs NNS (different L1s) NS vs NNS
Cambridge Learners Corpus (CLC)	-All levels -10 million -Commercial	Pos tagged Error tagged	NNS vs NNS
Longman Learners Corpus (LLC)	-All levels -Written essays -10 million -Commercial	Pos Tagged	NNS vs NNS
The Corpus of Academic Learner English (CALE)	-Advanced level -Various academic text types that are typically produced in university courses of English	Under development	NNS vs NNS (different L1s)
LONGitudinal DAtabase of Learner English (LONGDALE)	-from intermediate to advanced -range of text types -longitudinal data -spoken and written	Under development	NNS vs NNS (different L1s)
USE Uppsala University, Sweden (USE)	-Swedish university students of advanced level -written academic texts	- Plain text	NS vs NNS
Chinese Learner English Corpus (CLEC)	-Chinese students from five L2 proficiency levels -written texts -1 million words	-Error tagged	NS vs NNS
The ISLE corpus of non-native spoken English	-20 minute speech -German and Italian intermediate learners of English	-Orthographic -Phone-stress	NS vs NNS
PELCRA University of Lodz, Poland	-Polish learners of English at different levels of L2 proficiency -written texts	- Plain text	NS vs NNS

As illustrated in Table 4, researchers have spared great time and effort to compile learner language corpora but the learner corpora projects presented above are just a small reflection of the learner corpora projects around the world. Being one of the biggest learner corpora, HKUST learner corpus consists of 25 million words of essays and exam papers written by Chinese undergraduate learners of English in Hong Kong. One of the best-known and largely used learner corpora is ICLE corpus. The corpus contains argumentative essays written by university students with around 4,5 million words in total. With 16 sub-corpora it has, it can be used in combination with LOCNESS to compare native and learner English. The error and POS-tagged version of the corpus is now available. Granger, Dagneaux, Meunier, and Paquot, (2009) notes that the strength of ICLE comes from the fact that it was built according to very strict design criteria. Being comparable to the written corpus ICLE, LINDSEI is the first spoken corpus of learner English. The corpus has 20 sub-corpora with 11 of them completed and it comprises of interviews with advanced learners of English from different L1 backgrounds. NS and NNS comparison can be made with native corpus LOCNEC. CLC and LLC are commercial corpora with around 10 million words in total. The Corpus of Academic Learner English (CALE) is a specialized corpus of academic learner writing which is still under development. The corpus contains a range of academic texts written by mainly German learners and learners from different L1 backgrounds. Native reference corpora are Michigan Corpus of Upper-Level Student Papers (MICUSP) or the British Academic Written English corpus (BAWE). Another corpus, which is under development, is LONGDALE project. The corpus is in both spoken and written mediums with longitudinal data gathered from intermediate and advanced level learners from different L1 backgrounds. The project was initiated in 2008 with 117 EFL undergraduate students in their first year and the purpose is to follow the same students over at least three years. In addition to CALE and LONGDALE projects, another corpus in progress is the PELCRA Learner English Corpus. The corpus is aiming to present around 3 million written and 200,000 spoken data. The corpus represents texts such as essays, letters, MA theses and many other text types written by Polish learners of English at different proficiency levels. The Uppsala Student English Corpus (USE) consists of 1489 essays written by 440 Swedish university students of English. The texts types included in the project are evaluation, argumentation, discussion, culture, and literature with around 1 million words in total. CLEC contains over 1 million words

with a collection of written compositions by Chinese learners from different backgrounds and at different proficiency levels with 5 subcorpora. Lastly, the ISLE speech corpus is another spoken non-native corpus available like LINDSEI. The corpus includes approximately 20 minutes of speech from 23 German and 23 Italian intermediate learners of English with a total of around 12000 utterances. Recorded sentences contains different task types such as reading simple sentences, using minimal pairs, and giving answers to multiple choice questions. LINDSEI and its native counterpart LOCNEC, which make the database of the present study, will be explained in detail in the next sections.

2.3.1.1. LINDSEI: Louvain International Database of Spoken English Interlanguage

The Louvain International Database of Spoken English Interlanguage (LINDSEI) is the spoken counterpart to ICLE. It was launched at the Université Catholique de Louvain in 1995, five years after the start of ICLE. The aim of the project is to gather oral data through interviews produced by advanced learners of English from a variety of mother tongue backgrounds. The corpus has now 20 components in total. 11 of the subcorpora in Bulgarian, Chinese, Dutch, French, German, Greek, Italian, Japanese, Polish, Spanish have been completed and the others in Arabic, Basque, Brazilian-Portuguese, Czech, Finnish, Lithuanian, Norwegian, Swedish, Taiwanese and Turkish are in still in the process of compilation (Gilquin, 2012). It has a comparable native corpus that enables the comparison between NS and NNS. The first component was gathered from transcripts of 50 interviews with French learners of English with about 100,000 words of learner language in total. The other components of the project follow the same structure with 50 interviews containing three tasks: set topic, free discussion and picture description. Figure 1 provides the variables in the design criteria of LINDSEI.

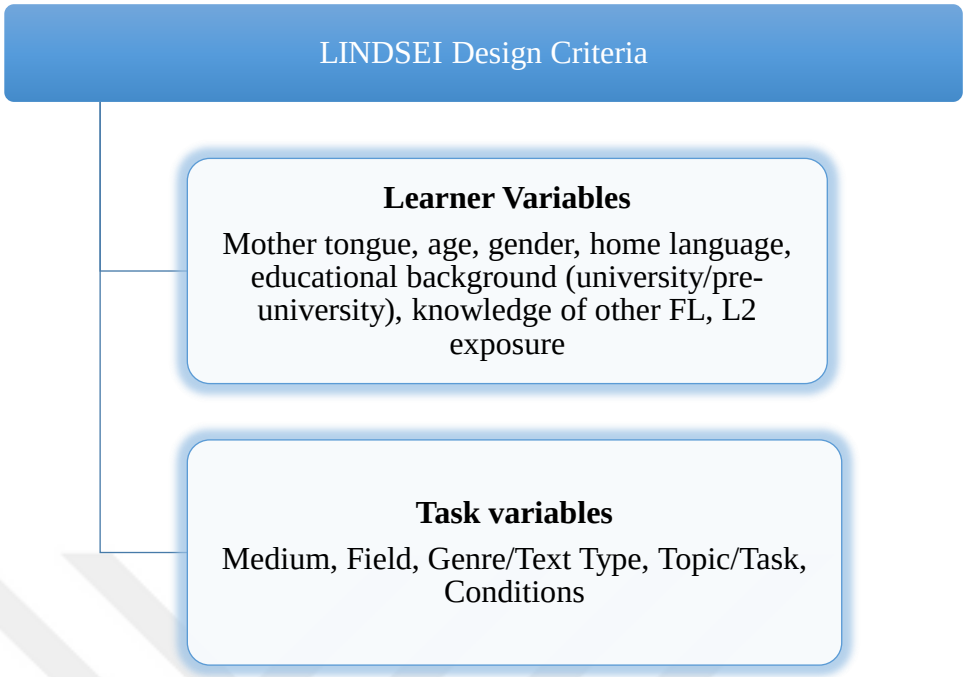


Figure 1. LINDSEI Design Criteria

LINDSEI-TR, the Turkish sub-component of LINDSEI, was gathered at Çukurova University. The spoken data compiled consists of 58 interviews with third- and fourth-year students at the Department of Foreign Language Teaching. The corpus totals 80,813 words and each interview lasts 12 minutes on average. LINDSEI-TR is explained in further detail in section 3.3.1.

2.3.1.2 LOCNEC: Louvain Corpus of Native English Conversation

The Louvain Corpus of Native English Conversation, LOCNEC is the comparable native counterpart to LINDSEI. The corpus includes 162,000 words of oral data in total produced by native speakers of English. In terms of compilation procedures, the same set of criteria regarding the format, the tasks design, recording and transcription were followed as in the case of LINDSEI corpus.

2.4 Stance: An overview

In the beginning of the twenty-first century, understanding the social and pragmatic nature of language, discovering the functions of it in the context it occurs have been an ongoing trend among researchers in linguistics and related disciplines. Stance, one of the linguistic mechanisms that reveals facts about the performative nature

of language has been relatively explored area in a wide variety of sub-fields of linguistics such as corpus linguistics, functional linguistics, discourse-functional linguistics, interactional linguistics as well as arising interests in closely related fields such as anthropology, social psychology, education and sociology (Englebretson, 2007). Although the term stance is not considered to be a monolithic term, the definitions offered to explain stance are quite broad and they change according to the individual backgrounds and interests of the researchers (Englebretson, 2007)

While in its broadest sense, stance, according to Biber and Finegan (1989) is "the lexical and grammatical expression of attitudes, feelings, judgements, or commitment concerning the propositional content of the message" (p. 124), it has been researched under different labels such as *evidentiality* (Chafe, 1986), *evaluation* (Hunston & Thompson, 2000), *hedging* (Hyland, 1996), *epistemic modality* (Hyland,1998) *affect* (Ochs, 1989), *appraisal* (Martin, 2000; White, 2001), *attitude* (Halliday, 1994). The labels referred in the literature have no clear-cut divisions and they seem to be overlapping in some aspects. For example, Hunston and Thompson (2000) refer to the judgements, feelings and viewpoint that a writer has about something using the term '*evaluation*'. Based on the definition offered by Biber and Finegan (1989), Englebretson (2007) comments that the definition proposes the subjective and evaluative nature of stance. Given the explanations, it can be clearly seen that the labels encapsulate each other.

Biber (2006) claims that stance expressions reveal different types of personal feelings and attitudes that the speaker has about the information given, certainty about the propositional content of the message, the way the speaker accessed the information and also the perspective they are taking towards the given utterance. Highlighting the importance of the need to find out the underlying principles of stance taking and to reach an agreement on its meaning, Du Bois (2007) regards stance "as a linguistically articulated form of social action whose meaning is to be construed within the broader scope of language, interaction, and sociocultural value" (p. 139). He extends the discussion providing a more detailed definition to stance taking from a discourse-functional view:

Stance is a public act by a social actor, achieved dialogically through overt communicative means, of simultaneously evaluating objects, positioning subjects (self

and others), and aligning with other subjects, with respect to any salient dimension of the sociocultural field. (p.163)

As it is clear from the definition, stance act includes three kinds of stance in itself. When taking stance, the speaker first evaluates the object, then positions the subject, and as the third step, aligns with other subjects. Figure 2 sums up the interactional acts performed when taking stance.

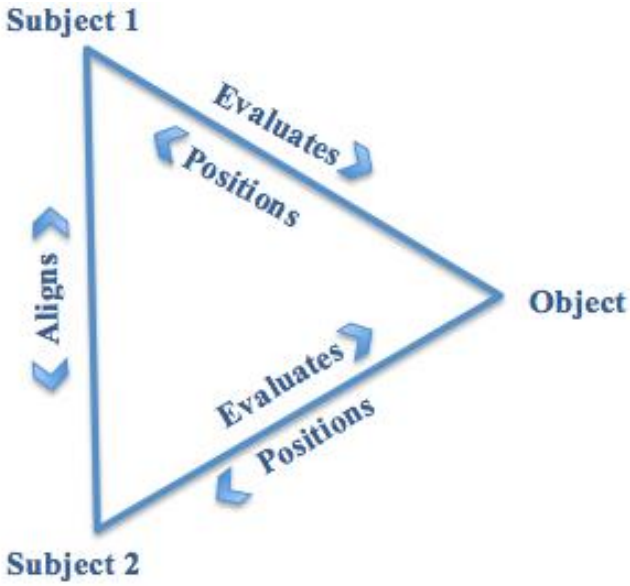


Figure 2. Stance Triangle. Adopted from DuBois (2007)

Similarly, Haddington (2013) defines stance as "the ways in which speakers talk and use language to express different kinds of positions, attitudes, and, feelings". The definitions of stance display a similarity in terms of the key elements they contain. However, when the linguistic form and the interactional effect are combined, stance can come into being in different forms (Kirkham, 2011). Three categories of stance adverbs that are put forward by Biber (2006) are *epistemic*, *attitude,* and *style*. According to this categorization, epistemic stance are divided into two subcategories as certainty (e.g. certainly) and likelihood (e.g. perhaps, probably). Attitude adverbs are related to the evaluation or assessment conveyed (e.g. importantly, amazingly), and style adverbs refer to the manner of giving information or the perspective that the interlocutor takes (e.g. frankly). As regards modal verbs, Biber (2006) proposes three categories as well related to their meanings: possibility/permission/ability, necessity/obligation, and prediction/volition. On the other hand, Kiesling (2009) puts forward a division of stance

into epistemic and attitudinal domains. Epistemic dimension relates to a person's certainty about his or her assertions and attitudinal stance refers to the way a person expresses his or her relation to the interlocutor which he calls the interpersonal stance. Kirkham (2011) notes that the two types of stance are linked to each other and they can occur together.

The variations in the definition of stance are also reflected in the appraches to the analysis of stance. Some recent research into stance in sociocultural terms has focused on social and cultural identity construction through the use of stance markers (Bucholtz & Hall, 2005; Olinger, 2011; Jaffe, 2009; Park & Takanashi, 2011, Morek, 2015). As Ochs (1996) argues, the linguistic structures that are deployed as epistemic and affective stances are regarded as the main linguistic resources so as to realize social acts and construct social identities. Another group of research into interactional linguistics studies the intersubjective nature of stance (Fitzmaurice, 2004; Almeida & Ferrari, 2012; Stone & Thompson, 2014; Precht, 2003; Keisanen, 2007; Kärkkäinen, 2006). Studying on the role of tag questions and yes/no interrogatives in stance taking, Keisanen (2007) concludes that stance is an interactional and intersubjective act rather than being subjective. Jaffe (2009) also contributes to this idea by noting that "stances are construed across turns", therefore, they are not the product of a single turn.

In addition to the studies conducted on stance by sociolinguistics and interactional linguistics, corpus linguistics has also greatly contributed to the identification and description of stance markers. One of the best known examples is by Biber and Finegan (1988) investigating stance adverbials in English through the use of LOB and London-Lund corpora containing 410 texts of written and spoken British English. Including six categories labelled as 1) honestly adverbials, 2) generally adverbials, 3) surely adverbials, 4) actually adverbials, 5) maybe adverbials, and 6) amazingly adverbials, the study made use of cluster analysis in order to detect the similarity and difference between the clusters in terms of their use of adverbial stance markers. The study presents two important findings. The first finding suggests that the use of adverbials connected to their stance style differ from the function regarding their literal meanings. The second finding puts forward that the frequency of stance markers is linked to their more or less specialized functions. That is to say, a frequent use of an adverbial stance seems to be connected with a generalized function while a less

frequently used adverbial associates with a more specialized function. Biber and Finegan (1988) add that the same stance styles are deployed in different genres for different purposes.

Although Biber and Finegan's (1988) study investigated stance in both spoken and written corpus, stance in academic research writing rather than in spoken language has been a dominating research area. Gathering the data from a corpus of American academic articles written in the fields of history and economics, Silver (2003) investigates the adverbial *evidently* both as a hedge and booster. The two corpora contain around 4 million words in total with the occurrence of the adverbial *evidently* 13 times in each. Taking *Longman Grammar of Spoken and Written English* (LGSWE) as a comparative model, the study concludes that while the adverbial often appears in economics corpus as a booster, in history corpus, it often appears as a hedge. Highlighting the need to analyse much larger samples to identify the hedge-booster distinction, Silver (2003) suggests that the use of the adverbials differs significantly in different disciplines.

Hyland (2005) sets out to study stance and engagement in academic research articles through hedges, boosters, attitude markers, and self-mentions. Making use of a corpus consisting of 240 research articles in eight disciplines, it is found out in the study that hedges are the most frequently used feature in writer positioning, and also the disciplines in the humanities and social sciences tend to take more involved and personal positions than the science and engineering related fields. Hyland (2005) also stresses the role of corpus to investigate stance by noting that 'the value of exploring such a large corpus is that it makes available many instances of the target features in a naturally occurring discourse' (p. 178).

Conducting a comparative study of stance in Chinese EFL learners of English, Jiajin and Manying (2008) investigate the epistemic, attitudinal, textual, and deontic stance markers used in argumentative essays of a shared topic written by 61 Chinese students. The corpus comprises two essays from each participant, being one essay in English and the other one in Chinese. The results of the study display the similarity of the use of stance markers in English and Chinese, stressing the L1 influence in writing of the learners as the analysis found out that the learners tend to use literal translation from their L1.

Aull and Lancaster (2014) examine stance markers gathering data from the University of Michigan and Wake Forest University. The first-year students' corpus contains 4032 essays from two universities. Upper level student data were obtained from Michigan Corpus of Upper-level Student Papers (MICUSP) and papers from Wake Forest University contained 615 papers, and these are compared with the scholarly written data obtained from Corpus of Contemporary American English (COCAA). Findings show that first year students tend to make use of approximative hedges, code glosses, concessions, and contrast expressions less frequently than the two advanced group while they overuse boosters and adversative connectors.

Even though above mentioned studies shed light to the understanding and identification of stance in written language, there have also been attempts to explain stance taking in spoken language.

2.4.1 Stance Markers in Spoken Language

Johnstone (2009) defines stance in terms of spoken language as follows: "Stance is generally understood to have to do with the methods, linguistic and other, by which interactants create and signal relationships with the propositions they utter and with the people they interact with" (p. 3). Although this definition is valid in written language from the writer-reader perspective, the features of stance taking acts in spoken language differ from the written language. Berman, Ragnarsdottir and Strömqvist (2002) establish that the differences between spoken interaction and writing stems from a number of factors such as "duration of the physical signal, the possibility of on-line feedback, mutual adaptation between sender and receiver, and the distribution of expressive features (words, tone of voice, gestures, facial expression, etc.)' (p.274). All these given characterizations result in different constructions of stance in discourse. While long-lasting feature of the signal in writing eases the process of metalinguistic reflection and long-distance editing, on-line feedback and mutual adaptation possibilities in spoken medium allows for a greater degree of recipient orientation (Berman, Ragnarsdottir & Strömqvist ,2002). This claim urges the need to investigate stance in spoken language as well as in written language. In this line, plenty of research has been carried out to provide a better and deeper explanation to the term stance.

Exploring the stance and evaluation in both spoken and written university registers, Biber (2006) studied on modal verbs, stance adverbs, and stance complement

clauses drawing the data from a sub-component of the TOEFL 2000 Spoken and Written Academic Language (T2K-SWAL) Corpus which contains 2.7 million words in total. The study focuses on classroom teaching and classroom management for the spoken part, and textbooks and course management for the written part. The study reveals a number of findings. The first finding suggests that stance markers are more frequently used in spoken language than in written language. The second finding shows that some stance features like prediction modals are connected to the management registers, regardless of spoken and written differences. The third finding reports that epistemic devices are more predominant in classroom teaching, but the directive meanings are predominant in classroom management. The last finding shows that stance devices in textbooks are relatively rare. They display no indication of personal attitude which Biber and Finegan (1989) calls as a 'faceless stance'.

In another corpus-based study, Precht (2003) examines stance moods in spoken British and American English. The study is based on the spoken section of the Longman Corpus of Spoken and Written English (LCSWE), which contains 1,5 million words for American conversation, and 1,4 million words for British conversation. The context for the oral data is comprised of conversations among related and non-related adults, conversation among family members, and conversations at work, with adult conversations making up the biggest proportion in the chosen corpus. Conducting multi-dimensional statistical analysis and factor analysis to find patterns in the use of stance markers, Precht (2003) suggests three moods of stance based on three factors: informal affect, boulomaic planning versus small talk, and hedged opinion. The study concludes that the British and American seem to have different socialization patterns in expressing evidentiality and affect. While the Americans deploy affect markers more frequently, the British tend to use more evidentiality markers to indicate involvement in general conversations among adults. This finding also stresses that it may lead to cross-cultural miscommunication because of different interpretations of the stance devices due to dialectic differences.

In a different study, Cheong (2012) researched stance features in L2 learners' oral presentations compiling a small corpus of classroom presentations of seven students at advanced level studying in an intensive English program. The presentations per each participant took around 7-9 minutes long. The analysis of the corpus indicates

that the speakers tend to use the same set of stance markers, showing a preference to use strong stance expressions. In addition, the study illustrates that the L2 students use stance markers to achieve three meta-functions: to structure their oral discourse, to interact with visuals, and to express their value judgements and feelings.

Conducting a study on spoken American English, Keisanen (2006) focuses on tag questions and negative yes/no interrogatives to identify stance use in interaction. Drawing the data from Santa Barbara Corpus of Spoken American English (SBSCAE), 42 speech events which last about 16 hours in total, with face-to-face interactions making up the biggest part in corpus are analysed. The results of the study imply that stance is an interactional achievement, rather than being subjective and also negative yes/no interrogatives. It also establishes that tag questions contain a high amount of linguistic and semantic material that can be exploited as recourse for showing the interrogative speakers' evaluative, affective or epistemic stance and there seems to be a quite restricted set of linguistic devices so as to form negative yes/no interrogatives and tag questions as they tend to appear in the present tense and to have third person singular subjects.

All the studies mentioned above have significantly contributed to the research on stance taking pointing out the different features of stance taking acts in spoken language which help identify the differences between written and spoken language. Epistemic stance, which is the focus of the present study, has been one of the most broadly explored categories of stance. The following two sections will introduce the epistemic stance and corpus based investigations of the term in L2 context.

2.4.2 Epistemic Stance

Every statement we make conveys not only the propositional content, but also it indicates the degree of certainty and uncertainty we have about the information being communicated (Brezina, 2009), and this relates to our epistemic stance. In speech, along with the propositional content of the message, other meanings attached to our system of values and identity are also communicated with our interlocutors (Hunston & Thompson, 2000). According to Du Bois' (2007) stance triangle model, when we take an epistemic stance, we first evaluate the information transmitted, then position ourselves, and finally align with the hearers. Regarding epistemic stance as a part of evidentiality which is seen as intertwined with epistemicity, Hunston and Thompson

(2000) note that the three basic functions that epistemic markers have in discourse are 1) expressing opinion, 2) maintaining relations between writer-reader / speaker-hearer, and 3) organizing discourse. Although the concept of modality or stance encircle numerous sub-types such as alethic (referring to the notion of truth), epistemic (relating to knowledge and belief), deontic (relating to duties) and boulomaic (referring to desire), epistemic modality is regarded as vital and central in naturally occurring interaction, and in especially spoken interaction (Brezina, 2009; Kärkkäinen, 2003).

Along with the various functions ascribed to epistemic stance, different definitions have also been proposed. In Coates' (1990) terminology, epistemic stance refers to the speaker's (lack of) confidence in the truth of the proposition communicated in the utterance. Biber and Finegan (1988) define epistemic stance as "the degree of certainty or evidence towards the content" (p. 30). Kärkkäinen (2006) suggests that epistemic stance relates to "marking the degree of commitment to what one is saying, or marking attitudes towards knowledge" (p. 19). Roseano, Gonza´lez, Borra`s-Comes and Prieto (2015) argue that epistemicity refers to the "degree of commitment a speaker has in the truth-value of a proposition" (p. 2).

Elaborating on the difference between epistemic status and epistemic stance, Heritage (2012) claims that epistemic stance can be conveyed through a rich variety of grammatical attainments of propositional content providing three example sentences:

1) Are you married?

2) You're married, aren't you?

3) You're married. (p. 6)

He goes on to argue that the three sentences exemplified above transmit the same message; however, the epistemic stance in the three sentences differs due to their different grammatical encodings. He explains that while the first sentence insinuates the questioner's having no knowledge about the marital status of the recipient, the second and the third sentence propose the questioner's "increasing commitment" to the possibility of recipient's being married.

According to Kärkkäinen (2006), spoken interaction is dominated by the expressions of epistemic stance, that's to say, epistemic stance is marked more frequently than attitudes, evaluations, expression of personal feelings, and emotions.

Therefore, numerous studies have devoted effort to explain epistemic stance in more detail in different contexts.

One of the most prominent studies is conducted by Biber (2006), who investigates the linguistic expression of stance both in spoken and written university registers through a sub-component of T2K-SWAL corpus and he proposes an important finding that stance is marked more frequently in spoken registers than the written registers. In addition, he provides the major grammatical stance devices deployed both implicitly and explicitly in interaction. Figure 3 shows the epistemic stance markers put forward by Biber (2006).

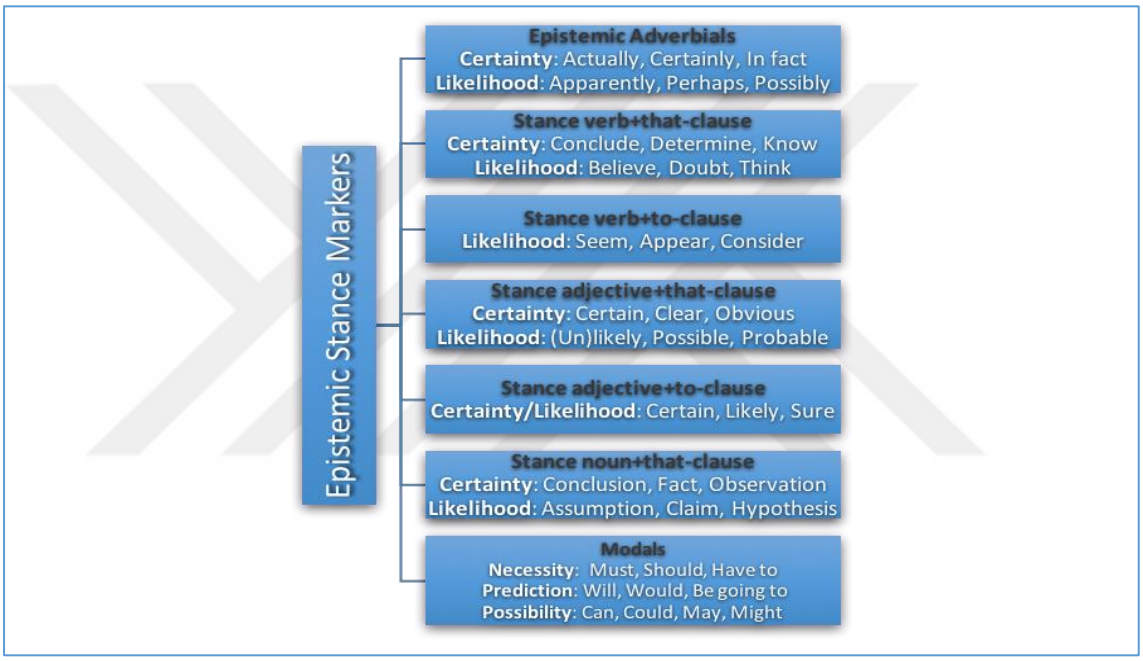


Figure 3. Epistemic Stance Markers. Adapted from "Stance in spoken and written university registers" by D. Biber, 2006 p.101, Copyright 2006 by Elsevier

In another study investigating native English conversation, drawing her data from Santa Barbara Corpus of Spoken American English, Kärkkäinen (2003) provides the most frequent grammatical classes that reflect epistemic stance in her database. Figure 4 presents the grammatical classes and the percentage of occurrences of epistemic markers in native American English speech.

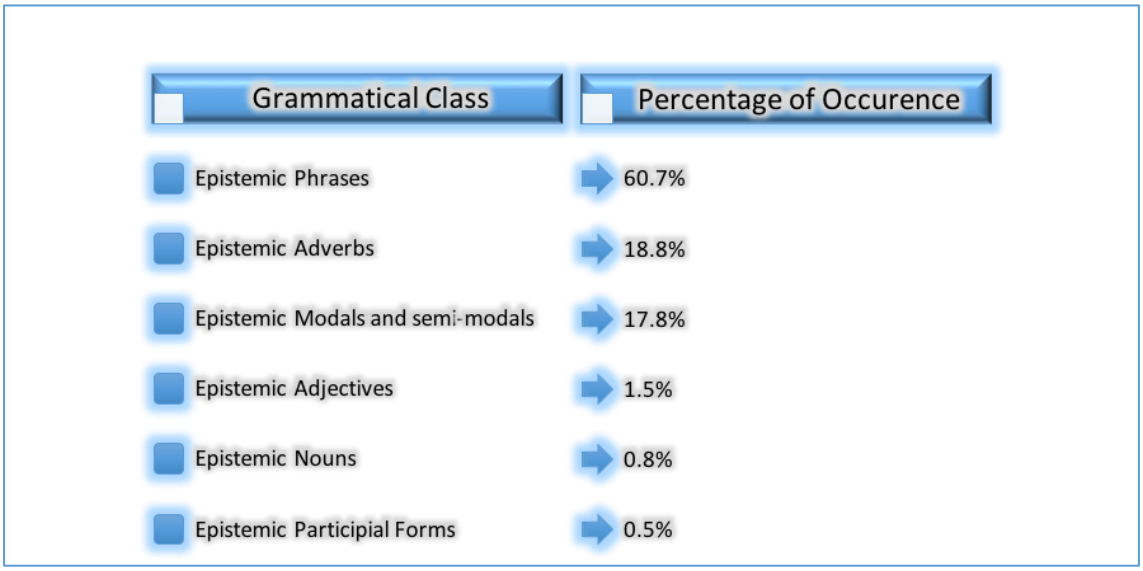


Figure 4. Grammatical Classes of Epistemic Stance in Order of Frequency. Adapted from "Epistemic Stance in English Conversation" by E. Kärkkäinen, 2003 p.37, Copyright 2003 by John Benjamins B.V.

As shown in Figure 4, epistemic phrases are the most common group, with the epistemic adverbs and modals following it. However, the epistemic adjectives, nouns, and participial forms are quite rare in the database used. The corpus-based study indicates that native American English speakers have a tendency to develop highly regular and routinized ways of epistemic stance taking.

Comparing general and academic English, Brezina (2009) explores the epistemic stance markers in speech using two corpora: British National Corpus for general English, and MICASE for academic English. Extracting a sub-corpus of highly interactive language, namely dialogues, from each corpus, the four epistemic markers (must, should, certain, and certainly) are analysed so as to evaluate Halliday's (1994) famous statement "The importance of modal features in the grammar of interpersonal exchanges lies in an apparent paradox on which the entire system rests – the fact that we only say we are certain when we are not" (p. 362). According to the results of the corpus analysis, Brezina (2009) comes up with three main findings: 1) in face-to-face interaction, it is very likely that every utterance is epistemically marked, 2) the certainty markers in question are linked to uncertainty, 3) the notion of uncertainty relates not to the (psychological) subjective uncertainty, but reflects intersubjective uncertainty.

All three significant studies in the relevant literature offer profound insight into the nature of epistemic stance in terms of native speech and provide strong basis for the well-established idea that corpora can be used effectively to investigate the phenomenon of stance (Hunston, 2007).

Exploration of stance on L1 has gained ground considerably for understanding the phenomenon. Mostly inspired from the studies in L1 context, L2 context also offers a promising area to clear the path to reveal more features of stance. As stated by Gablasova, et al. (2015), there is still a need to get more information about how L2 speakers employ epistemic stance markers in spoken discourse.

2.4.3 Epistemic Stance Markers in L2 spoken Corpora

Despite the fact that corpus-based epistemic stance by L2 speakers has been mainly researched in written language, (e.g. Wharton, 2012; Chen, 2012; Ağçam, 2014) there have also been some studies, though limited in number, to explore epistemic stance markers in L2 spoken corpora.

Accordingly, one of the first leading studies is carried out by Kärkkäinen (1992), which analyses epistemic stance devices used by Finnish learners of English through a contrastive study. She investigates epistemic modality through a corpus of 48 simulated task-oriented conversations between Finnish learners of English and a native speaker of English, which totals around 7 hours and 75,000 words. To enable the comparison, a small corpus of Finnish-Finnish and English-English recordings are compiled. The findings of the study highlight that non-native speakers employ fewer epistemic stance expressions when compared to the native group and also the expressions preferred by non-native group of students differ from the ones that the native group favours. As a pedagogical issue, she emphasizes the importance of teaching learners how to exploit the epistemic devices in interaction.

In a different EFL context, Letica (2009) aims to compare the epistemic modality used in Croatian and English spoken corpora compiled from the same students. The first corpus is gathered from the picture descriptions tasks in English produced by 33 Croatian students, and the second corpus is compiled from the same type of task by the same students in Croatian. Both corpora contain approximately 18,000 words in total. As for the frequency and range of epistemic markers in both

corpora, the study reveals that the participants make use of fewer epistemic devices in their L2 than in their L1, while the range of epistemic devices are found to be slightly wider in English than in Croatian. Correlational analysis implies that the students that use epistemic markers more frequently in L1 tend to use these markers in their L2 more frequently. Regarding the grammatical distribution of epistemic markers, the study finds out that modal auxiliaries are the least frequent class in their data, which is in contrast with the findings of previous research (Kärkkäinen, 1992). In addition, no significant correlation between proficiency as another factor in the use of epistemic markers and the range of devices is found.

As for the Swedish context, Aijmer (2009) studies *I don't* know and *I dunno* through the Swedish component of LINDSEI and its comparable counterpart LOCNEC. Arguing that the pragmatic markers are multifunctional and they need to be evaluated in their immediate context with regard to epistemic and affective stance, speech management and politeness, the study concluded that while non-native speakers often use *I don't know* as a speech management signal, native speakers prefer to use the marker not to ask questions in a direct way.

A more recent study by Gablasova et al., (2015) analyses adverbial, adjectival, and verbal epistemic stance expressions using a sub-component of the Trinity Lancaster Corpus of Spoken L2 introduction. The study contains 132 advanced level L2 speakers from various L1 backgrounds who participated in four different speaking tasks: presentation, discussion, interactive task, and conversation. In connection with the types of epistemic stance, most occurrences were found to be verbs. When the effect of task was taken into account, some differences were also reported in the distribution of epistemic markers across tasks, with the largest difference found between the monologic and dialogic tasks. Depending on the results of the study, the authors point out that the epistemic stance markers used by advanced L2 speakers adjust their epistemic stance according to both social and individual requirements of the tasks.

As for the Turkish EFL context, corpus-based epistemic stance related research through learner corpora remains relatively limited to some studies with a focus on written interlanguage (Can, 2009; Kilimci, 2009, Ağçam, 2014;) with an exception of a comparative study on recurrent word combinations in both spoken and written learner corpora (Şahin-Kızıl, 2013; Şahin-Kızıl & Kilimci, 2014).

In this respect, the literature outlined above implies a need to further investigate the use of epistemic stance markers in L2 in a more systematic way making use of the potentials of corpora, with a particular emphasis on spoken language.

2.5 Corpus and Language Teaching

Corpora as large systematic and computerized collections of naturally occurring written/spoken language have been acknowledged to have an undeniable impact not only on different branches of linguistics but also on SLA.

Myles (2005) concerns the relationship between SLA and corpora in the following words:

> If we are to make generalizations about learner development, we need to be able to capture the various stages that learners go through. For this we either need longitudinal data of a number of learners over a lengthy period of time, or we need very large cross-sectional datasets, so that the number of learners in each well-defined stage is big enough for us to be confident that the results of the analysis are generalizable (p. 374).

The lines quoted above emphasize that corpora have potential to describe interlanguage with the technological advances that made compilation, storage and management of corpus data relatively easy.

In terms of language teaching, the idea that the language learners should study genuine examples of language have gained acceptance. Sinclair (1990) contributes on the idea by noting that: “I am convinced that it is essential for a learner of English to learn from actual examples, examples that can be trusted because they have been used in real communication”, which stands in stark contrast to the generative grammarian approaches to language teaching. In spite of the acceptance of the potential of corpora in language teaching, the applications of corpus based research in the language classroom have remained limited. In Römer’s (2006) words: “Despite the progress that has been made in the field of corpus linguistics and language teaching, the practice of ELT has so far been largely unaffected by the advances of corpus research” (p.121).

Meunier (2011) proposes four reasons for the lack of native corpus-based methods and tools in instructional settings. The first reason stems from the lack of interrelation between linguists and teachers. Meunier (2011) extends on the reason

claiming that this may be because of teachers' being unaware of or not trained about the possibilities of corpora, or because of *ivory tower effect*, which concerns the perception that "linguists work in their offices and have no idea of what teaching is about". The second reason is that there are too many influential approaches to second language learning and corpus-based method is just one of them. The third one relates to the frequency, effects of which have been made clear in SLA, but far less clear in language teaching. Finally, the last reason presented is due to the "lack of empirical studies exploring the actual impact of corpus methods on the learning outcomes".

Although there is a lack of corpus-based applications in language learning and teaching, their effect on pedagogical areas is not totally invisible. Recently, a wide range of corpus-based reference works dictionaries, books on grammar and vocabulary based on native corpus are available to the teachers and students. However, in spite of the positive effect of the native corpus research on language teaching, it has been claimed that it cannot guarantee a completely successful EFL learning and teaching as native corpora do not provide the possibility to show which words and structures are difficult for learners and to what degree (Granger, 1998). With regard to instructional applications of learner corpora, Granger (2009) asserts that learner corpus research has not achieved its goal yet since it has brought up quite few pedagogical applications. De Cock (2010) argues that this is especially true when it comes to spoken learner corpus research and adds that a lot more research has to be conducted on spoken learner corpora so that spoken learner corpus informed teaching materials can be developed.

Römer (2008) puts forward that the use of corpora in language teaching contains the use of corpus tools and corpus methods, which leads to a distinction between the direct and indirect pedagogical applications. While the indirect applications can answer questions what and when to teach, direct applications can be used so as to "assist in the teaching process". Figure 5 illustrates the indirect and direct uses of corpora in language teaching.

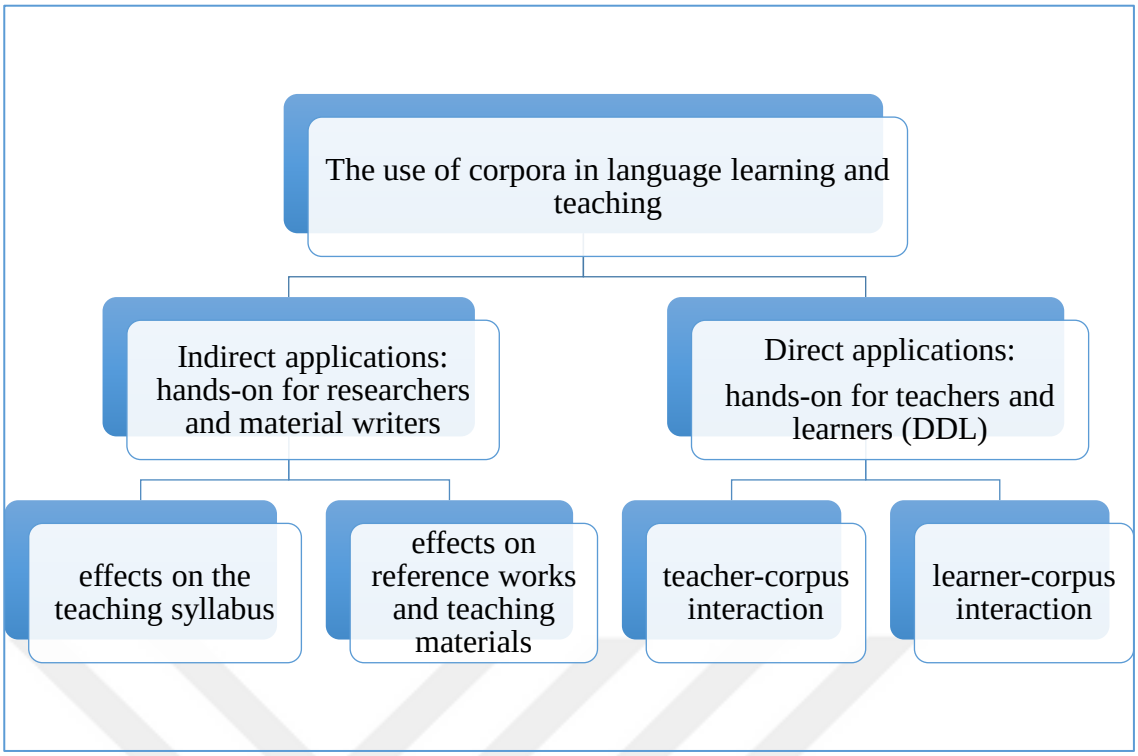


Figure 5. Applications of Corpora in Language Teaching. Adopted from "Corpora and Language Teaching" by U. Römer 2008 Corpus Linguistics. An International Handbook (Volume 1) p.113, Copyright 2008 by Walter de Gruyter

As shown in Figure 5, indirect applications of corpora relate to teaching syllabus, reference works, and teaching materials with access to corpora by researchers and material writers, while direct applications are available to teachers and learners through Data Driven Learning. Leech (1997) suggests three focuses to the direct uses of corpora: "teaching about", "teaching to exploit", and "exploiting to teach". McEnery and Xiao (2010) state that "teaching about" and "exploiting to teach" are regarded in connection with the students of linguistics and language programs, "teaching to exploit" refers to all language learners "who are expected to benefit from the so-called DDL or discovery learning".

DDL which constituted the methodological source for the current study, is an approach developed by Tim Johns in 1991. In this approach, through the use of concordance lines, learners discover facts about the language being learnt from the authentic examples of the target language in the corpus data. As Johns (1991) sets forth on the use of DDL in classroom, learners "often notice things that are unknown not only to the teacher, but also to the standard works of reference on the language" (p.3). Sripicharn (2004) conducts a study to compare the strategies employed by native and

non-native speakers working on the same concordance material and finds out that while native speakers tend to rely on their intuitive knowledge and ignore the concordance data, L2 learners of literature performed data driven strategies when they analyse language features through concordance.

Apart from DDL, another corpus-based approach to language teaching is the "three I's" method, (Illustration, Interaction, and Induction) which was proposed by Carter and McCarthy (1995) instead of "three Ps" (Presentation, Practice, and Production) as the former is considered to be more compatible with more corpus based applications to teaching. In three I's method, Illustration refers to looking at real data, interaction refers to discussion and exchange of ideas and observations, and Induction refers to making a rule for the target feature (McEnery &Xiao, 2010). Highlighting the importance of spoken language in syllabus, McCarthy (1998) argues that the L2 learners should be exposed to authentic spoken data to be fluent speakers of the target language, but three P's method doesn't allow the learners to understand functions of the language items in discourse.

However, DDL and three I's method have been reported to cause some problems with the learners as the approaches adopt an inductive approach, (Meunier, 2002) and inductive teaching may not appeal to some students due to the different learning styles they have. Especially for the work on phraseology through concordance lines by learners, Flowerdew (2009) extends the three I's method by offering a four I's model, where Intervention as a stage between Interaction and Induction is added. In doing so, Flowerdew aims to eliminate the difficulties that the students may encounter in an inductive teaching process. Intervention stage therefore involves some hints or prompting questions by the teacher such as "do you notice any difference between x and y?". Flowerdew (2009) comments that rather than seeing deductive – inductive approaches as contradictory, Intervention in the form of clues and prompts can be provided to students in order to both mediate between inductive-deductive approaches and lower the difficulties that the learners have when interpreting concordance lines. The approach with the fourth stage constitutes the basis for the instructional part of the present study.

As it is clear, the two methods presented above favours the use of concordance lines in language teaching. Tribble (2012) suggests however that the use of concordance

data in the classroom necessitates mediation by the teacher and the lines to be provided to the learners should be selected carefully to enhance learning. In the same manner, Meunier (2011) states that if the corpus methods are to be integrated to classroom, the activities must be "learner-centred, context-dependent and culture-bound".

In sum, it can be claimed that the use of corpora in language teaching has a promising potential since it has had a huge informative effect on syllabus design, the material writer and also to the teacher and the learner. In addition, it has been acknowledged that corpora and particularly learner corpora, can inform and facilitate language learning and teaching.

2.5.1 Teaching Stance Markers

Despite the role of stance markers both in written and spoken language has been well-attested in the extant literature, research on pedagogical applications of it have been relatively neglected and restricted to a limited number of studies (Tenuta, Oliveira, & Orfano, 2014).

Conducting a corpus based analysis on the Chinese EFL learners' use of stance in writing, Jiajin and Manying (2008) note as a pedagogical implication that the learners must be taught stance taking strategies according to the customs of target language and they should learn the vocabulary specific to their own field of study. In terms of spoken language, Wierzbicka (2006) puts an emphasis on the teaching and learning of epistemic modality with the following words:

Given the role of English in the contemporary world, we must also recognize the needs of learners of English all over the world, who, to be proficient in English, have to learn not only new words and grammatical constructions but also new meanings and new ways of speaking. They have to learn that the need to qualify one's statements and differentiate one's degree of "epistemic commitment" is greater in English than in most other languages and that "speaking English with the appropriate degree of conviction" (Holmes 1983) can present a great difficulty for students of English as a second language. (p. 251)

One of the few studies aimed to observe the instructional effects of teaching stance to L2 learners is carried out by Fordyce (2014). He seeks for the effects of explicit and implicit instruction on 81 Japanese EFL learners' use of epistemic stance.

The sample size for the implicit group is reported as 44 and for the explicit group 37. The participants were also divided according to their proficiency levels. The study lasts four weeks which contained two 45 minute sections for each week, the first two weeks dealing with the genre of opinion in both spoken and written language, and the latter two weeks dealing with the genre of description again in both medium. After given implicit and explicit treatments, two tasks were produced by the learners: writing a description of a photograph, and writing their opinion on a discussion topic. Conducting a pre-test, post-test and a delayed post-test on the students, the results of the study reveal that explicit intervention seems to have a larger impact on learners' use of epistemic stance markers than the implicit intervention. It is reported that no significant correlation between proficiency levels and explicit and implicit groups was found. In addition, the author emphasizes the role of *noticing* in pragmatic developments of L2 learners.

In a different study, Hita (2008) presents a sample lesson plan to teach modality. Based on the systemic functional linguistics framework, enabling learners to practice all four skills, giving oral production a priority, the paper attempts to teach the communicative and functional aspects of English. Additionally, the study highlights that different types of modality should be introduced to the learners explicitly at different stages of proficiency and adds that modality is a widely ignored part of language by teaching methods of EFL. Previous research not on particularly stance, but on L2 pragmatic development shows a tendency to favour explicit interventions (Takanashi, 2001; Li, 2012).

Regarding the integration of corpus linguistics in language teaching, Almeida, Brito and Toledo (2014) suggest a method of teaching epistemic modality and evidentials in scientific language through the use of *evycorpe*, which is a native corpus. The steps applied in teaching are as follows: the students are required to identify stance expressions from the corpus, then they share opinions, then they are required to complete a set of creative tasks. Presenting some examples of corpus-based activities to teach and learn stance, the study points out that through a corpus-based methodology, students are able to understand epistemic stance markers and evidentials, and they can achieve to employ these markers in a communicative setting. In terms of hedges, specifically, Ko (2014) analyses the clause, adverbial and adjectival hedges frequently

used by native speakers of English through COCA and GloWbE corpora. The study concludes that practitioners should made use of corpus linguistics as a methodology in class.

When the above-mentioned studies are taken into account, it can be clearly seen that the relevant literature implies a great need to further investigate the effect of teaching stance markers as they constitute a large part of L2 pragmatic development. This becomes even truer when it comes to research on teaching stance markers through learner corpora.

2.5.2 Teaching Stance Markers through Learner Corpora: A Meeting Point

As Nesselhauf (2004) states "most learner corpus studies claim to have some relevance for language teaching and typically end with a short section stating pedagogical implications of the study". This seems to be a valid claim for the available research carried out on teaching stance markers through learner corpus analysis.

Although it is a widely held view that the stance markers play a crucial role in interaction, the functions of the stance markers derived from the studies based on learner corpora analysis have not been conveyed to language teaching. Kärkkäinen (1992) suggests that the lack of epistemic stance expressions in learner speech may derive from the lack of explicit teaching as stance markers are difficult for learners to acquire due to their multiple-meanings. She adds that if epistemic stance devices are presented to learners as extra information, the learners are likely to ignore the features completely. However, it is emphasized that the learners can express themselves "fully and subtly" if the learners are taught how to use stance markers strategically through exposure to a wide range of functions that the markers have. Yet, Granger (2009) comments that features of learner language revealed through learner corpus analysis doesn't necessarily entail an action in the instructional settings. Instead, De Cock (2011) puts forward that the factors like learner needs, objectives of teaching and teach-ability of the stance related features must be taken into account in order to decide which one should or shouldn't be taught. Efstathialdi (2010) recommends that so as to enable students to better understand the semantic differences of stance markers and actual use of target language, various types of oral or written practice based on concordance lines can serve learner needs.

As it is clear from the literature sketched above, teaching stance markers through learner corpora has been limited to the pedagogical implications sections of the studies. To the best of the author's knowledge, the present study stands as one of the first attempts to convey the results of learner corpus research on stance markers to classroom settings. Given the potentials learner corpora has on teaching stance markers, there seems to be a great need to take the research beyond the presented implications derived from corpus analysis and carry the results obtained from learner corpora into the instructional settings. To this end, addressing the gaps identified in the related literature, the current study sets out to analyse epistemic stance markers in spoken language of L2 learners through LINDSEI corpus contrasted with the native English language through LOCNEC corpus. As to the second aim of the study, the effect of explicit teaching of epistemic stance markers to L2 learners is tested depending on the corpus-informed results of the analysis.

2.6 Chapter Summary

This chapter presents four parts. Firstly, an introduction to corpus linguistics and its historical development is presented. Second section provides the basic features of learner corpora and major projects from around the world. The third section deals with stance taking and epistemic stance and the relevant studies carried out in learner corpora with a focus on spoken language. The next section explains the connection between corpus linguistics and its use in language teaching. The chapter ends with the revision of the studies conducted on teaching stance markers, with a particular emphasis on learner corpora.

CHAPTER III

RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the methodology pursued to provide the answers to the research questions of the study which adopts a-two-phase research agenda integrating two interrelated studies being study 1 as the corpus-based analysis and study 2 as the corpus-informed intervention. The chapter is composed of 3 sections. The first section presents information on the design of the overall research by portraying key conceptions referred to in the present study. The second section outlines the corpora investigated in the study and describes the methodology followed in identifying the epistemic stance markers in both corpora including the descriptions of the frequency analysis procedure and the corpus-analytic tool used to extract data for Study 1. The last section provides information on the Study 2, which is a quasi-experimental study, integrating information on the sample, data collection tools, steps followed in intervention stage and data analysis procedures.

3.2 Research Design

It has been widely acknowledged that involving corpus linguists, language teachers and learners in the systematic development and utilization of corpus-informed activities in the classroom is crucial to open up new avenues for pedagogical applications (Campoy, Bellés-Fortuno, & Gea-Valor, 2010; Mukherjee & Rohrbach, 2006). Starting off the very gap between corpus linguistics and English language pedagogy (Mukherjee & Rohrbach, 2006), though gradually bridging, this research project sets two overall goals: 1) to uncover and describe the epistemic stance markers in Turkish EFL learners' spoken language; 2) to measure the change, if any, occurring in epistemic stance marker use in L2 learners' oral performance as a result of corpus-informed treatments. To achieve these goals, the research project elaborated here is composed of two interrelated studies following methodologies from corpus linguistics and the field of language pedagogy, thus, representing an intersection between linguistic insights and their application to English language teaching. For the sake of clarity,

methods followed in each study are presented separately using the labels study 1 for the linguistic analysis part and study 2 for the treatment based part of the research project.

Setting out to explore the use of epistemic stance markers in Turkish EFL learners' spoken English, study 1 made use of a well-known learner corpus within a corpus-based approach which is defined as the analysis of "the systematic patterns of variation and use for those pre-defined linguistic features" (Biber, 2010 p. 196).

Although the number of methods that can be applied to investigate epistemic stance markers remains limited by the use of corpora; in analytic terms, an electronic corpus offers a considerable number of possibilities. Word frequency counts or combinations of words form the core of many corpus studies and as stated by Gries (2010), "branches of linguistics that have been using corpora or text databases have always been among the most quantitatively oriented sub-disciplines of the field" (p. 5).

Therefore, study 1 is quantitative in nature in inquiring epistemic stance markers in learner speech. However, Hunston (2007) claims that using a quantitative approach to investigation of stance may pose some problems since individual words and their functions do not have a simple correspondence. She further explains that counting forms and functions are not the same things and no matter how effective it is to use quantitative approach to analyse stance, it must be accompanied with a qualitative approach. Hence, study 1 also briefly touches on qualitative interpretation of the data to avoid presenting the results in decontextualized numbers and to provide richer insights about the stance use in learner language.

In order to provide strong support to the claims to be made about Turkish EFL learners' interlanguage, a comparable native spoken corpus was also included in the study 1. Thus, the present study follows the principles of CIA as it sets out to make a comparison between L1 and L2.

Granger (2009) defines CIA as a method that "comprises of quantitative and qualitative comparisons between native language and learner language (L1 vs L2) and between different varieties of interlanguage (L2 vs L2) (p. 5). She adds that while the comparisons between L1 and L2 enable researchers to discover distinctive features of learner language, the second type aims to test the generalizability of features of learner language when compared with different learner populations and language situations. In spite of the criticisms the first type of comparison received, it can be claimed that L1 vs

L2 comparisons are highly valuable as they can shed light on the features of learner language that has never been touched upon. Similarly, Granger (2002) notes that "one can engage in close investigation of interlanguage in order to understand the system underlying it and concurrently or subsequently compare the interlanguage with one or more native speaker norms in order to assess the extent of deviation" (p.13). However, the use of learner corpora in linguistics research may have raised some questions about issues such as generalizability and representativeness since learner data is heterogeneous and it involves so much variation in terms of EFL/ESL (Granger, 2002). So as to eliminate this problem, study 1 makes use of two corpora which are designed according to very strict design criteria.

Using the data obtained from corpus-based linguistic analyses, study 2 which forms the second phase of the research aims to test the impact of explicit teaching on the use of epistemic stance markers through concordancing.

For this reason, study 2 adopts a corpus-informed approach to language teaching. A corpus-informed approach refers to an approach "where corpus research findings and corpus data itself can be used to design curricula and materials" (Camiciottoli, 2010, p. 96). The potential of corpus-informed teaching and learning has been widely acknowledged as explained in chapter 2. Whereas, Granger (2004) puts forward that "the number of concrete corpus-informed achievements is not proportional to the publications advocating the use of corpora to inform pedagogical practice" (p. 136).

In measuring the degree of change, if any, occurring in epistemic stance marker use in L2 learners' oral performance as a result of corpus-informed treatments, study 2 also adopts a quasi-experimental design which is defined by Cook (2015) as "testing the casual consequences of long-lasting treatments outside of the laboratory" (p. 1). Cook (2015) adds that the experiments attempt to determine if there was a difference occurred after the treatment, but not to clarify the reason why the difference occurred. The study 2, therefore, exploits the data obtained from quantitative quasi-experimental technique, which involves a pre-test, post-test, and a delayed post-test design. Following the treatments, a questionnaire was employed to check the student beliefs and opinions about the treatment administered. In brief, the study 2 adopts a corpus-informed

approach to language teaching, and combines experimental technique to test the effect of explicit teaching of stance markers on L2 learners' spoken language.

Following section elaborates on the materials, methods and data analysis procedures exploited for Study 1.

3.3 Study 1: Corpus-based Linguistic Analysis

3.3.1. Materials: LINDSEI & LINDSEI-TR

LINDSEI project as a spoken counterpart to ICLE or ICLE's talkative sister (De Cock, Gilquin & Granger, 2010) was started in 1995 at the University of Louvain in Belgium. The Project aimed at collecting a corpus of spoken data produced by advanced learners of English from several mother tongue backgrounds. The first component of the corpus was gathered in French and it contained 50 transcribed interviews with 100,000 words of learner language in total. Recently, compilation of 11 of the components has been completed and they are publicly available, and 9 of them are under development. (Gilquin, 2012) So, 20 sub-corpora in total are included in the project which has expanded it in a large degree.

All the subcomponents of LINDSEI follow the same structure with 50 interviews containing three tasks which are set topic, free discussion and picture description. In addition, transcription and mark-up procedures of the interviews are conducted according to the same guidelines with each interview linked to a profile that depicts information about the learner, the interviewer and the interview (See Appendix A). This information allows researchers to study the possible impact of certain factors on interlanguage (Gilquin, 2012).

3.3.1.1 Learner variables

LINDSEI has been compiled according to very strict and explicit design criteria. Among the eight variables taken into account, mother tongue variable stands as the basic variable for the division of sub-corpora.

The Turkish sub-component, LINDSEI-TR was compiled at Çukurova University drawing the data from the third and fourth year students studying in the Department of English Language Teaching (Kilimci, 2014). Therefore, the participants are regarded as advanced learners based on the external criteria. The component

contains 58 interviews with a total of 80,813 words which are orthographically transcribed but not POS tagged yet (Kilimci, 2014). Regarding the age and gender information for the interviewees, the average age is reported as 22 with 19 male, and 39 female participants (Kilimci, 2014).

Students' English background knowledge is another variable in the component. This variable gives information about their university education and before university education. Table 5 presents the student's educational background in English in LINDSEI-TR.

Table 5. English background knowledge: LINDSEI-TR

	Years	No. of students	Percent
Before University Education	Less than 7	4	6.8%
	7-10	35	60.3%
	10-13	18	31.3%
	More than 13	1	1.7%
University Education	4	42	72.4%
	5	15	25.8%
	More than 5	1	1.7%

As shown in Table 5, before university education, majority of the students received English education at least for seven years. As for the university education, 42 students making up 72% of the participants received four years of English instruction. Therefore, the participants are considered as advanced level learners depending on the years spent on learning English.

Another variable included in the project is the knowledge of other foreign languages. Table 6 shows the distribution of other foreign languages spoken by learners with the number of the students and rounded percentages.

Table 6. Distribution of other foreign languages in LINDSEI-TR

Second Foreign Language	No. of Students	Percent
German	44	75.8%
French	12	20.7%
Other	2	3.5%
Third Foreign Language		
-	47	81.3%
French	3	5.1%
German	3	5.1%
Dutch	3	5.1%
Arabic	1	1.7%
Hungarian	1	1.7%

As Table 6 shows, second foreign language spoken by most of the students is German, and this is proceeded by French. Regarding the third foreign language, majority of the students (81%) do not speak a third language, and the ones who speaks a foreign language constitutes a very small percentage.

In terms of languages spoken at home, the majority of the students (88%) are reported to use Turkish, 9% per cent of the students speak Arabic, and only 3% of them speak Kurdish. Finally, the last variable regarding the time the participants spent in an English-speaking country, 56 out of 58 students have never been to an English-speaking country (Kilimci, 2014).

3.3.1.2 Task Variables

As regards task variables of LINDSEI-TR, it followed the same structure with LINDSEI. The informal interviews lasted 13 minutes on average with the longest interview being 18 minutes and the shortest being 10.41 minutes (Kilimci, 2014). The whole subcomponent for the student turns contains 80,813 words with around 1000 words for each interview.

Each interview in the subcomponent involved three tasks. For the first task which is *set topic*, the three topics among which the interviewers were asked to choose to talk about for 3-5 minutes are as follows:

Topic 1: An experience you've had which has taught you an important lesson. You should describe the experience and say what you have learnt from it.

Topic 2: A country you have visited which has impressed you. Describe your visit and say why you found the country particularly impressive.

Topic 3: A film/play you've seen which you thought was particularly good/bad. Describe the film/play and say why you thought it was good/bad.

The second task being *free discussion* contained an informal conversation about learners' lives related to general topics such as their university life, hobbies, and future plans. In the last part of the interview, the students were required to do a picture description task, so they were provided with a 4-picture sequence and asked to retell the story. See Appendix B for the full version of picture tasks.

Table 7. LINDSEI task variables summarized

TASK VARIABLES	*LINDSEI*	
Medium	Spoken	Typically lower level of self-monitoring than written registers; spoken corpora have not been extensively collected and analysed previously
Field	Education/Academia	Typical learner environment that is familiar to the interviewees
Genre/Text Type	Informal Interview	Similar to spoken conversation due to the informality, thus previous findings on spoken conversation may (cautiously) be used for comparison and explanation, restrictions makes for valid comparisons with subcorpora
Topic/Task	A personal topic set beforehand, informal chat prompted by the interviewer, picture description	Encourages implicit performance (attention to topic rather than language), few constraints on language use, close to natural linguistic behaviour, enough restrictions for sub-corpora to be comparable
Conditions	No reference tools available; non-surreptitious recording; each interview should last for at least 15 minutes	Absence of reference tools creates a more authentic situation, and promotes continuous language use and topic awareness rather than explicit attention to form

During the interview, the students were not allowed to write down any notes or use a dictionary to increase the spontaneity of the language. In addition, although some thematic control was taken on the topics, there wasn't control over their choice of the linguistic forms (Kilimci, 2014).

3.3.1.3 Reference Corpus: LOCNEC

LOCNEC (the Louvain Corpus of Native English Conversation) was compiled as a comparable spoken counterpart to LINDSEI which contains interviews with native speakers of English. This fact enables the comparison of the native language and interlanguage so that the universal and L1-related features of oral language can be identified (Gilquin, 2012). Apart from the fact that the interviews were recorded at Lancaster University, UK, LOCNEC follows the same design criteria as LINDSEI and all the students participated in the project are university students majoring in English. In terms of task variables, it applies the same pattern, which means the students were first asked to talk about pre-identified topics followed by free discussion and picture description task. The interviews in LOCNEC for the student turns make up 118,553 words in total.

3.3.2 Procedures

Before the contrastive analysis of the two corpora, some procedures were followed to be able to conduct a proper data analysis.

To form a basis for the linguistic analysis of two corpora under investigation, first step was to prepare a list of target linguistic forms employed as epistemic stance markers. In deriving such a list, previous research on the topic (Biber, 2006; Kärkkäinen (2003) which reports the linguistic items utilized to signal epistemic modality was consulted. Although the present study mainly made use of the list provided by Kärkkäinen (2003) which presents the most common epistemic stance markers in spoken American English, all the epistemic stance markers listed by Biber (2006) were searched and checked in the native corpus used in this study to make sure whether they display any occurrence in native spoken English or not. The preliminary searches showed that some epistemic markers, particularly, adjectival epistemic stance expressions (e.g. certain, likely, obvious) showed only one or no occurrence at all. That was because Biber's (2006) list involves stance markers in both spoken and written language. Therefore, the markers that displayed 1 or no occurrence in the list were eliminated from the linguistic analyses. Eventually, a list with a total of 32 most common epistemic stance markers in native language were obtained. Table 8 illustrates the epistemic stance markers forming the core of the corpus-based analyses.

Table 8. Epistemic stance markers investigated

Epistemic Stance Markers	
1. I think	17. Sure (adverb)
2. Maybe	18. True (adjective)
3. Of course	19. I know
4. Actually	20. I know+comp.
5. In fact	21. I remember(ed)
6. I don't know	22. Probably
7. I don't know+comp.	23. Must
8. I said	24. Definitely
9. she said	25. Possibly
10. I thought	26. Might
11. I don't think	27. Should
12. I guess	28. Perhaps
13. Will	29. I'm (not) sure
14. May	30. Looks like/to me
15. Would	31. Apparently
16. Could	32. Certainly

As shown in table 8, the list of epistemic devices investigated in this study consists of thirteen epistemic phrases, eleven epistemic adverbials, seven modals, and one epistemic adjective.

3.3.3 Data Analysis

3.3.3.1 Quantitative Analysis

The first step for the data analysis of the study 1 adopts a quantitative approach to investigate the use of epistemic stance markers in spoken language. This step allows for the identification of differences and similarities between native and non-native spoken language regarding the epistemic stance markers.

Upon getting a list of epistemic stance markers, frequency information on epistemic stance markers in both corpora, which are LOCNEC and LINDSEI-TR, was obtained using the AntConc 3.4.3m (Anthony, 2014).

Being a corpus-analytic software, AntConc 3.4.3m is a freely available concordance program which allows for different types of search in a corpus through seven different tools that can be exploited for both quantitative and qualitative analyses of corpora. Concordance tool, for instance shows the search results for the target linguistic item in a KeyWord in Context (KWIC) format. Using the concordance plot tool, it is possible to see where the targeted item in text is located. File view tool enables

a more detailed investigation of the results generated by other tools of the program. Clusters/N-Grams bar allows for the analysis of the whole corpus the clusters in different number of words. Collocates option shows the collocates of the search item so that investigations of non-sequential patterns in language can be identified. Keyword list option can be used for an ESP study because it reveals the unusually frequent or infrequent words in a corpus when compared with the words in a reference corpus. Finally, wordlist tool presents an ordered list of words in the corpus and provides information about frequency. So, this study exploited that concordance program to be able to reach the information related to the frequency of each epistemic stance marker in both corpora.

Being a part of quantitative analysis, this study also employed Log-Likelihood statistics to explore the data on overused and underused occurrences of epistemic stance markers. Paquot and Bestgen (2009) defines it as a test that "compares the observed frequencies with expected frequencies" (p.6). Therefore, if the difference between the observed and expected frequency of a word is large, the log-likelihood ratio is close to zero and higher log-likelihood value refers to a more significant difference between two frequencies. Other than identifying overused/underused linguistic mechanisms, conducting a log-likelihood statistics was helpful to determine which stance markers should be included in the intervention stage. At this point, the critical values (G2) provided were exploited. Figure 6 shows the critical values to be considered in frequency analysis.

The higher the G2 value, the more significant is the difference between two frequency scores. For these tables, a G2 of 3.8 or higher is significant at the level of $p < 0.05$ and a G2 of 6.6 or higher is significant at $p < 0.01$.

95th percentile; 5% level; $p < 0.05$; critical value = 3.84
99th percentile; 1% level; $p < 0.01$; critical value = 6.63
99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83
99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13

Figure 6. Critical Values in Log-likelihood Statistics.

As seen in the figure, the critical values indicate the significance of the difference between the frequency scores of linguistic items in corpus under investigation. As McEnery, Xiao and Tono (2006) assert "By convention, the general

practice is that a hypothesis can be accepted only when the level of significance is less than 0.05 (i.e. $p<0.05$). In other words, one must be more than 95 per cent confident that the observed differences have not arisen by chance" (p.55). Therefore, a cut-off value of 10.83 was selected to determine the stance markers to be taught to the learners, since this can be interpreted as we are more than 99.9 per cent sure that the difference of frequencies observed in two corpora is statistically significant. Therefore, stance markers that indicated a value lower that 10.83 were excluded from Study 2, which is the interventional phase of this research.

Moreover, some functional properties of the epistemic stance markers analysed were included in the present study. The aim of integrating functional properties was a broader and more detailed description to the functions of epistemic markers in interlanguage. In addition, this functional analysis was used to exclude the stance markers with non-epistemic meaning. This was performed for each stance marker that could express some other thing than certainty or likelihood in both corpora. For instance, the modal verb *must* can be used both with its epistemic and non-epistemic meaning as in the examples below extracted from LINDSEI-TR:

Epistemic meaning - *<B> (mm) ... (eh) of course you know . (er) her friends must be pretending that they like the picture </B>*

Non-epistemic meaning - *<B> I I said to myself yes I must study and be a teacher I can't stand you know <overlap /> </B>*

Hasko (2013) notes that "the results of quantitative corpus analysis can be explicated and illustrated through the interpretive power of qualitative methodology beyond the bare statistics of occurrence" (p.4). When it comes to the notion of stance, it constitutes a complex part of human language, which is also highly complex in itself. Therefore, it is crucial that stance markers have to be analysed within the specific context considering the close relationship between its form and function. As Hunston (2007) puts it, quantitative research into stance has to be complemented by a qualitative analysis, and since quantitative analysis should be only a starting point to identify the functions of stance markers, this study also elaborated on the functional features of the pre-identified items.

As for the functional classification of epistemic stance markers, evidence from different previous research has been consulted (Chafe, 1986; Biber, 2006; Kärkkäinen,

2003; Brezina, 2012; Gablasova et al, 2015; Aijmer, 2004; Baumgarten & House, 2010; De Cock, 2004). The epistemic markers in learner language contrasted with the native language in spoken context were analysed in relation to their meaning and position in the interactional context.

3.4 Study 2: Corpus-Informed Intervention

Study 2 aims to explore the effect of (if any) corpus-informed intervention on the use of stance markers in Turkish EFL learners' spoken English.

3.4.1 Participants

As the idiosyncratic features of research setting and participants have a determining role on the overall research design, description of the research site would provide a clear understanding of the underlying variables of the present study. The present study has been conducted at the Department of English Language and Literature, Fırat University in the fall term of 2016-2017 academic year. The research site accepts students based on the results of a nation-wide university entrance exam a part of which is measuring English proficiency of the students. Those scoring well in this exam are subject to another English proficiency exam prepared and administered by the university itself in accordance with the Common European Framework of Reference for Languages (CEFR). The students are required to get 60 or above out of 100 from the proficiency test. Those failing the exam are exposed to 30 hours of English instruction in one week through one academic year, in the first term studying integrated skills and focusing on the separated skills in the second term and are expected to finish the year with the high intermediate level, which is B2 according to CEFR qualifications. The students who get 60 and over are considered to be at B2 CEFR level and gain the right to enrol in the first year of the undergraduate program in the Department of English Language and Literature. English is the medium of instruction in the research site and all of the courses offered, except one, are literature-focused, which means that the participants of this study had no opportunity for further language learning let alone learning specific uses (e.g. stance markers) in spoken English.

Drawn in accordance with the convenience sampling procedures which "involves choosing the nearest individuals to serve as respondents" (Cohen, Manion, & Morrison, 2000 p.102), participants of the present study are 39 Turkish EFL learners at

their first year. 31 of the students are female and 8 of them are male and their ages range from 18 to 22. Except for one student who is Arabic, 38 out of 39 participants are Turkish and all the participants are native speakers of Turkish. Furthermore, all the participants included in the study reported that they have never been to an English-speaking country before. Regarding the years they spent learning English before university and while studying at university, on average, the participants were exposed to English for 8,5 years at school and 1,7 years at university. As for the languages spoken at their home, 33 students reported that they only speak Turkish, 3 of them speak both Turkish and Kurdish, and 3 of them speak Turkish and Arabic.

Regarding the English proficiency level of the participants, contrary to the supposed B2 level according to the proficiency test of the university, the results of the Oxford Quick Placement Test (2004) showed that 8 students are at elementary level (A2), 30 students are at intermediate level (B1), and only two students are at high-intermediate level (B2). The proficiency levels of the participants were validated with a self-assessment grid on speaking skill, which showed that while some students underestimate and some others overestimate their proficiency level of English, approximately half of the participants were aware of their abilities in speaking. Table 9 summarizes the characteristics of the participants.

Table 9. Characteristics of participants summarized

	No. of Students	Percent
Gender		
Female	31	79.5%
Male	8	20.5%
Age		
18-22	39	100%
Nationality		
Turkish	38	97.5%
Arabic	1	2.5%
Native Language		
Turkish	39	100%
Proficiency level		
A2	8	20.5%
B1	30	76.9%
B2	1	2.6%
Total	**39**	**100%**

3.4.2 Data Collection:

The data was collected through a total of 7 types of research instruments; a proficiency test including a Self-Assessment Grid on speaking proficiency level, feature-specific speaking tests (as pre-test, post-test and a delayed post-test), instructional materials and questionnaires.

Before the intervention stage, the students were given Oxford Quick Placement Test (2004), which is a paper and pen test of English language proficiency developed by Oxford University Press and Cambridge ESOL. The two parallel versions available for the test were administered. The 60 items in multiple choice formats assess reading, grammar and vocabulary. The test consists of two parts: the first part (questions 1-40) is taken by all the students and intended for those who are at intermediate level or below. The second part is only for students who scored more than 36 out of 40 in the first part.

Along with the proficiency test, the participants were given a self-assessment grid designed by Common European Framework Reference (CEFR). So as to determine their self-observed proficiency in speaking, only the parts including *speaking interaction* and *speaking production* were given to the students. Thus, the general English proficiency level of the students was validated with the students' self-assessed proficiency levels on speaking.

Then, the participants took a pre-test before the treatment to identify which epistemic stance markers they employ without any intervention in their oral production which enabled the researcher to see if there was a significant difference after the teaching sessions. The pre-test was a spoken test which was adopted from the speaking sections of different sample tests of IELTS which is accepted as an internationally and officially valid test. Two students participated in each session to allow for some interaction between the students. In addition, so as to prevent the students from feeling stressful during the tests, the participants were allowed to choose their own pairs.

Pre-test included 3 main parts and students were asked to speak individually for the first two parts and have a discussion for the last part. Therefore, the questions were repeated for the second student in the room. (See Appendix C). The first part was an introductory part including questions about the participant's name, hometown, and beliefs about learning English. In the second part, the participants were shown a picture that required them to guess what is likely to happen / to be happening in the picture. For

example, the students were shown a picture with a crowd of people with surprised faces looking and pointing at somewhere, and they were asked to make some predictions about the possible event. For the last part, the students were required to talk to each other on a pre-identified situation providing them with a set of pictures to help them generate some ideas on the given topic. One example from the test is illustrated below:

> *A young man on holiday in **North America** wants to buy **a present** to take home to **his parents**. Talk together about the different presents he could **buy**, and say which would be **best**.*
>
> *Here is a picture with some ideas to help you.*

On each part, students were required to talk about for 2-3 minutes. Pre-test was administered three days before the treatment. Three sets of exam papers in the same format but with different topics and pictures were prepared. All the three parts in the three sets contained questions with comparable difficulty. The purpose of preparing three sets of exam papers was twofold: first, the participants didn't know which of the three sets of exam papers they would encounter in one test. Second, they were assigned a different set in each three exams, namely the pre-test, post-test, and delayed post-test. Therefore, the tasks were rotated in each test in order to eliminate the familiarity effect and to prevent the questions from limiting the participants' answers or comments in each test.

The instructional sessions lasted 5 weeks including 12 sessions in total. In the sessions, a power-point presentation prepared by the researcher was used. The slides contained explanations on the functional properties of epistemic stance markers, on their positions in the sentence through extra genuine examples. In addition, the students were provided with a handout with tasks in it which followed the steps of 4 I's approach. Three days after all the sessions were completed, the students took the post-test to see the change, if there was any, in their use of epistemic stance markers. In the post-test, students were asked to attend the test with the same participant that they attended the pre-test. Each group were assigned a different set of exam paper than the one they were assigned in the pre-test. Three weeks later, a delayed post-test was conducted to see the retention effect of the instructions given. For each test, the participants' speech was audio recorded upon taking their consent.

As for the last research instrument, the participants were given a post-instruction perception questionnaire (See Appendix F & G for English and Turkish versions) which was adopted from to the relevant literature (Yoon & Hirvela, 2004) considering distinctive features of the present study. The survey aimed at obtaining the personal information of the participants together with the information on their general computer use. Also, it aimed at finding out the participants' attitudes towards three domains, which are on 1) general use of the corpus, 2) difficulties of using the corpus and 3) the use of corpus to learn epistemic markers. For each domain, the participants were asked to indicate their degree of agreement on a 7 point Likert scale (where 1 stands for strongly disagree, 2 for disagree, 3 partly disagree, 4 for no opinion, 5 for agree, 6 partly agree and 7 stands for strongly agree). Considering the proficiency levels of the students, the questionnaire was translated into Turkish for the students to provide answers more clearly and the items were rechecked by another instructor for any mistakes in the Turkish version. Also, the questionnaire was piloted with a similar group of students, who were familiar with corpus applications, to see if some improvement was needed for the actual implementation The next section explains the procedures of the instructional phase.

3.4.3 Procedures

The instructional phase attempted to conduct an explicit teaching of epistemic stance markers through a concordance program. In total, 19 epistemic stance markers were assigned for teaching sessions after examining the findings obtained through Log-likelihood statistics, which allowed for eliminating the epistemic markers of which ratio was lower than the pre-determined critical value of 10.83 as explained in section 3.3. In every session, 1-2 epistemic markers were studied via the concordance program. Before the sessions began, a freely available and easy-to-use concordance program was to be determined. Therefore, the spoken section of the British National Corpus (BNC-BYU) was chosen. The 12 sessions took place in a computer laboratory which provided the participants with internet access.

Before the teaching sessions, two sessions lasting 90 minutes in total were allocated for the introduction and a brief explanation of the purpose of the study. In this part, the participants were also provided with the crucial information about why the use of a wide range of epistemic markers are vital in speech Therefore, this introductory

part attempted to raise the learners' awareness of the importance of epistemic stance markers in spoken production ad it helped to keep the participants engaged and motivated during the sessions. In addition, since a corpus was to be used as a learning tool in the teaching sessions, the BNC corpus was introduced to the learners and a step by step use of the corpus was explained in detail. Each student was provided with special assistance when they had any kinds of difficulty in using the program. Regarding the purposes of the study, the *conversation* genre under the spoken section of BNC was used so that the participants would be exposed to genuine examples of informal native speech. After making sure that all the participants could use the concordance program efficiently through some trial searches done with the researcher, a handout (See Appendix D) was given to each learner to use throughout the sessions that they could follow the steps of the approach adopted clearly. The students were exposed to the explicit teaching of two epistemic markers in each session.

As for the teaching methodology, the 4 I's method proposed by Flowerdew (2009) was followed. In the illustration part, the students were asked to search for and look at the data for the epistemic marker in question and write down five sample sentences generated by the concordance program. The learners were encouraged to write down the example sentences which they could understand more easily in order not to make them feel demotivated due to the loads of unknown vocabulary. In the interaction part, by examining the examples they wrote down, the students were required to discuss with their friends the linguistic and functional features of the target epistemic marker within the context and share their ideas. For the intervention part, the students were directed some prompt questions and clues to help them understand the form and function of target item such as "How do you think the epistemic marker contributes to the meaning of the sentence? / What message do you think the speaker tries to give by including this epistemic marker in his /her interactional turn? What do you notice about the position of the stance marker in the sentence?". Finally, as for the induction part, students were asked if they can find out a general functional pattern in which the target epistemic marker occurs. Following that, each epistemic stance marker was explicitly taught to the learners through instructional materials (For a sample, see Appendix E) explaining its different functions and position in the sentence. The students were also provided with some example sentences taken from corpus to back up their learning with extra information. As the last step, the participants were shown a picture

through power point presentation and they were asked to make up a short dialogue with their pair making use of the target epistemic markers studied in the session. The pictures chosen required the learners to make predictions and enable them with a context in which they could use the epistemic markers taught. For instance, in one of the sessions, the students were shown two groups of inventions by men and women and they were asked to predict which inventions were by men and which ones are by women. Following that, they were asked to form a short dialogue containing the target epistemic markers. It was made sure that all the students wrote down their dialogues on their handout. However, as it was not possible for all students to role-play their dialogue due to time limitations, the volunteering participants role-played their dialogue in front of their classmates. The same procedures were followed in teaching of each stance marker.

3.4.4 Data Analysis

The study 2 aimed at measuring the effectiveness of explicit teaching of stance markers on the same group of participants through a pre-test, post-test, and a delayed post-test. Therefore, in order to identify the differences between the results of the three tests, statistical analyses were conducted through a paired samples t-test. This enabled the researcher to see if there was a significant difference in the learner's performance in relation to the use of epistemic stance markers in their oral production after the teaching sessions.

Firstly, the audio-recording of the interviews conducted for the pre-test, post-test and delayed post-test were analysed in two different ways: 1) by examining the overall frequency or each marker that the participants employed in their speech and 2) by investigating the range of each epistemic marker used.

For the first measure of instructional impact, the overall frequency of each epistemic stance marker was counted. However, at this point, there was a need to conduct manual analysis to make sure whether the marker was used in its epistemic meaning or not. This was especially necessary for the modal verbs; for example, the modal verb 'would' could be used as a marker of epistemic certainty or a modal verb explaining volition.

Regarding the second measure, while epistemic stance markers were analysed according to only frequency in the previous literature, in this study, each student was

given an *epistemic range score*, which was labelled as *epistemic type score* by Fordyce (2014), by identifying the different epistemic markers deployed by a participant. For each marker, the participant got 1 point. For example, if the participant used three different epistemic markers such as *I think, I don't know,* and *perhaps*, s/he got an epistemic range score of '3'. (Fordyce, 2014) The scores given to be able to represent the range of epistemic stance markers used in their oral production as one of the main purposes of the instructional phase was to enable the students to make use of a wider variety of epistemic markers.

The data of two measures for each three tests, which are overall frequency counts and epistemic range score, for each participant for three time points were entered into Statistical Package for Social Sciences (SPSS version 22).

Firstly, a paired samples t-test was conducted to see the overall effect of the instruction and student gains in epistemic stance marker knowledge and use in spoken production. Then to understand whether the participants retained their knowledge after the treatment or not, a paired samples t-test was run comparing the post-test and delayed post-test results of the participants. As for the residual effects, the pre-test and delayed post-test scores were compared again by running a paired samples t-test.

Secondly, an analysis of covariance (ANCOVA) was run to see whether the results of the post-test, which is the independent variable of the study, depend on any other variable like proficiency level or not. ANCOVA, as noted by Larson-Hall (2010), is an effective analysis used when the focus is "on the effects of a main variable with the effects of other interval-level variables factored out" (p.357). It is especially useful when the participants are "not starting on equal footing", which forms an external variable and ANCOVA allows the effects of that variable (such as age, proficiency level) to be separated out from the response variable" (p.357). As participants' scores from proficiency test in this study indicated that the students were not homogeneous in their level of proficiency which could have impact on their overall command on speaking in English, controlling proficiency level as a variable would lead to have sound claims about the instructional effect.

Thirdly, differences of student scores in three different task types were calculated through descriptive statistics to see in which task students preferred to

employ epistemic stance markers more frequently and possible reasons were suggested depending on the nature and content of each task.

Lastly, the questionnaire results of participants' attitude towards the use of concordance for learning epistemic markers in speech was analysed. The questionnaire applied was composed of three domains which were related to the strengths, drawbacks, and to general nature and use of the corpus consulted in the study. The results were analysed using descriptive statistics.

3.5 Chapter Summary

This chapter presents the methodology utilized in the present study. The first section explains the nature of the study explaining the key notions referred to in this research. The second section introduces the corpora used for investigating the frequency distributions and linguistic features of epistemic stance markers. The last section outlines the quasi-experimental study, which aimed at the explicit teaching of epistemic stance markers through concordancing, including the descriptions of the participants, data collection tools, steps followed in the actual treatment and data analysis procedures.

CHAPTER IV

FINDINGS AND DISCUSSION

4.1 Introduction

This chapter explains the findings of the present study. The chapter includes 2 sections. The analysis reported in the first section, which is Study 1, is three-fold: 1) frequency analysis of epistemic stance markers in LINDSEI-TR, 2) frequency analysis of epistemic stance markers in LOCNEC, 3) contrastive analysis of epistemic stance markers in two spoken corpora under investigation in terms of overuse/underuse. The second section presents the finding of Study 2, which attempted to convey the results gained from the corpus analysis into pedagogical settings making use of explicit teaching methodologies and concordance lines. The second section presents the results of the corpus-informed intervention in five parts: 1) overall effects of the explicit treatment in teaching epistemic stance markers through corpus-based learning tools, 2) effects of the explicit treatment on the frequency and range of epistemic markers employed by Turkish L2 learners, 3) individual analysis of epistemic stance markers before and after the intervention, 4) use of epistemic markers in different task types by L2 learners, 5) learner reactions towards the use of concordance lines in learning epistemic markers in spoken language.

4.2 Study 1: Corpus-based Linguistic Analysis

The study 1 attempts to conduct a corpus–based analysis of epistemic stance markers in spoken interlanguage contrasted with the native spoken language following the framework of CIA. Through the quantitative analysis, the aim is to identify the most frequent epistemic stance markers employed by Turkish L2 learners and native speakers of English. In addition to the raw and normalized frequency counts in both corpora, Study 1 illustrates information on the overused and underused epistemic markers when contrasted with the native spoken language

4.2.1 Epistemic stance markers in LINDSEI and LOCNEC: Quantitative Analysis

This part of the Study 1 sets out to provide answers to the first two research questions of the study, namely, the most frequently used epistemic markers in native spoken English and in L2 production of Turkish learners of English are clarified.

Using AntConc 3.4.3m, the raw frequency of epistemic markers in LINDSEI-TR and LOCNEC were calculated. Gries (2010) comments that 'a higher frequency occurrence of some element in some corpus (part) does not automatically [show] that the element observed more often is more frequent because the observed frequencies are of course dependent on the sizes of the corpus parts that are compared.' In parallel with this, normalized frequency counts of the epistemic markers were calculated. In order to identify the similarities and differences between two corpora in terms of overuse and underuse of epistemic markers, Log-likelihood results were also calculated. Table 10 presents the overall frequency of epistemic markers in question in both corpora according to their grammatical class.

Table 10. Grammatical classes of epistemic stance markers in LINDSEI-TR vs LOCNEC

Grammatical Class	**No. of occurrence**	
	LINDSEI-TR	LOCNEC
Epistemic Phrases	327	967
Epistemic Adverbs	273	594
Modals	54	267
Epistemic Adjectives	7	69
Total	**661**	**1897**

According to Table 10, it can be seen that both corpora has the same order of frequency of epistemic markers in use in terms of their grammatical class. Epistemic phrases such as *I think, I know, I guess* are the most common group in the data. They are also labelled as *parenthetical clause* or *comment clause* in the relevant literature but with somehow wider meanings covering epistemic modality (Kärkkäinen, 2003). Epistemic phrases are followed by epistemic adverbs (maybe, actually, etc.) and modals. Epistemic adjectives are the last type of occurrence with a very low frequency in both corpora. Epistemic nouns like *it is my (guess/opinion) + that clause* are

excluded from the analysis as they have only one or two occurrences in both corpora. Figure 7 visualizes the percentages for the grammatical distribution of epistemic stance markers in both corpora.

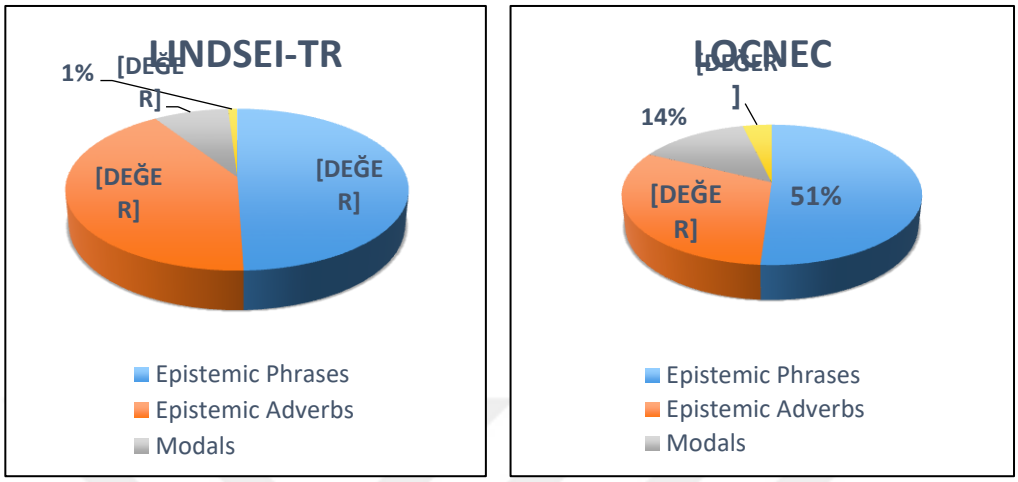


Figure 7. Grammatical Distribution of Epistemic Stance Markers in LINDSEI-TR vs LOCNEC

Although this finding is in contradiction to the corpus-based study by Biber et al. (1999), who investigated stance in spoken English and found out that the modals are the most common epistemic markers, the order of frequency related to the grammatical classes of epistemic markers found in the present data is in line with previous studies reporting the epistemic phrases as the most common group (Kärkkäinen, 2003; Letica, 2009; Recsky, 2006). Three possible reasons to the infrequency of modals in conversational English are suggested by Kärkkäinen (2003):

1) Being restricted to clause-medial position, inability to move
2) Being implicit rather than explicit markers of subjectivity, so more explicitly personalized markers are preferred by speakers,
3) Being ambiguous regarding sematic meaning, so other markers that unambiguously express epistemic stance are used

In terms of L2 production, the three reasons offered seem to apply to spoken language of non-native users to a great extent. The findings of the present study show that not only native, but non-native speakers tend to prefer epistemic phrases and adverbs more in their speech. This can be due to the fact that these two categories

involve epistemic devices that are relatively more flexible in terms of their position in sentence, and are unambiguous and more explicit. Depending on the idea that acquiring the stance markers, and performing them in speech is difficult for even advanced level learners, (Gibbs, 1990; Fordyce, 2014) L2 learners' preference towards epistemic phrases and adverbs rather than modals may be linked to their desire to sound more fluent when speaking in the target language.

Although the findings of the study seem to be in parallel with a number of previous studies in terms of grammatical classifications of epistemic markers, it can be seen that the study presents interesting findings when the epistemic markers are analysed separately in non-native and native speech. As the first step of analysis, epistemic stance markers in learner corpus were investigated. Table 11 provides the most common epistemic stance markers in the spoken data of Turkish L2 learners and the number of occurrences for each stance marker as well.

Table 11. The most common epistemic markers in LINDSEI-TR

Marker	**LINDSEI-TR (No. of occurrences)**		
I think	202	Sure (adverb)	7
Maybe	102	True(adj)	6
Of course	60	I know + comp	6
Actually	58	I know	5
In fact	38	I remember(ed)	4
I don't know	27	Probably	3
Would	25	Must	2
I said	24	Definitely	2
S/he said	16	possibly	2
I thought	13	Might	1
I don't think	11	Should	1
I guess	11	Perhaps	1
Will	10	I'm (not) sure	1
May	8	Looks like/to me	0
I don't know + comp	8	Apparently	0
Could	8	Certainly	0
			Total: 661

As shown in Table 11, it is seen that so-called advanced level Turkish learners of English exploit the epistemic stance markers in their speech. Yet, when the range of markers found in the data are analysed with the frequency counts, it becomes evident

that the pattern they follow to express their commitment to the truth-value of the proposition is very restricted and routinized.

A closer look at Table 11 shows that the most frequent epistemic marker applied in the speech of L2 learners is *I think* with a total of 202 occurrences. This finding gains support from a great number of studies in which the speakers showed a clear preference to employ this marker in their speech (Kärkkäinen, 2003; Letica, 2009; Aijmer, 2004; Gablasova et al., 2015). It is pointed out in the literature that *I think* is a signal of uncertainty especially when it is placed in the medial or end position of the sentence (Aijmer, 2001). Considering that there are a great number of examples in LINDSEI-TR in which *I think* is used in medial and final position, it can be claimed that learners tend to express more uncertainty in spoken interaction (Aijmer, 2004). The examples below represent the uses of *I think* in medial and final position.

(1) *<B> they should learn this (mm) . they should acquire the language not second language I think </B>*

(2) *<B> and (eh) you have to go to bank and anything else and (eh) the life in dormitory I think (eh) you have to get along with lots of people (em) there are lots of people and they are really different from you (eh) in terms of personality or different things and (em) you learn to get along with those people and (eh) during this time you really (eh) have difficulties I think </B>*

One possible reason for high frequency of *I think* in learner language is proposed by Aas (2011) who claims that the learners may not have the knowledge of how to express their epistemic stance or weakening their responsibility in other ways depending on their proficiency level. This may build support for the high occurrence of *I think* in Turkish L2 learner speech.

Following *I think* as an epistemic phrase, *maybe* is the most frequent stance adverbial in the data. When the frequency table for LINDSEI-TR analysed, it can be seen that after *I think*, epistemic stance adverbials *maybe, of course, actually,* and *in fact* constitutes the most frequently used epistemic markers employed by Turkish L2 learners. *Maybe* stands as the most frequent epistemic adverbial with 102 occurrences. *Of course* with 60 occurences, *actually* with 58 occurrences and *in fact* with 38 occurrences are the followings of *maybe*. Whereas, stance adverbials, *sure, probably,*

definitely, possibly, apparently perhaps, and *certainly* display just a few occurrences in the data. This finding doesn't correspond to the findings of the study conducted by Letica (2009) who found out that *probably* is the most frequent stance adverbial used by Croatian L2 learners of English.

Although the study indicated that *probably, maybe,* and *of course* are among the most frequently used epistemic markers, it seems that Turkish L2 speakers tend to deploy a somehow different set of epistemic adverbials, for instance in the Turkish L2 learner data, *probably* is used only 3 times. Additionally, while Letica (2009) reports that in her data, *I think* and *probably* accounts for almost 60% of the epistemic markers employed by Croatian L2 learners, in LINDSEI-TR, four epistemic devices *I think, maybe, of course* and *actually* accounts for %60 of the data. On the other hand, in the study carried out by Nozawa (2015) to investigate epistemic hedging devices used by Japanese EFL learners, it is shown that *I think* and *maybe* are the most common epistemic markers. So, this supports the findings of the present study.

Concerning *of course,* it has been found out to be a high frequency stance marker in learner speech in the relevant literature and the learners tend to misuse that marker especially in companion with yes/yeah as a response to an opinion expressed by the interlocutor, which may make learners sound rather over-emphatic and impolite (De Cock, 2004). In the non-native data, there are instances of *yes/yeah of course* identified as misuse, which confirms the claim suggested in the previous research.

(3) *<B> and (eh) I learned (eh) very important things in my life (eh) my own money I was (eh) paying and then also I was a child I can I can say I learned the life at that time and at first time I was (eh) sad and (eh) sorry for going that school but by the time I learned it is really important for me and now I think I am very strong </B>*

<B> yes of course </B>

(4) *<B> <overlap /> because we're Turkish and we're different they they were interested <overlap /> in in us </B>*

<B> <overlap /> yeah of course </B>

The next most frequent marker in LINSDEI-TR is *I don't know* with 27 occurrences, and the data displays 8 uses of *I don't know + complement clause*. Aijmer

(2004) notes that this marker indicates that “speakers are not taking full responsibility for what they are saying”, and it can be used together with other uncertainty markers like *maybe, I think.*

(5) *<B> then the painter agree (eh) he draws her like a beautiful woman a very beautiful (em) but not realistic (eh) the picture is far more far more different from the woman her= herself then (eh) that woman gets the picture to his to her house and shows it to her friends she wants them to be jealous of her and they seem so: I don't know maybe they were laughing (em) or . I can see their faces but (em) it's uglier that the woman in the picture is very different from the woman than real one so (eh) the woman the all women are . looking at he= at the picture and the woman and they're surprised because there is no nothing (em) related to each . (mm) so we can conclude that women like showing off <laughs> </B>*

In addition, it is claimed in the relevant literature that the “difficulty of expressing themselves in a foreign language interferes with the encoding process as a whole” (De Cock, 2004, p. 29) as in the following example suggested by Aas (2011) from the Swedish subcomponent of LINDSEI:

(30) because he’s pumping money into the business <breathes> so: and she’s she’s ill as well she has a: I don’t know what you call it (eh) . in the lungs (LINDSEI-SW) (p. 97)

Although this usage is reported to be the most common usage of *I don’t know* in the German component (Götz & Schilk, 2011), it is reported by Aas (2011) that the Swedish component has only four instances with retrieval problems and the learners make use of the word combination *what do you say*. This is in agreement with the findings of the present study as the examples of the use of *I don’t know* as a retrieval strategy is not found in LINDSEI-TR. An interesting finding is that, instead, Turkish learners tends to prefer the word combination *how can I say* when they have difficulty in retrieving a word from the memory or in expressing themselves in the target language. This can result from either mother tongue effect or the instructional differences between cultures. In example 6 below, the speaker makes of the word combination *how I can say* to express that s/he is experiencing a retrieval difficulty and combines *I don’t know* to emphasize that s/he lack sufficient knowledge.

(6) *<B> (erm) .. how can I say in English (em) P D R </B>*

<B> <foreign> yan dal yapyyorum hocam ben </foreign> </B>

<B> how can I say in English I don't know </B>

In terms of modals, *would* is the most frequent modal verb (n=25) to express epistemic stance. This finding is in agreement with Kärkkäinen's (2003) study who found that *would* is the most frequent modal verb functioning as an epistemic device with a different order of frequency of other modal verbs: *would, must, might, could, will, may,* and *should*. The following modal verbs in the data in order of frequency are *will* (n=10), *may* (n=8), *could* (n=8), *must* (n=2), *might* (n=1), *should* (n=1). In the present data, the learners in fact make use of modals to a great extent. However, within the scope of the study, when the epistemic uses of the modals were extracted from the data, the numbers considerably fell down. Furthermore, several reasons were suggested above in relation to the low frequency of modal verbs as epistemic markers. As for LINDSEI-TR, it appears that instead of employing modal verbs to express un(certainty), the learners tend to make use of markers which are not ambiguous and have flexible use in sentence such as adverbs and epistemic phrases. Another reason for the learners' low usage of modals as epistemic markers may stem from being not aware of the modals' epistemic meanings. For instance, *must* has only two occurrences in the data expressing epistemic stance out of 20 instances, but in almost all cases it is used to express deontic modality. Two uses of *must* in LINDSEI-TR to express epistemic commitment can be seen from the following examples:

(7) *<B> (mm) ... (eh) of course you know . (er) her friends must be pretending that they like the picture </B>*

(8) *<B> and the other one maybe is thinking I should have some artist to paint my (eh) portrait too or I just should do some kind of thing and the other must be thinking of oh I should pose like that like that I have learned something . I don't think that they appreciate much they don't look very (eh) honest to me </B>*

Although *I think* is found to be the most prevalent epistemic marker in LINDSEI-TR, other epistemic phrases such as *I said* (n=24), *s/he said* (n=16), *I thought* (n=13), *I don't think* (n=11), *I guess* (n=11), *I know + complement clause* (n=6), *I know*

(n=5), I remember(ed) (n=4), and looks like/to me (n=0) are found in learner speech respectively. Kärkkäinen (2003) claims that *I said* and *s/he said* are categorized as hearsay evidentials which are most common with the utterance verb *say* that occur in introductions of direct reported speech and adds that direct reported speech is regarded as a common interactional device to convey stance towards past actions or reported speaker. There are 40 occurrences of this type of markers in total in LINDSEI-TR. This is in contrast with the findings of Kärkkäinen (2003) and Biber et al. (1999) in which they found the most frequent verb after *think* is *say.* In the Turkish case, *know* appears to be the second most frequent verb after *think*. Kärkkäinen (2003) states that when *I said*/ *s/he said* are used in conversation, the speakers embed themselves or a different speaker into the discourse and they function as a 'change in footing'. As for *I don't think*, Bybee (2007) suggests that in all tokens of *I don't think*, speakers express their epistemic stance to hedge their opinion and manage polite disagreement in conversation. This seems to be the case for LINDSEI-TR as in the following example:

(9) *<B> actually before coming here (er) I would expecting the[i:] university to be much more fun <overlap /> </B>*

<B> but I didn't have it here . (eh) I'm happy that I'm going to be graduating this year </B>

<B> <u>I don't think</u> that there will be . lots of things that I will <X> <overlap /> that I will miss <X> </B>

For the case of *looks like/ to me or seems like/to me,* it can be seen in Table 11 that they do not occur in the data, which is in contrast to the findings of Kärkkäinen (2003) and Letica (2009). Especially, Letica (2009) reports the verbs *look* and *seem* are among the most frequent verbs used by Croatian L2 learners.

Lastly, for the group of epistemic adjectives, only two examples were found in the data: *true* (n=6) and *I'm sure* (n=1), which represents very low frequency. This finding is also in parallel with Kärkkäinen's (2003) study which suggests that adjectival forms to express epistemic stance are often linked to written language, rather than conversational language.

All in all, it can be seen that learners tend to use a rather limited set of epistemic markers in their speech, with the prevalence of *I think*. Although different types of

epistemic markers are found in the spoken production of Turkish L2 learners, they tend to deploy epistemic phrases and epistemic adverbs in their speech more frequently.

4.2.2 Epistemic stance markers in LOCNEC: Quantitative Analysis

Since an analysis of the epistemic markers in native speech separately could provide a basis to make a comparative analysis of learner and native corpora, this section focuses on the most frequent epistemic markers in native speaker English. Table 12 displays epistemic markers with their raw frequency in the native corpus.

Table 12. The most frequent epistemic markers in LOCNEC

Marker	**LOCNEC (No. of occurrences)**		
I think	432	I know + comp	30
Actually	221	True(adj)	29
I don't know	166	Certainly	27
Would	133	S/he said	25
I thought	120	I said	24
Probably	93	possibly	22
Maybe	86	Apparently	21
Might	63	Must	19
Definitely	49	Looks like/to me	18
I don't think	46	I guess	11
I don't know + comp	43	May	9
I know	43	I remember(ed)	9
I'm (not) sure	40	In fact	6
Could	36	Will	5
Perhaps	35	Sure (adverb)	2
			Total: 1897

As the table illustrates, it can be seen that native speakers make use of epistemic stance markers in their spoken production. However, when looked at the frequencies given, it can be suggested that they tend to rely on a restricted and routinized set of patterns in speech, which is also claimed by Kärkkäinen (2003). It was explained in the previous section that in native spoken English like the non-native spoken English, epistemic phrases and epistemic adverbs are the most frequently used epistemic markers to convey (un)certainty towards a propositional content. It was observed that epistemic phrase *I think* is dominating the native spoken language with 432 occurrences. This finding is in line with the studies conducted by Kärkkäinen (2003) and Baumgarten and House (2010). Baumgarten and House (2010) state that instead of talking about events and actions, people prefer to talk about their feelings and attitudes to express their

degree of belief, opinion or subjective evaluation towards the propositional content. Therefore, *I think* is frequently used in spoken language as an epistemic marker. What's more, Aijmer (1997) differentiates between the tentative and deliberative use of *I think*. In tentative use, the marker is employed to express the speakers' uncertainty functioning as a hedge, and in deliberative use, *I think* is followed by the complementizer *that* with a boosting function. In LOCNEC, the use of *I think* with complementizer *that* occurs only 31 times out of 432 occurrences in total. Thus, it can be claimed that the native speakers tend to use this marker more in its tentative function where they express their uncertainty towards the proposition.

Table 12 shows that the second most frequent epistemic marker in the data is *actually*. This marker is not available in Kärkkäinen (2003). However, Biber and Finegan's (1988) study investigating stance adverbials in different text types both in written and spoken English shows that *actually* frequently occurs in informal speech in particular. Aijmer (2002) points out that the most typical use of *actually* is in final and mid position and a closer look at LOCNEC shows that the use of *actually* in final and mid position is very frequent while the initial position is very rare. This finding could confirm that rather than being a discourse particle, *actually* is used as an epistemic marker. Furthermore, Quirk et al. (1985) state that *actually* can occur with various functions and positions. They argue that it can both have an emphatic function and disjunctive function which can 'state the sense in which the speaker judges what he says to be true or false' (p. 621). When it is used in pre-verbal position, *actually* functions as an emphasizer strengthening the truth value of the clause in which it is employed. (Aijmer, 2002) When looked at LOCNEC, a great number of examples can be found in which *actually* is used in pre-verbal position. See the following example:

(10) *<B> the problem is I haven't got any experience at the moment of er . . of doing <X> I <u>actually</u> did an interview with erm one of the sports reporters on my local paper <\B>*

(11) *<B> yeah .. and the climate as would be different <laughs> <\B>*

<B> <laughs> <\B>

<B> yeah <\B>

<B> [erm I've thought about it a lot <u>actually</u> <\B>

As a similar adverb to *actually, in fact* is also used as an epistemic marker in speech. Although they are very similar in their core meanings, it is seen that in fact occurs just 6 times in LOCNEC. The reason why can be attributed to the fact that it is reported to be more common in formal speech (Aijmer, 2002). Since the interviews in LOCNEC involve informal speech, it is very infrequent in the data. In terms of other stance adverbials found in the data, *probably* is the second most frequent epistemic stance adverbial with 93 occurrences, and *maybe* (n=86), *definitely* (n=49), *perhaps* (n=35), *of course* (n=32), *certainly* (n=27), *possibly* (n=22), *apparently* (n=21), *in fact* (n=6), and *sure* (n=2) are the following epistemic stance adverbials with their occurrences indicated in parentheses. This seems to be in agreement with Biber et al. (1999) who presented the most common epistemic adverbials with the following order: *probably, maybe, of course, certainly, definitely and perhaps* as expressing epistemic doubt and certainty. Under the label of epistemic actuality, they found that *actually* is a very common epistemic marker with 800 occurrences per million words, which confirms the findings of the current study. Similarly, Kärkkäinen (2003) finds out that *maybe, probably, apparently, of course,* and *definitely* are the most common adverbials conveying epistemic stance of the speaker. Kärkkäinen (2003) explains that spoken language involves on-line planning, and the syntactic mobility and the freedom of occurrence of the adverbs can ease the process of that planning. In turn, this can explain their considerable use in spoken language.

I don't know is the next most common epistemic phrase in the data with 166 occurrences. It is also commonly used with a complement clause (n=43). This finding is also in line with Kärkkäinen (2003). As Baumgarten and House (2010) put it:

With regard to the expression of speaker stance, I don't know in L1 British and American English thus signals the speaker's insufficient knowledge, his/her avoidance of the expression of an overt, unequivocal stance towards the subject matter, and a hearer-oriented discourse organization in terms of an opening-up of the conversational floor to the other participants. (p.1194)

Therefore, it is not a surprising finding that it is commonly used in spoken language as it yields a number of stance related functions. In example 11, *I don't know* is used as an epistemic marker functioning as yielding the floor to other speakers and in 12, it is used to avoid an overt expression:

(12) *<B> so I like children so I went into primary school teaching that was like .. and it was easy here to be able to do a degree in theatre studies which was my little . my little pet subject so <\B>*

<B> it was just like oh I can come and enjoy myself and as well get a degree and have a job at the[i:] end of it <\B>

<B> [so that's good thing about it <\B>

<B> I don't know <\B>

(13) *<B> erm .. I don't know probably I mean somewhere with still like a good education system I don't think I'd want to go to like a third world country <\B>*

The next most common epistemic stance marker in the current database is *would* with 133 occurrences. The other modal verbs found in the data in order of frequency are: *might* (n=63), *could* (n=36), *must* (n=19), *may* (n=9), *will* (n=5), and *should* (n=1). Kärkkäinen (2003) finds a slightly different order of frequency among the modal verbs: *would, must, might, could, will, may and should.* The fact that *would* occurs more than the other modal verbs can be related to the fact that it is very commonly used to express prediction (Biber et al., 1999). In LOCNEC, 133 out of 164 instances of *would* was marked as expressing epistemic uncertainty which confirms the claim proposed by Biber et. al (1999). Concerning *will,* they argue that its function as prediction and volition is highly ambiguous. The ambiguous uses of *will* were excluded from the study to be able to draw a clearer picture of epistemic stance. That could be the reason of infrequency of will in the database. As for *must,* the authors find that *must* is used to mark logical necessity rather than personal obligation in conversation. This seems to be true for the current data since out of 30 occurrences, it marks logical necessity 19 times. Concerning *could* and *might,* it was reported that the logical possibility function is much more common in conversation. Although this seems true for *might* in the current data, where 63 out of 65 occurrences mark logical possibility, this doesn't seem to be the case for *could* as only 36 out of 153 occurrences were marked as conveying epistemic stance (logical possibility) rather than permission and ability functions, which stands in contradiction to findings by Biber et al. (1999). Finally, *should* occurs only once in LOCNEC in relation to its epistemic use, which was also found to be very infrequent by Kärkkäinen (2003).

Although *I think* and *I don't know* are found to be the most frequent epistemic phrases in the data, it is seen in Table 12 that native speakers also make use of other epistemic phrases in their speech. The phrases found in the data in order of frequency are as follows: *I thought* (n=120), *I don't think* (n=46), *I know* (n=43), *I'm (not) sure* (n=40), *I know + complement* (n=30), *S/he said* (n=25), *I said* (n=24), *looks like/to me* (n=18), *I guess* (n=11), and *I remember(ed)* (n=9). Kärkkäinen (2003) finds a very different pattern in her database: *S/he said, I said, I guess, I thought, I remember(ed), looks like/to me, I don't think, I know* and *I know +complement.* Here, this difference can be linked to the fact that the corpus Kärkkäinen draws her data involves face-to-face everyday conversation in American English through SBSCAE corpus and the data in LOCNEC contains informal interviews. This also explains the infrequent use of *I guess* as it was suggested to be more common in American English than British English. The epistemic phrase *I thought* is claimed to indicate a change of epistemic stance (Kärkkäinen, 2012) as can be seen in the example extracted from LOCNEC:

(14) *<B> erm . . at the time when I lived on campus I thought I don't want to live out . because it's so nice being on campus <\B>*

Furthermore, in a different study, Kärkkäinen (2010) states that besides *I think* and *I don't know*; *I know, I thought*, and *I don't think* are among the most frequent epistemic phrases and epistemic phrase *I remember* is less frequent, which supports the findings of the present study. She also includes *I'm sure* into the category of epistemic phrases and argues that it is a less frequently used marker in native speech. In the current database, both the negative and affirmative version of *I'm (not) sure* are included (n=40) as it conveys a clear (un)certainty about the content of the information and that could be the reason why it does not appear as a very infrequent marker in LOCNEC. Additionally, it is clear that the native speakers tend to prefer to use the adjectival expression (*I'm sure*) as an epistemic phrase rather than using the adverbial form (*sure*). Finally, among the epistemic adjectives, *true* is the only example with 29 occurrences. Other epistemic adjectives showing one or zero occurrences in the database are excluded.

In sum, although the native spoken language seems to densely rely on the use of epistemic stance markers, it can be claimed that native speakers tend to follow a routinized pattern in conversation (Kärkkäinen, 2006).

The next section will elaborate on the findings of the contrastive analysis of epistemic stance markers in spoken language in terms of overuse and underuse of particular markers to provide a more clear comparison between native and non-native speech.

4.2.3 Epistemic stance markers in LINDSEI vs LOCNEC

Even though separate analyses of two corpora provide insight into understanding the epistemic uses of the language in native and non-native speech; a rather distinctive picture arises when the two corpora are analysed contrastively. Table 13 presents the raw and normalized frequency of each stance marker in LINDSEI-TR and LOCNEC with the purpose of reaching more reliable comparisons between two corpora with different sizes (i.e. LINDSEI-TR = 80,813 words / LOCNEC = 118,553 words). The results were normalized per 10,000 words 'to allow comparison across corpora of different sizes' (Hyland, 2010). Finally, the last column in the table presents the log-likelihood value of each stance marker so as to enable the identification of overused and underused epistemic markers in LINDSEI-TR in relation to LOCNEC. When calculating the log-likelihood ratio, the cut-off level was chosen at 10.83 which provides the differences between two corpora are significant at $p < 0.001$, which is interpreted as the difference between two corpora happening by chance is less than 0,1%. This means we can be 99,9% certain that the result is meaningful. The values higher than 10.83 are highlighted.

Table 13. Epistemic stance markers in LINDSEI vs LOCNEC

Marker	LINDSEI	LOCNEC	NORM(LIN)	NORM(LOC)	LL
I think	202	432	25	36	-20.3
Maybe	102	86	13	7	+14.3
Of course	60	32	7	3	+22.7
Actually	58	221	7	19	-49.2
In fact	38	6	5	1	+39.8
I don't know	27	166	3	14	-65
Would	25	133	3	11	-45.4
I said	24	24	3	2	+1.75
S/he said	16	25	2	2	-0.04
I thought	13	120	2	10	-63
I don't think	11	46	1	4	-11.7
I guess	11	11	1	0.9	+0.8
Will	10	5	1	0.4	+4.16

May	8	9	0.9	0.7	+0.30
I don't know + comp	8	43	0.9	3.6	-14.8
Could	8	36	0.9	3	-10.1
Sure (adverb)	7	2	0.8	0.1	+5.1
True(adj)	6	29	0.7	2	-8.9
I know + comp	6	30	0.7	3	-9.5
I know	5	43	0.6	4	-21.6
I remember(ed)	4	9	0.4	0.7	-0.5
Probably	3	93	0.3	8	-75.4
Must	2	19	0.2	2	-10.1
Definitely	2	49	0.2	4	-37.6
possibly	2	22	0.2	1.8	-12.7
Might	1	63	0.1	5	-57
Should	1	1	0.1	0.1	0.07
Perhaps	1	35	0.1	3	-29
I'm (not) sure	1	40	0.1	3	-33.9
Looks like/to me	0	18	0	1.5	-18.7
Apparently	0	21	0	3	-21.8
Certainly	0	27	0	2	-28

When the both corpora were analysed with the normalized frequency counts, it becomes apparent that there are certain similarities and differences between two corpora in terms of their exploitation of epistemic stance markers in speech. Concerning the similarities, it is shown in the table that *I said* has 3 normalized frequency in non-native corpus and 2 in native corpus. The log-likelihood value (+1,75) shows that *I said* is slightly overused by non-native speakers but the result is not significant. Another similar stance marker formed with the utterance verb say is *s/he said.* The normalized frequency in both corpora is calculated as 2 with the log-likelihood value -0,04, which suggests that there is not a significant difference in terms of overuse/underuse. *I guess* is the next marker which doesn't display a significant difference between two corpora. The normalized frequencies are 1 and 0,9 for the learner and native corpora respectively with the log-likelihood value +0,8. The value shows that it is only slightly overused by learners, whereas Aas (2011) surprisingly notes that it was significantly overused in Swedish and Norwegian subcomponents of LINDSEI. Here, especially for the learner corpora, the infrequent use of *I guess* can be linked to the effects of input and materials the learners are exposed to. Although it covers the same meaning with *I think,* it is used very infrequently in Turkish EFL learner speech, which suggests that the learners may be making use of the markers they are more familiar with. As a modal verb, *will* has 1

normalized frequency in non-native speech while it has 0,4 in native speech. The log-likelihood ratio is +4,16, which tells that it is overused by non-native learners but the difference is not significant. Although the difference is not significant, this finding is in fact surprising as the data shown in the table only includes the epistemic uses of the markers. It seems that Turkish L2 learners exploit the epistemic use of *will* more than the native speakers. This can also suggest that the native speakers prefer other modal verbs to explain their degree of commitment in speech. The normalized frequency for *may* is 0,9 in learner corpus and 0,7 in native corpus with the log-likelihood ratio +0,30, which shows that the difference is not significant. This shows that the speakers in both corpora choose to employ the marker with its permission/ability function rather than its epistemic meaning. In addition, this result is not interesting since *may* was found to be extremely common in academic prose and very rare in conversation by Biber et. al (1999, p. 487). Another modal verb which represents a similarity between two corpora is *could* with 0,9 and 3 normalized frequencies in learner and native corpora respectively. The log-likelihood value is -10,1, which means that this marker is underused by non-native learners of English. This result indicates that the difference is significant at $p < 0.01$; however, because it shows a ratio under the cut-off level chosen for the purposes of study, this marker was not highlighted as an indicator of significant difference between the two corpora. Furthermore, the cause of underuse of this modal verb may stem from the learners' being more familiar with its permission and ability function. In its adverbial form, *sure* has 0,8 normalized frequency in learner corpus and 0,1 in native corpus with the log-likelihood ratio +5,1. This implies that it is overused by non-native speakers, yet the difference is not significant at $p < 0,001$. As the only epistemic adjective found in the data, *true* has 0,7 normalized frequency in learner corpus and 2 in native corpus. The log-likelihood value is -8,9, which means that although it is underused by Turkish EFL learners, the difference is not significant. The normalized frequencies for *I know + complement clause* as an epistemic marker is 0,7 in learner corpus and 3 in native corpus with the log-likelihood value -9,5. This means although it is underused by learners, the difference is not significant at $p < 0,001$. Next, *I remember(ed)* has 0,4 normalized frequency in learner corpus and 0,7 in native corpus, and the log-likelihood value is -0,5, which means that the difference is not significant at all. It was reported by Kärkkäinen (2010) that among the epistemic parentheticals like *I think* and *I know, I remember* is a less frequent marker of epistemic commitment. *Must*

has 0,2 and 2 normalized frequency in learner and native corpus respectively and the log-likelihood ratio (-10,1) shows that it is underused by learners but the difference is not significant at $p < 0,001$. The normalized frequencies of *should* is 0,1 for both corpora with the log-likelihood value 0,07, which doesn't indicate a difference in terms of overuse/underuse.

To summarize, the epistemic markers that imply similarities between LINDSEI-TR and LOCNEC are as follows according to their grammatical classes: epistemic phrases (*I said, s/he said, I guess, I know + comp, I remember(ed)*), modals (*will, may, could, must, should*), *sure* as an epistemic adverbial, and *true* as an epistemic adjective.

Even though it was indicated that there are some similarities between native and non-native corpus concerning the use of epistemic stance in speech, significant differences were also identified. As Table 13 displays, *I think* as a very common epistemic phrase has 25 normalized frequencies in learner corpus and 36 for native corpus with the log-likelihood value -20,3. This means that it is underused by Turkish EFL learners at a significant level. This stands as a surprising finding as it was reported to be an overused epistemic marker by a number of studies (Baumgarten & House, 2010; Aas, 2011; Huang, 2014; Aijmer, 2004). Especially, Aas (2011) puts forward that in the Swedish and Norwegian subcomponents of LINDSEI compared to LOCNEC, *I think* is overused by the learners, which provides a fruitful resource to compare the findings of the study with the other subcomponents of LINDSEI. Kärkkäinen (2003) identifies three main functions of *I think* as an epistemic stance marker: 1) starting point function either to bring in speaker's personal perspective along with topic shifts and presentation of new perspectives on the same topic or to display an increased involvement and commitment; 2) a marker of on-line planning; 3) or a marker indicative of the completion of the turn at talk. In addition, Baumgarten & House (2010) states that '*I think* is the most common means of overtly encoding a subjective perspective in discourse, but it is not possible to conclude that L2 speakers use the collocation indiscriminately as an all-purpose, anywhere element'. Aijmer (2004) notes that 'learners use vague and uncertain markers to express uncertainty or hesitation and not for face-saving or to signal politeness'. Depending on the varied functions proposed in the literature, Turkish learners' underuse of this marker can be explained with their

lack of awareness of the different functions of *I think*. This finding calls for the need to raise the learners' awareness about the different uses of *I think.*

Similarly, *I thought* and *I don't think* are the underused patterns in the database. I *thought* occurs with 2 normalized frequency in LINDSEI-TR and with 10 in LOCNEC. The log-likelihood ratio is calculated as -63, which refers to the fact that it is significantly underused in learner speech. The normalized frequencies calculated for *I don't think* are 1 and 4 for learner and native corpus respectively with a log-likelihood value of -11,7. However, this combination was also found to be significantly overused in Swedish component of LINDSEI by Aas (2011). *I know* occurs with 0,6 normalized frequency in learner corpus and 4 in native corpus. The log-likelihood ratio (-21,6) reveals that the marker is significantly underused in the speech of Turkish learners. Normalized frequency of *I'm (not) sure* in learner corpus is 0,1 and 3 in native speech with the log-likelihood ratio -33,9, which reflects a highly significant underuse in L2. Finally, *looks like/to me* as an epistemic phrase shows up with zero occurrences in learner corpus and 1,5 occurrences in native corpus with a log-likelihood ratio -18,7. This means that it is also highly underused in learner corpora. These findings suggest that Turkish learners of English need to be informed about the other combinations of the common lexical verbs such as *think* and *know* and also about the different ways of expressing epistemic (un)certainty.

Maybe as an epistemic adverbial occurs with 13 normalized frequency in learner corpus and 7 in native corpus with a log-likelihood value +14,3. Considering the results, it is seen that *maybe* is overused by non-native learners and the difference is significant. The similar findings are proposed by Gablasova & Brezina (2015) who found out that maybe is overused by L2 speakers. The reason of overuse of *maybe* in learner corpus can be due to the mother tongue effect. In Turkish (belki), it is a very frequently used marker used to express uncertainty and to lower the personal commitment to the information proposed in interaction. In parallel with this, it could be useful for learners to make them aware of its uses in L1 context.

Of course stands as a very frequent marker in both LINDSEI-TR and LOCNEC. When the normalized frequencies in both corpora are taken into account, it is seen that it has 7 occurrences in learner speech while it has 3 occurrences in native speech. The log-likelihood value (+22,7) shows that it is highly overused by L2 speakers. This finding

gains support from De Cock (2004). Investigating the French learners of English, she reports that non-native speakers tend not only to overuse this marker but also misuse it by employing the marker as *yes/yeah of course* 'to answer a request for information or to respond to an opinion expressed by another speaker'. She adds that major learner dictionaries emphasize that using this combination as a response to a request for information 'would sound as if you think the answer to this question is very clear and you think the person is stupid to need to ask you'. (p.242) Considering the fact that these kinds of examples are also found in LINDSEI-TR, it is clear that the L2 learners need to revise their knowledge on the uses of *of course.* In addition, an interesting finding from the learner corpus shows that it is employed as a hedge rather than as a boosting device to increase the speaker's commitment towards the proposition. In the example below, the participant uses it to lower his/her commitment and to avoid sounding too assertive:

(15) *<B> it it this is my opinion <u>of course</u> but I think like that (mm) </B>*

Another epistemic adverbial *actually,* has 7 and 19 normalized frequencies in learner and native corpora respectively with the log-likelihood ratio -49,2. This reveals that *actually* is a highly significant underused item in learner corpus. Given that this item is more peculiar to spoken language, it becomes clear that the Turkish EFL learners should be informed on this particular item. A very similar epistemic adverbial *in fact* has 5 normalized frequency in learner corpus while it has 1 in native corpus. The log –likelihood value (+39,8) suggests that it is highly overused in learner speech. It was mentioned in the previous section that *in fact* is more common in formal speech. It is possible to claim that as Turkish learners of English are much more exposed to formal language in instructional settings, they tend to transfer their knowledge of this marker to speech. Aijmer (2002) argues:

'while in fact reinforces the opposition to the preceding proposition by implying that a stronger claim needs to be made, actually seems to signal that some slight reinterpretation or adjustment of perspective needs to take place which may be seen as a shift in common ground'.

The L2 learners doesn't seem to be aware of this distinction between the functions of *actually* and *in fact.* Furthermore, a number of other adverbials seem to be underused by Turkish EFL learners. *Probably* as a very frequent marker in spoken

English, has 0,3 normalized frequency in learner corpus while it has 8 in native corpus. The log-likelihood value is given as -75,4, which reflects a highly significant underuse in learner speech. Similarly, *definitely* occurs 0,2 times in LINDSEI and 4 times in LOCNEC with the log likelihood value -37,6. *Possibly* has 0,2 normalized frequency in learner data and 1,8 in native data with the log likelihood value -12,7. *Perhaps* occurs 0,1 times in learner corpus and 3 times in native corpus with a log likelihood value -29. *Apparently* doesn't occur in learner corpus but it occurs 3 times in native corpus. The log-likelihood value is calculated as -21,8. Finally, *certainly* is also not used in learner corpus but it has 2 occurrences in native corpus according to normalized frequency counts with the log likelihood value -28.

As can be seen clearly, excluding *maybe, of course,* and *in fact,* the comparison of two corpus reflects that all of the epistemic adverbials in the database are underused by L2 speakers. Given that the epistemic adverbs constitute 40% percentage of all epistemic markers deployed in LINDSEI-TR, it can be claimed that the Turkish learners heavily rely on a very few adverbials to qualify their commitment to the truth of proposition.

Given the normalized frequencies of *I don't know* as 3 in learner corpus and 14 in native corpus with the log-likelihood ratio -65, it is apparent that this marker of uncertainty is highly underused by L2 speakers. Again, this doesn't correspond to the findings of Aas (2011) who indicated that *I don't know* was overused by Swedish learners of English. Baumgarten & House (2010) argues the different uses of I don't know in L1 and L2 with the following words:

> While the L1 speakers use *I don't know* to express their subjective perspective on the proposition or sequence of propositions in terms of the avoidance of the overt expression of an unequivocal stance, the L2 speakers use *I don't know* to overtly state their insufficient knowledge and to signal their being momentarily at a loss for words. (p. 1197)

Aijmer (2009) asserts that the end position of *I don't know* has the function of 'yielding the floor or fulfilling the desire of the interviewee to close a topic', plus it signals uncertainty or avoidance of taking responsibility. Given the evidence from previous research, it seems necessary to inform the learners about the different functions of the marker.

Would constitutes one of two modal verbs in the study which represents a significant difference between two corpora. In terms of its normalized frequencies, it has 3 occurrences in learner corpus and 11 in native corpus with the log-likelihood ratio -45,4. Similarly, *might* has 0,1 normalized frequency in learner corpus while it has 5 in native corpus with the log-likelihood value -57. This finding reveals that the learners are not aware of the epistemic meanings of modal verbs. Also, the fact that epistemic meanings are very difficult to acquire for even advanced learners may explain the reason of underuse.

An overall look on the employment of epistemic stance markers in the speech of Turkish EFL learners displays that the L2 learners follow a more restricted pattern than L1 speakers in expression of epistemic stance. Table 14 summarizes the overused and underused patterns in LINDSEI-TR.

Table 14. Overused and underused epistemic stance markers in LINDSEI-TR

LINDSEI-TR	
Overused epistemic markers	*Maybe, of course, in fact*
Underused epistemic markers	*I think, actually, I don't know, would, I thought, I don't think, I don't know + complement, I know, probably, definitely, possibly, might, perhaps, I'm (not) sure, looks like/to me, apparently, certainly*

As the table displays, the most of the epistemic markers in LINDSEI-TR is underused in relation to LOCNEC with the exclusion of only three epistemic markers overused: *maybe, of course, in fact.* The overall findings of the study show that the L2 learner tend to make use of a narrower range of epistemic markers in their speech when compared with the native speech. One reason that most epistemic markers are underused by Turkish EFL learners could be due to collective conscience and cultural characteristics of Turkish. It has proposed in the extant literature that the 'explicit instruction is more effective than implicit instruction in developing learners' use of epistemic stance in writing in the short- and long-term' (Fordyce, 2014). However, it is also reported that epistemic stance in spoken communication of L2 learners is an area which requires more research (Gablasova et al., 2015). Given that pragmatic competence plays a crucial role in knowledge of an L2, a corpus-informed approach to stance markers can provide well-based steps for practitioners to take when teaching

epistemic stance taking in conversation. In line with this purpose, the Study 2 reveals findings on whether the explicit teaching of stance markers identified through contrastive analysis is more effective in L2 educational settings.

4.3 Study 2: Corpus-Informed Intervention

Study 2 set out to explore the effect of, if any, corpus-informed intervention on the use of epistemic stance markers in Turkish EFL learner's spoken production.

The findings of the present study are presented in five parts: 1) participants' overall knowledge of epistemic stance markers at three different time points (measured by their performance on pre-test, the post-test, and the delayed post-test; 2) effects of corpus-informed intervention on the sub-levels of epistemic marker frequency counts and epistemic range scores; 3) the individual epistemic stance markers before and after the treatment; 4) learners' preference of epistemic marker use in different task types; 5) participants' attitudes towards the use of concordance in learning epistemic stance markers in spoken English.

4.3.2 Overall Learning Effects of Concordance Use on Learning Epistemic Stance Markers

To measure the learners' overall progress after the treatment, sums of the frequency counts and range scores of epistemic markers at three different tests employed were calculated for each participant. The paired t-test comparison of the pre-test results and post-test results pointed that the learners showed a significant progress ($p=.002 < .05$) in epistemic marker use in general.

Table 15. Pre-test vs. Post-test: overall scores

	N	Mean	S.D.	T	Df	Sig.
The pre-test overall	34	9.64	4.55	-3.38	33	.002*
The post-test overall	34	13.29	7.74			

*Note. * $p < .05$.*

As the table shows, learners' overall epistemic stance marker knowledge (both frequency and range) increased significantly in the immediate post-test after the 12-session treatment. This finding is in line with the findings of Fordyce (2014), who found

out that the learners showed a sharp increase in their range of epistemic markers used in writing after the explicit intervention. At this point, considering that the learners had different proficiency levels measured by the scores they obtained from the proficiency test given before the treatment, Analysis of Covariance (ANCOVA) was conducted to determine if the proficiency level has a statistically significant effect on post-test overall scores once the pre-test results are controlled. The independent variable was the immediate overall post-test scores, and the dependent variable was the proficiency levels of the learners obtained by the proficiency test administered before the intervention. Descriptive statistics indicated that out of 39 participants, 8 of them were at A2 level, 30 of them were at B1 level and only one student was in B2 level. Since there was only one participant at B2 level, this participant was excluded from the analysis and ANCOVA was conducted to compare the participants in two groups, which are A2 (lower level) and B1 (upper level) level of proficiency. Participants' overall scores on the pre-test were used as the covariate in the analysis.

Preliminary analyses were conducted to make sure that the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate were not violated. Upon confirming if the assumptions hold, the results of the ANCOVA indicated that there was no significant difference between the lower level group and upper level group once the pre-test results are controlled $F (1, 30) = .91$, $p = .34$, partial eta squared = .03. Table 16 presents the results obtained from ANCOVA.

Table 16. Analysis of covariance for overall epistemic marker use by proficiency level

Source of Variance	Sum of Squares	df	Mean Squares	F	Partial Eta Squared
Pre-test overall	553.341	1	553.341	13.220	.306
Proficiency	38.202	1	38.202	.913	.030
Error	1255.699	30	41.857		

As shown in the table, no significant difference was found between the participants at A1 level and participants at B1 level. This could be interpreted as that the effect of the treatment was independent of the proficiency levels of the participants. Therefore, it can be claimed that both groups of students according to their proficiency level benefited from the intervention. This finding is consistent with the findings of

Fordyce (2014) who reported that the variety of epistemic markers utilized by high, mid, low proficiency groups in the explicit intervention was promoted in the immediate post-test.

In a different study, that is not in parallel with the findings of this study, Oh and Kang (2013) propose that epistemic modality demonstrates itself in the written language as a variant feature for different proficiency levels. In the same line, Hu and Li (2015) find out that the learners at different proficiency levels show different reactions to the use of epistemic devices in writing and the epistemic markers they prefer to employ in their writing differs. Although this finding is beyond the scope the present study, it can be proposed that both the lower level groups and higher level groups benefited from the explicit intervention. However, since no studies exploring the effect of explicit teaching of epistemic stance markers in particularly spoken language was found, it was not impossible to compare the findings obtained from the current study.

To measure the overall effect of instruction, the overall results of the post-test and delayed post-test were compared to see how much of the epistemic marker knowledge was retained after 4 weeks following the immediate post-test. Results of the paired samples t-test showed that the learners regressed 4 weeks after the intervention. Table 17 presents the results.

Table 17. Retention of epistemic stance marker knowledge

	N	Mean	S.D.	T	Df	Sig.
The post-test overall	27	14.48	7.92	3.41	26	.002*
The delayed post-test overall	27	10.07	4.83			

Note. * $p < .05$.

Table 17 shows that students' overall retention of epistemic stance markers between the immediate post-test and delayed post-test decreased significantly ($p = .002 < .05$). This finding was expected owing to the fact that the treatment was not on written language but on spoken language. Although the tasks were chosen to stimulate the learners to use epistemic markers of certainty and uncertainty in nature, the learners were not guided or urged to use the target devices in their oral production during the

tests. Therefore, it can be asserted that the students' awareness of the target forms decreased after the treatment as the content of the tests were not controlled and the learners were not explicitly exposed to the epistemic stance devices after the treatment.

In addition, to examine the residual effects of the treatment (the comparison of overall pre-test and delayed post-test results), another paired samples t-test was run. It was hypothesized that the learners would regress in the delayed post-test, but they might still be better when compared to their pre-test results. T-test results indicated that even though the means in the delayed post-test were only slightly higher than the means in the pre-test, the difference between the two time points were not statistically significant ($p > .05$). This finding could be attributed to the fact that the students may have felt under pressure both metacognitively and psychologically during the speaking tests. Similar findings are reported by Jones and Carter (2014) who measured effectiveness of PPP and III methods in learning spoken discourse markers. They state that the PPP approach was effective in improving the ability of the learners to use discourse markers in the short term but they were not retained in the long term. In another study conducted by Fordyce (2009), which compares the epistemic stance markers in written and spoken language of L2 learners, it is argued that while the learners were better at expressing their epistemic stance in writing, which is off-line in nature, they had more difficulty in exploiting the same variety of epistemic stance markers in speaking which requires on-line processing.

In sum, it seems that the participants of the present study have benefited from the explicit treatment through concordancing in general. However, it can be argued that the treatment was more efficient in the short term as the epistemic stance marker knowledge was not retained in the long term. Therefore, it implies that the target epistemic forms have to be revised within a regular time line (Jones & Carter, 2014).

4.3.3 Effects of Intervention on the Frequency Counts and Range of Epistemic Stance Markers

Although comparisons of the overall scores of the learners at three time points provides insight into the effectiveness of the treatment employed in this research, examining and comparing the data separately for the frequency counts and range of the epistemic markers used in three tests might draw a clearer picture to understand the learner gains.

Firstly, comparisons were made to see the differences, if any, between the frequency counts of epistemic forms in pre-test, post-test, and delayed post-test. The paired samples T-test results demonstrated similar findings to the overall scores of the learners explained in the previous section.

Table 18. Comparison of data in terms of frequency counts

	N	Mean	S.D.	T	Df	Sig.
The pre-test frequency	34	6.88	3.73	-2.73	33	.010*
The post-test frequency	34	9.20	5.93			
The post-test frequency	27	10.11	6.13	3.12	26	.004*
The delayed post-test frequency	27	6.81	3.57			

Note. * $p < .05$.

As shown in Table 18, when the frequency counts of target epistemic forms were compared for pre-test and post-test scores, it can be seen that the learners tended to deploy epistemic devices more frequently after the treatment and the difference was statistically significant ($p < .05$). That result can be explained as the corpus-informed teaching through concordancing can be effective in raising the learners' awareness of the target forms. In the relevant literature, it is compromised that the explicit teaching methodologies may have an influence on the language of L2 learners 'which they may not always pick up from the English-speaking environment' (Jones & Carter, 2014).

When it comes to the comparison of frequency counts in post-test and delayed post-test, t-test results showed that the participants utilized much less epistemic forms in number in the delayed post-test and the mean difference was significant ($p < .05$). That means when the learners are not exposed to the epistemic devices regularly, they tend to express their stance using less epistemic forms in spoken communication.

However, depending on the in depth analyses of individual differences, it can be claimed that the range of epistemic markers employ in their speech could be a better indicator of epistemic stance taking achievement in speech as the main goal of the treatment was to increase the range rather than the frequency of the epistemic markers used by EFL learners. After the treatment, it was expected that the learners would be able to convey their epistemic stance with a wider variety of epistemic devices and with

more precision rather than depending on only one or two forms heavily. Examples below from the pre-test and post-test results of one participant in the study could provide a better interpretation to the issue:

Participant 18: Pre-test scores : Frequency score: 11 Epistemic range score: 4

The epistemic devices used: *I think* (4 times), *of course* (3 times), *maybe* (2 times), *actually* (2 times)

Participant 18: Post-test scores: Frequency score: 13 Epistemic range score: 7

The epistemic devices used*: I think* (5 times), *probably* (3 times), *maybe* (1 time), *actually* (1 time), *apparently* (1 time), *I'm sure* (1 time), *of course* (1 time)

As seen in the example above, although the participant still keeps relying on using *I think* heavily, she used 3 different markers after the intervention, which means that she improved her ability to express her stance with some other epistemic forms that she was not aware of before. Another example from another participant reflects the difference after the intervention more clearly:

Participant 6: Pre-test scores: Frequency score: 6 Epistemic range score: 2

The epistemic devices used: *I think* (2 times), *maybe* (4 times)

Participant 6: Post-test scores : Frequency score: 8 Epistemic range score: 5

The epistemic devices used: *I think* (2 times), *would* (2 times), *might* (2 times), it looks like to me (1 time)

As can be inferred above, while the participant only depends on using *I think* and *maybe* in the pre-test, she preferred to employ other epistemic devices, especially the modals in the post-test correctly, which are claimed to be difficult linguistic items to acquire their epistemic meaning for L2 learners. (Palmer, 2001)

The examples given above indicates that other than the frequency scores of the participants, also the epistemic range scores of the students should be examined separately. Table 19 presents the results for paired-samples t-test on epistemic range.

Table 19. Comparisons of pre-test and post-test results on epistemic range scores

	N	Mean	S.D.	T	Df	Sig.
The pre-test range	34	2.76	1.15	-4.23	33	.001*
The post-test range	34	4.08	2.16			
The post-test range	27	4.37	2.18	3.60	26	.001*
The delayed post-test range	27	3.25	1.60			

Note. * $p < .05$.

Table 19 indicates that there was a highly significant difference on the epistemic range scores of the participants after the intervention ($p < .001$) and the learners regressed 4 weeks after the post-test.

In the previous section, it was found that the participants' overall scores did not show a statistically significant progress when the pre-test and delayed post-test results were compared. Namely, the learners were very close to their entry levels in their delayed-post-test scores and they tended to forget the forms they have been exposed to in the treatment sessions. However, when the range scores of the participants were compared for the pre-test and delayed post-test, the results were different. Table 20 shows the results.

Table 20. Residual effects of the intervention on epistemic range scores

	N	Mean	S.D.	T	Df	Sig.
The pre-test range	31	2.77	1.23	-2.28	30	.030*
The delayed post-test range	31	3.32	1.53			

Note. * $p < .05$.

As can be understood from the table, even though the students showed a regression between the post-test and delayed post-test, their range scores in the delayed test were still better than their pre-test range scores. Depending on the results given above, it can be argued that the learners had gains from the explicit teaching of stance markers through corpus-based tools widening the variety of epistemic forms they make use of in their oral production.

4.3.4 Individual Epistemic Stance Markers in Pre-test vs. Post-test

For Study 2, it was determined through the corpus analysis that 19 stance markers should be assigned to the explicit treatment. When the pre-test results were explored for the individual epistemic stance markers, it was observed that the students used 269 epistemic stance markers in total with only 10 different types of devices. Figure 8 provides the frequency counts of individual epistemic forms employed by the participant of the study.

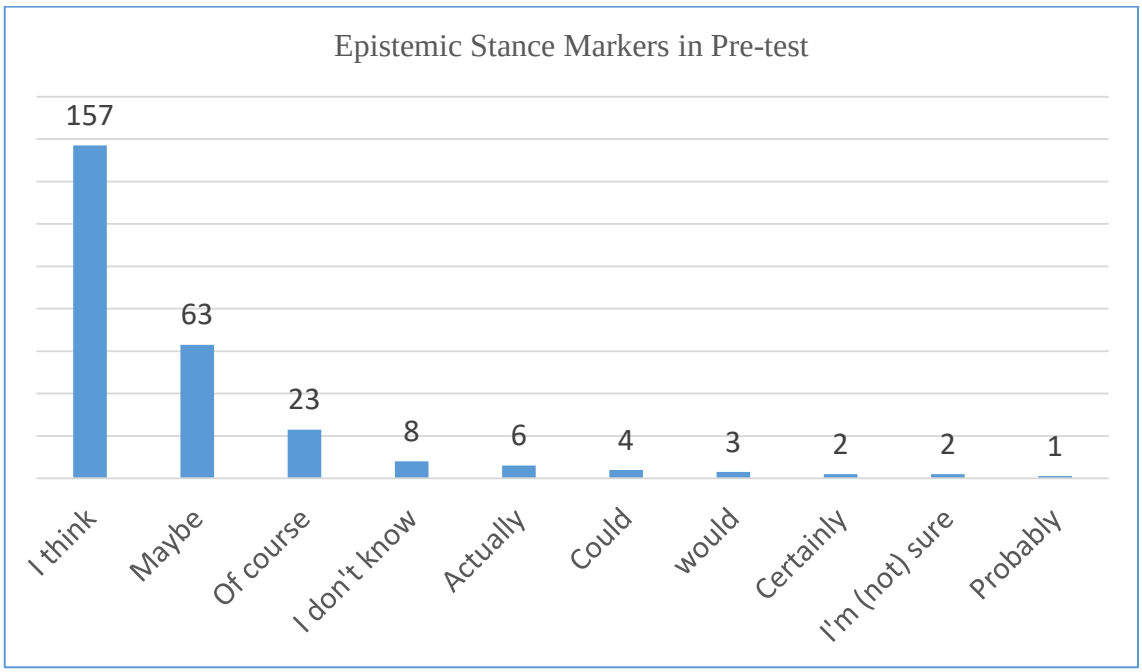


Figure 8. Individual Epistemic Stance Markers in Pre-test

As seen in Figure 8, in the pre-test, the students exploited only 10 epistemic markers, especially *I think* and *maybe* with higher loadings. Out of 269 epistemic markers in total, the participants used *I think* 157 times, *maybe* 63 times, *of course* 23 times, *I don't know* 8 times, *actually* 6 times, *could* 4 times, *would* 3 times, *certainly* 2 times, *I'm (not) sure* 2 times, and *probably* only once. These results are very similar to the findings obtained from the analysis of LINDSEI-TR. It was found in LINDSEI-TR that *I think, maybe* and *of course* dominated the spoken language of L2 learners. In the relatively small sample participated in this study, the situation was the same. However, the EFL learners in LINDSEI-TR did not mark their stance with certainly while it was used by the participants included in this study although it was marked with only 2 frequency.

Overall, this finding concerning the individual stance markers deployed by Turkish EFL learners presents crucial facts about learner corpora. Depending on the frequently marked epistemic devices found in the speech of the participants of the study, it can be argued that the learner corpora are very strong predictors of the language the second language learners use and that means findings obtained through the analysis of learner corpora is highly representative of the linguistic forms commonly used by different groups of learners.

Given the results from the individual epistemic marker analysis in the pre-test, the same type of analysis for the post-test would reveal some other facts about the effectiveness of adopting an explicit teaching method through the use of concordance lines in teaching epistemic markers to second language learners. The manual analyses on the post-test showed that while the learners used 269 markers in the pre-test, the learners employed 321 epistemic markers in total. Figure 9 provides the results.

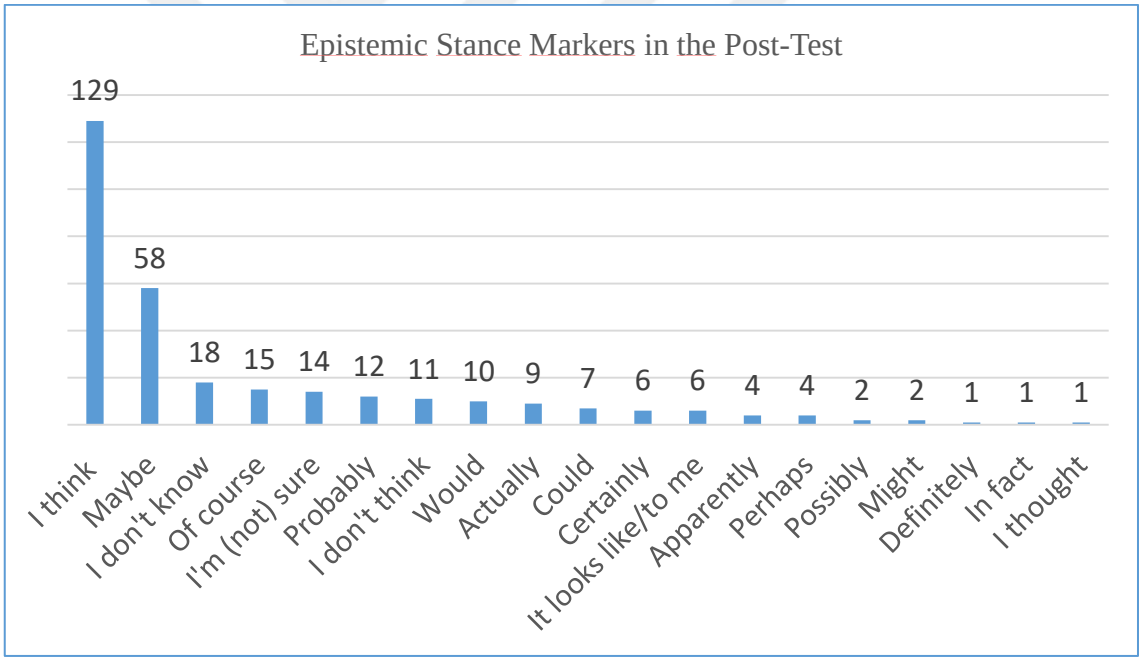


Figure 9. Individual Epistemic Stance Markers in Post-Test.

As seen in the graph, although the participants still used *I think* and *maybe* very frequently, all 19 epistemic forms addressed in the explicit teaching sessions were used by the participants. After the intervention, the students not only increased the frequency of epistemic markers they used in oral production, but they also started to mark their certainty or uncertainty with different epistemic devices. It can be observed by

examining the figures above that while the participants exploited 10 different epistemic types before the intervention, they were able to use 19 different epistemic types in the post-intervention test. Moreover, it should be noted that the intervention urged the participants to mark their (un)certainty by other epistemic forms like *I guess*, and *it is possible* that was not addressed in the intervention stage. This finding confirms the claim proposed by Fordyce (2014) that *noticing* is a highly vital factor in pragmatic development of L2 learners. In return, this means, it is obvious that the EFL learners are able to put into practice the linguistic forms in their oral production, which are already in their schemata, but needed to be activated through explicit teaching methodologies.

4.3.5 Learners' Epistemic Marker Use According to the Task Types

All three tests conducted (pre-test, post-test, delayed post-test) had the same format which consisted of three parts. In the first task, the students were directed an open-ended question like 'Do you think English will be useful for you in the future?' as a warm-up question that could lower their anxiety during the spoken test. In the second task, the students were given a set of pictures that required them to make predictions on what is likely to happen. In the last task, a situation was read aloud to the students. Then they were given a set of pictures as clues to make them predict the possible options to the given situation. (See Appendix C) In order to examine in which task type the students tended to employ stance markers more frequently and with a wider variety, descriptive statistics was run for each task type. Table 21 provides the results for the frequency counts of epistemic stance markers in three task types.

Table 21. Frequency information of epistemic devices in three task types

	N	Min.	Max.	**Mean**	Std. Deviation
Task 1 frequency	39	0.00	9.00	**3.51**	2.01
Task 2 frequency	39	1.00	23.00	**9.10**	4.68
Task 3 frequency	39	1.00	23.00	**7.87**	5.34

When the means for different task types are investigated, in task 1, it is seen that the participants employed 3 epistemic forms on average. In task 2, they used approximately 9 forms, and in the last task, nearly 7 epistemic forms were used. As seen, Task 2 is the one which the participants used epistemic stance devices the most frequently. In addition, the overall range of epistemic markers deployed in each task was analysed. Table 22 displays the results.

Table 22. Overall range of epistemic devices in three task types

	N	Min.	Max.	**Mean**	Std. Deviation
Task 1 range	39	0.00	4.00	**2.01**	.95
Task 2 range	39	1.00	7.00	**3.56**	1.77
Task 3 range	39	1.00	7.00	**3.05**	1.55

The range of epistemic markers in the first task was about 2, in the second task 3,5 and in the last task, it was about 3. Again, the findings are similar to the frequency information although the range scores in task 2 and task 3 are quite close to each other. In task 2, the learners used approximately 3,5 epistemic types, which outnumbered the range score in other tasks.

In Task 1, since the students were asked to give an answer to one open ended situation, they may have not felt the need to express their certainty or uncertainty as the answer for them was quite certain that the English would be useful for them in the future. However, there is one point in this part worth considering. In the pre-test, 17 out of 39 students answered this question with the epistemic marker *of course*. As explained in section 4.1.3, it is claimed in the extant literature that non-native speakers tend to misuse this marker by employing it as *yes/yeah of course* 'to answer a request for information or to respond to an opinion expressed by another speaker (De Cock, 2004). However, in the post-test, only 6 participants misused *of course* but instead they preferred to use markers such as *certainly*, and *I'm sure*. This issue about *of course* reveals three crucial things. One of them is that the learner corpora is a quite strong predictor of EFL learner language from different backgrounds as the similar findings were reported from the analyses of LINDSEI-TR and other subcomponents like French (De Cock, 2004). Secondly, the LINDSEI corpus was composed of spoken production of advanced learners and in the present study, the participants were at A2 and B1 level. It indicates that the Turkish learners have difficulty in using *of course* with its correct epistemic meaning regardless of their proficiency levels. The last issue concerns the fact that explicit teaching methodology adopted in the present study helped the students to become aware of their errors in the use of the marker *of course* and to correct it.

With regard to Task 2 and 3, the reason why the participants preferred to deploy more epistemic forms in Task 2 might be related to the task requirements. In the last task, as explained above, the participants were read aloud a situation and a set of pictures to use as ideas when having some guesses on the given situation. The students

first needed to comprehend the text read to them, then to utter their predictions through the picture clues. Whereas, in Task 2, the students were again given a picture but this time they needed to predict on the possible things that could happen / could be happening in the pictures. Hence, the student responses to Task 2 seemed to be relatively more flexible as the responses to task 3 were slightly more controlled.

4.3.6 Learner Attitudes Towards the Use of Concordancing in Learning Epistemic Stance Markers in Spoken Language

The last question this research project sets out to answer was "What are the Turkish EFL learners' perceptions of the use of concordancing in learning epistemic stance markers?". Necessary data for the answer was collected through a questionnaire which consisted of two parts: 1) background information 2) reactions to using the BNC Corpus. The first part involved 12 questions that were related to general computer and Internet use and language preference of the participants. The second part, the actual scale, consisted of 30 items divided into three domains: strengths of corpora in learning epistemic stance markers, negative sides of corpora in general use, and general and future use of corpora. In total, 31 students completed the questionnaire.

As for the first part of the questionnaire, the first question was "In general, do you like to use computer?" and the 84% of the students answered yes and 16% of the students answered no. So, most of the students had positive attitudes towards using computer in general. Second question was about how often the participants use computer for personal purposes. 35% of the students answered they use computer a few times a week, 25,8% of them answered a few times a day, 19,4% of them answered a few times a day, 3,2% of them once a week, 3,2% of them once a month, and 12,9% of them answered they use computer for personal purposes rarely. The third question was again about how often they use computer for their school work. It was found that 29% of the students answered once a week, 19,4% of the students answered a few times a day, 19,4% of them once a day, 16,1% of them a few times a day, 9,7% of them once a month, and 6,5% of them answered rarely. As for the next question which was related to what language they prefer to use when using computer for personal purposes. 64,5% of the students said they use Turkish, and 35,5% of them said they use both English and Turkish when they use computer for personal purposes. The sixth question concerned how much of their total computer time was in their native language. 74,2% of the

students reported that they use Turkish for 75% of their total computer time. 12,9% of them said they used Turkish for 50%, 12,9 of them said they used Turkish for 25% of their total computer time. 100% of the students reported that they had access to Internet at their home or dormitory for the next question. Seventh question was “Do you use the Internet to improve your speaking ability?”. 77,4% of the students answered positively and 22,6% of the students answered this question negatively. Also, 90,3% of the students stated that they found the Internet useful for improving their speaking ability, and 9,7% of them did not find it useful. For the next question, 84% of the students reported that they haven’t heard about corpora before they participated in this research, and 16%of them reported that they haven’t heard about corpora before. However, none of the students have used corpora before the intervention part of the study.

In sum, it can be claimed that most of the students owned positive attitudes towards the use of the Internet for personal purposes and for their field-related studies, but they were not familiar with corpus-based language learning tools.

As explained above, the actual questionnaire attempted to measure the student reactions to the use of BNC corpus in learning epistemic stance markers in spoken language. It included a total of 31 statements on a 7-point Likert scale. For the analysis, mean scores of the responses were calculated. For the ease of interpretation and enhancement of the data presentation, responses were coded into three main categories as “agree”, “disagree” and “no opinion” by assigning all positive answers into “agree” and all negative responses into the “disagree” categories.

The first 8 questions in the questionnaire was related to using concordance lines for learning epistemic stance devices. The results are presented in Table 23.

Table 23. Perceptions towards the use of BNC for learning epistemic stance markers

	AGREE %	DISAGREE %	N.O. %	p
1. Using the corpus is helpful for learning the meaning of EM's	83.9	16.2	0.00	.000
2. Using the corpus is helpful for learning the usage of EM's	87.1	12.9	0.00	.000
3. Using the corpus is helpful for learning the function of EM's	86.7	13.3	0.00	.000
4. Using the corpus has improved my understanding of certainty in spoken language	87.1	22.6	3.2	.000
5. Using the corpus has improved my understanding of uncertainty in spoken language	90.3	6.5	3.2	.000
6. I believe that I can express my stance appropriately when speaking after this instruction	87.1	9.7	3.2	.000
7. The use of concordance lists challenged me to actively make generalizations about the function of a marker.	87.1	12.9	0.00	.000
8. Concordance was useful for learning the stance markers in spoken language.	93.6	6.4	0.00	.000

As Biber (2006) stated, stance is a vital mechanism of language that is marked more frequently in spoken language than written language. However, it can be inferred from the corpus-based analyses that the non-native learners of English do not employ stance markers very frequently in their speech. Kärkkäinen (1992) suggests that this lack of epistemic devices in learner language may result from the lack of explicit teaching since they are very difficult to acquire because of having multiple-meanings. As seen in Table 23, most of the students found corpus very useful to better understand the usage, function and meaning of epistemic markers after they were introduced with the corpus-based tools to learn epistemic forms in speaking. Walsh (2010) proposes that the language learners encounter a lot of problems in speaking and listening and corpus can be very useful for learners to cope with these problems. According to the responses of the participants of the study, the corpus-informed intervention managed to help the students learn to express their stance in speech. In addition, they stated that the corpus helped them to understand how the certainty and uncertainty is expressed in spoken communication.

The next domain in the questionnaire was on the difficulties the learners had when using concordance lines. This domain included 9 items in total. Table 24 provides the student responses to the items.

Table 24. Perceptions on the difficulties on using BNC

	AGREE %	DISAGREE %	N.O. %	p
9. I have some difficulty in using the corpus due to time and effort spent on analysing the data	26.6	66.7	6.7	.000
10. I have some difficulty in using the corpus due to unfamiliar vocabulary on concordance/collocate output	51.7	45.1	3.2	.000
11. I have some difficulty in using the corpus due to cut-off sentences in concordance output	43.3	50	6.7	.000
12. I have some difficulty in using the corpus due to too many sentences in concordance output	45.1	51.6	3.2	.000
13. I have some difficulty in using the corpus due to the limited number of sentences in concordance output	19.3	74.3	6.5	.000
14. I have some difficulty in analysing concordance output	41.9	54.9	3.2	.000
15. I have some difficulty in analysing output for EM's	38.8	58	3.2	.000
16. I have some difficulty in performing the search technique	40.1	56.6	3.3	.000
17. The real texts in the corpus are too difficult to understand	33.3	50	16.7	.000

In this part, the learner responses seem to fall in two sides. It is obvious that for most of the items here, while nearly half of the students reported that they have difficulty in using/ analysing concordance output, the other half said that they did not find it that difficult to search for an item in the corpus. When the responses to the question "I have some difficulty in using the corpus due to time and effort spent on analysing the data" are examined, it is seen that 66,7% of the students disagreed that the use of corpus was difficult because of time and effort spent on analysing the data. In addition, the learner responses to the question number 13, showed that 74,3% of the students did not agree on that the sentences in the concordance output was limited. However, the findings from this domain are consistent with some other studies

commenting on the student perceptions towards the difficulties of using corpus in language learning. Girgin (2011) investigated the effectiveness of using corpus-based tools on grammar learning with lower level Turkish EFL learners. It is suggested in the study that most of the learners had differing, and uncertain opinions about the difficulty of using concordance lines in learning grammar. Additionally, it is stated that the Turkish learners reported that they needed assistance from the researcher as to analyse the concordance output. In parallel, Aşık, Vural and Akpınar (2015) explored the Turkish EFL learner beliefs towards the use of DDL activities through COCA corpus. It is put forward that the learners found corpus very time consuming, hard to understand the data because of the unknown vocabulary, and not very user-friendly. Depending on the learner perceptions reported in these studies, the Turkish EFL learners have common difficulties in using and analysing the corpus data. Considering that the texts in a spoken corpora include pauses, cut-off sentences and a lot of hesitations, it is not surprising that the texts challenged them to understand the stance markers in general through concordance output. However, the corpus can be helpful to motivate students in terms of speaking English as Walsh (2010) asserts “when we look at a corpus, we find that native speakers also hesitate a lot, are not always coherent, frequently use shorter turns, and may use a fairly narrow range of vocabulary” (p. 336)

The last domain of the questionnaire was on attitudes towards the general nature of using corpora. Table 25 displays the learner responses to the items in this domain. This part included 13 items in total.

Table 25. Perceptions towards the general use of BNC for learning epistemic stance markers

	AGREE %	DISAGREE %	N.O. %	p
18. The corpus is more helpful than a dictionary for my English speaking ability	70.9	25.9	3.2	.000
19. The searching technique was easy to learn to use as a reference when I practice speaking	74.2	22.6	3.2	.000
20. I understand the purpose of using the corpus in this treatment	90.3	3.2	6.5	.000
21. When I need to get prepared for a spoken performance, I search for help in the corpus	80	16.7	3.3	.000
22. When I search for information in the corpus, I usually get the information that I need	80.6	9.7	9.7	.000
23. I use the corpus when practicing for other courses	41.9	45.2	12.9	.000
24. As I have learned more about the corpus, I have come to like them more	80.7	9.6	9.7	.000
25. I will use the corpus for my English speaking skill in the future	70.8	16.1	12.9	.000
26. Learning about the corpus has increased my confidence about speaking in English	77.4	12.9	9.7	.000
27. If I had used the corpus earlier, I would have had a better performance on speaking	87.1	9.7	3.2	.000
28. Overall, the corpus is a very useful resource for my English speaking	80.7	16.1	3.2	.000
29. The corpus should be introduced in all EFL departments	90.4	6.4	3.2	.000
30. I will recommend the corpus to other students at Fırat University or elsewhere.	80.6	9.7	9.7	.000

When the percentages of learner responses to the general use of corpus in language learning and especially for speaking skill were taken into consideration, it is observed that the participants reacted highly positively to the use of concordance lines in the classroom. Interestingly, for the item 23, it looks like students are not decisive about whether they would use corpus for their other courses. Nearly, while 41,9% of the students reported that they would consult corpus for their other courses, 45,2 of them stated they wouldn't use the corpus for practice in other courses and nearly 13% of them stated no opinion. However, broadly speaking, it can be argued that the participants adopted quite positive attitudes towards using corpora for language learning purposes.

Moreover, in item number 20, nearly 90% of them stated that they understood why the corpus was exploited in this research. It is vital for the language learners to understand the purpose and meaning of the tasks they need to complete. McCarthy (2004) highlights the usefulness of corpus-informed materials as in the following words:

> Successful learning is all about motivation. Corpus-informed materials motivate because teachers and learners can be sure that the language they are practicing is modern, used in everyday situations, targeted to situations they are likely to find themselves in, and corresponds to what they will hear and see in real conversations, movies, radio and TV shows, newspapers, books, Internet texts, and magazines. It is not artificial or invented language, but consists of the most widely used words, phrases, and grammar. (p. 17)

Therefore, it can be maintained that especially for speaking skill, which is an area in language that the learners have all types of difficulty, providing students with corpus-informed materials can help them become more efficient in oral communication.

All in all, by examining the findings from the intervention and the questionnaire, it is observed that students have benefited from corpus-based learning activities, and also hold pretty positive attitudes towards using it in the classroom. However, it should be noted that the language learners need assistance in learning how to exploit corpus not only when learning epistemic markers, but also for other forms of language.

4.4 Chapter summary

This chapter describes the findings obtained through the contrastive corpus analysis and corpus informed intervention in order to detect the effect of teaching spoken epistemic markers to Turkish L2 learners through concordance lines. To this respect, the first section first presents the findings of the frequency analysis of epistemic devices in LINDSEI-TR, then the findings of frequency analysis of epistemic devices in LOCNEC, lastly, reports on the findings of the contrastive analysis of epistemic stance markers in two corpora in terms of overuse/underuse. The second section firstly contains the overall findings on the effect of corpus-informed intervention to teach epistemic markers in spoken language. Then it reports on the effect of explicit teaching methodology adopted on the frequency and range of epistemic markers in the speech of Turkish L2 learners. Next, it reports the findings from the individual epistemic stance

markers before and after the intervention. After that, it explains the findings on the learners' use of epistemic stance markers in different task types. Finally, it presents learners' attitude towards the use of concordance lines in learning epistemic stance markers in spoken language.

CHAPTER V

CONCLUSION

5.1 Introduction

This chapter summarizes the conclusions drawn from the present research. The chapter initially provides a general overview and findings of Study 1 (corpus based analysis) and Study 2 (corpus informed intervention) concisely. Pedagogical implications and limitations of the study are presented. Finally, suggestions for further research are discussed.

5.2 Overview of the Current Study

Adopting the principles of CIA, the present study has investigated the use of epistemic stance markers in spoken language, which are regarded as one of the fundamental and central part of naturally occurring interaction (Brezina, 2009; Kärkkäinen, 2003). It is also suggested by Kärkkäinen (2006) that epistemic stance is much more expressed in spoken language than attitudes, evaluations, expression of personal feelings and emotions. However, in spite of the fact that epistemic stance explorations in L1 have revealed considerable information about stance, it was reported that there is still a great need to explore how L2 learners exploit the features of epistemic stance especially in spoken language. To this end, the first main objective of the study was to uncover and describe the epistemic stance markers in Turkish EFL learners' speech and then compare and contrast them with the native speakers' use of epistemic stance markers quantitatively along with presentation of brief discussions on the functional properties of individual markers. Finally, the data drawn from the two corpora was examined in terms of overused/underused occurrences of epistemic devices. Therefore, two corpora were analysed for Study 1: LINDSEI-TR and LOCNEC.

Additionally, it was suggested in the relevant literature that research on the pedagogical applications of epistemic stance markers have been neglected and limited to only a number of studies (Tenuta, Oliveira, Orfano, 2014). In this respect, the second main objective of the study was to measure the effect of a corpus-informed explicit teaching methodology in teaching epistemic stance markers in spoken language through

the use of concordance lines. Exploiting the spoken section of BNC corpus, the data for Study 2 was collected from 39 Turkish EFL learners at three different time points (pre-test, post-test, delayed post-test). The data was analysed for 1) the overall effect of the intervention, then for 2) its effect on the frequency and range of the forms addressed in the intervention, 3) individual epistemic markers, 4) epistemic marker use in different task types. Finally, the learner attitudes towards the use of concordance lines in learning epistemic markers in oral communication were measured through a Post-Instruction Perception Questionnaire. The next section presents the general findings obtained from Study 1 and Study 2 respectively.

5.3 General Findings: Study 1 Corpus-Based Contrastive Analysis

Analyses of grammatical classes representing the epistemic stance markers showed that, epistemic phrases and epistemic adverbials are the most frequently marked forms in both native and non-native corpora. It was found that modal verbs are the next frequent grammatical class representing epistemic modality and the least frequent epistemic markers were adjectives. This finding is in line with Kärkkäinen (2003) who claims that as epistemic phrases and adverbs include comparatively more flexible, explicit and unambiguous devices, they are more commonly found in spoken language.

The raw frequency analyses on LINDSEI-TR indicated that the speech of Turkish EFL learners is dominated by the use of the markers *I think* and *maybe*. Regarding *I think,* it was found that the marker was mostly used in the medial and final position, which signals uncertainty (Aijmer, 2001). Therefore, it can be claimed that L2 learners tend to express more uncertainty in speaking. In terms of epistemic phrases, while *I think* and *I don't know* were commonly employed in LINDSEI-TR, other epistemic phrases like *I know*, *I don't think*, *looks like/to me*, *I'm (not) sure, I thought* was not found to be very frequent in learner corpus. As for *I don't know*, it was reported in the literature that the German learners mostly use this phrase when they have difficulty in expressing themselves in the target language (Götz & Schilk, 2011). However, interestingly, Turkish learners prefer the word combination *how can I say* when they have trouble in retrieving a word from their memory and they seem to employ *I don't know* to emphasize either that they are not taking the full responsibility of their propositions or that they lack sufficient knowledge on the propositional content (Aijmer, 2004).

Regarding the epistemic adverbials, although the learners used *maybe, of course*, *actually* and *in fact,* other epistemic adverbs like *probably*, *definitely*, *perhaps*, *apparently*, and *certainly* were found to be deployed with very low frequencies. However, regarding the epistemic adverbials *in fact* and *probably,* the study yielded interesting findings. As for *in fact,* which is reported as more common to formal speech, it was found that it is a frequent stance adverbial in learner speech, whereas it was not found to be very common in native speech. Contrarily, although *probably* is a common epistemic stance adverbial in native data, it was quite infrequent in learner data. This points that L2 learners are not aware of the features peculiar to spoken interaction. In addition, the study showed that Turkish EFL learners tend to misuse *of course* together with *yes/yeah* as a response to a thought conveyed by the interlocutor, which may cause learners to sound over-emphatic and not polite (De Cock, 2004). Lastly, in terms of modals, Turkish learners prefer to use modal verb *would* more than the other modals to express their degree of commitment to the truth value. It was found that modals in general were much less frequently marked in learner speech by their epistemic meanings. This finding points that the learners are not aware of the epistemic uses of the modal verbs. Overall, Turkish EFL learners seem to use a highly restricted and routinized set of epistemic stance markers in oral communication.

The analyses of native corpus reflected a very different pattern when compared to the learner language. Although *I think* was the most frequent epistemic device in native corpus, it was ascertained that the following most frequent stance markers were quite different than the learner data. As for *I think,* it was revealed that the native speakers tend to employ the marker more in tentative function as a hedge. *Actually* as the second most frequent marker, was found to be employed as an epistemic marker emphasizing the truth value of clause rather being a discourse particle in native language (Aijmer,2002). Additionally, the native speakers seem to make use of other epistemic stance adverbials such as *probably*, *maybe*, *definitely*, *perhaps*, *of course*, *certainly*, *possibly*, *apparently*, *in fact,* and *sure* very commonly in informal speaking, which is in agreement with the findings of Biber et al., (1999). It was put forward in the related literature that the stance adverbials are considerably used in spoken language as the spoken language requires on-line planning, syntactic mobility and freedom of occurrence (Kärkkäinen, 2003). Analysis of other epistemic phrases in native corpus showed that although *I think* and *I don't know* are very common epistemic markers in

native spoken language, other combinations of the lexical verbs *think* and *know* are commonly employed (*I thought, I don't think, I know*). In addition, it was found out that the native learners use epistemic markers like *s/he said, I said, looks like/to me, I guess, I remember(ed*). This pattern according to the decreasing frequency counts of the epistemic phrases reflects a highly different pattern from Kärkkäinen (2003). This points that the different genres of the corpora result in different patterns of epistemic markers used since Kärkkäinen (2003) drew her data from face-to-face conversations in American English and the data from LOCNEC contained informal interviews. Finally, regarding the epistemic modals, *would* was the most frequent modal in native corpus. This is in line with Biber et al., (1999). Following that, *might, could, must,* and *may* was found to be frequent modals of epistemic (un)certainty in speech. However, *will,* and *should* was found to be not very frequently marked in epistemic meaning, which is in agreement with Kärkkäinen (2003).

Contrastive analysis of LINDSEI-TR and LOCNEC showed that although the native speakers make use of a routinized and limited pattern in expressing epistemic (un)certainty, it was seen that natives employ a comparatively bigger number of devices in conveying their epistemic stance in speech. The contrastive analyses also showed that although there are similarities between both corpora, there were also differences worth considering in terms of overuse/underuse. The epistemic markers reflecting similarities were: epistemic phrases (*I said, s/he said, I guess, I know + comp, I remember(ed)*), modals (*will, may, could, must, should*), *sure* as an epistemic adverbial, and *true* as an epistemic adjective. Among these epistemic phrases, *I guess* was surprisingly found to be very slightly overused by Turkish learners, which was not significant, whereas Aas (2011) notes that it was significantly overused in Swedish and Norwegian subcomponents of LINDSEI. The infrequency of *I guess* was attributed the effects of input and materials that the learners are exposed to as this marker was found more common in American English.

Along with the similarities, some significant differences were also found between two corpora. *I think* was found to be underused by Turkish learners, which is surprising as it was reported as overused by L2 learners (Baumgarten & House, 2010; Aas, 2011; Huang, 2014; Aijmer, 2004). Although *I think* was reported to have several functions such as starting point function together with topic shifts and expression of

increased commitment, on-line planning function, and turn completion function, Turkish learners were found to use the marker to express uncertainty or hesitation rather than employing it for face-saving and politeness purposes as Aijmer (2004) notes. This points the need to increase the learners' awareness about the different functions of *I think*. The same issue was observed in the use of *I don't know*. While L1 learners seem to use it with a wider range of functions like expressing subjective perspective or avoidance of taking responsibility, opening-up the conversational floor to the other participants (Baumgarten & House, 2010), L2 learners seem to use the marker mainly to express their lack of information on the propositional content. Underuse of the marker by Turkish learners could reflect that they need to be informed about the different functions. Considering that other combinations of common lexical verbs such as *I know, I thought, I don't think* was underused by Turkish learners, it is obvious that the Turkish learners are not aware of other possible ways of expressing their epistemic (un)certainty. With regard to epistemic adverbials, except for *maybe, of course, in fact,* Turkish learners were found to underuse all the other epistemic stance adverbials which are common in native spoken language. In terms of modal verbs, Turkish learners were found to underuse *would* and *might* which reflected significant differences between two corpora. It shows that the Turkish learners need much more instructional input on their epistemic meanings, which was reported to acquire for even advanced learners.

The differences detected between two corpora and the underuse of the most epistemic forms by learners suggests that Turkish learners are not familiar with the features of spoken language which could be one of the factors that keep them away from foreign-soundingness. It was discussed that transfers from L1 and effects of instructional input could be possible factors of the difference.

5.4. General Conclusions: Study 2 Corpus-Informed Intervention

Analyses on the overall learning effects of corpus-informed intervention to teach epistemic stance markers in spoken language indicated that the learners made a significant progress after the intervention. In other words, the learners' overall knowledge of epistemic stance markers was promoted after the explicit teaching, which is in line with findings of Fordyce (2014) who found that a sharp increase in the range of epistemic markers employed in writing was seen after the explicit intervention. Taking into consideration that the participants of the study was composed of learners at

A2 and B1 proficiency levels of English, ANCOVA statistics was run to determine if the proficiency level had a significant effect on the post-test scores of the participants once the pre-test results were controlled. The results from ANCOVA indicated that there were no significant differences between the lower level group and upper level group. Therefore, it was maintained that the both groups of proficiency have benefited from the corpus-informed intervention. This finding was in line with Fordyce (2013) and in contrast with some other studies, which found that epistemic marker use varies with different proficiency levels (Oh & Kang, 2013; Hu & Li, 2015). Through the comparisons of post-test and delayed post-test scores, it was observed that the learners regressed 4 weeks after the intervention, which was an expected finding since the spoken language is harder to control than the written language. Analysis of the residual effects showed that while the means in the delayed post-test were slightly higher than the pre-test means, no significant difference was found. It was suggested that when the learners do not visit the target forms regularly, they tend to forget the items they have learnt to exploit in spoken interaction. The analyses according to the frequency of the target forms displayed similar results that the learners showed an increased performance on the post-test, they regressed in the delayed test, and there was no significant difference between the pre-test and delayed post-test results. It was also reported by Jones and Carter (2014) that the explicit teaching methodology was more effective in the short term. However, that changed when the data was analysed for the range of epistemic stance markers used at three time points. The learners performed a significant progress in the range of epistemic forms they used in post-test. It was again observed that they regressed in the delayed-test. However, the analysis of residual effects of the corpus-informed intervention pointed that although the learners regressed in the delayed post-test, they were still better than their pre-test epistemic range scores 4 weeks after the intervention. Therefore, it was suggested that the learners retained their knowledge of epistemic stance markers in terms of the range of epistemic markers they deployed in spoken language.

When the individual epistemic stance markers employed in pre-test and post-test were analysed, it was seen that while the learners used only 10 epistemic stance markers in the pre-test with a total number of 269 epistemic forms, they made use of all the epistemic devices addressed in the intervention in the post-test with 321 epistemic forms in total. Both the frequency and range of the epistemic markers was increased after the

corpus-informed treatment. Moreover, it was observed that the learners used some other epistemic stance markers such as *I guess,* and *it is possible,* which were not included in the intervention sessions. This points out that when the learners are explicitly exposed to the target forms, they are able to exploit them in their oral production. As a final remark for this section, individual epistemic stance markers investigated showed that in the present study with 39 participants, which is a relatively smaller sample when compared to LINDSEI-TR, it was determined that *I think* and *maybe* dominated the oral production of the participants. This finding reveals the vital role of corpora in investigating the learner language. Even in a small sample of learners, it was claimed that the learner corpora have great potential in predicting some patterns about L2 learners' speech.

In terms of the epistemic markers used in three different task types, the learners seemed to employ more epistemic stance markers in Task 2 in terms of both frequency and epistemic range. It was suggested that the learners were able to express their epistemic commitment in the second task, which required the students to make predictions on the given pictures, and was similar to picture description. In Task 1, which required the students to provide answers to an open-ended question, the learners made use of relatively small number of epistemic stance markers. However, while *of course* was misused by 17 out of 39 students in the pre-test, they were able to correct this error replacing their answers with other epistemic devices like *certainly*, and *I'm sure*. This suggested that a corpus-informed teaching methodology was not only effective in increasing the frequency and range of the stance markers used in spoken language, it was also effective in urging the students to correct their errors. In Task 3, the frequency and range of the epistemic forms were comparatively less. The possible reason discussed was the level of control in the task.

Finally, the questionnaire results for the first dimension showed that the learners had quite positive attitudes towards the use of concordance lines in learning the meaning, usage and function of the epistemic stance markers and they reported that the corpus was useful to improve their understanding of the (un)certainty in spoken English and to learn epistemic stance markers in English. The second dimension concerned the difficulties the learners had in using concordance lines. The learner responses reflected somehow conflicting opinions. While approximately one half of the students reported

that they had difficulties when using the corpus and analysing the concordance output, the other half stated they didn't have any trouble. Therefore, it was suggested that the learners were indecisive in whether the corpus was user-friendly or not. It was discussed that learners' unfamiliarity with the corpora may have challenged them to conduct searches in the corpus and analyse the output. The last dimension of the questionnaire was related to general use of BNC. The learners reacted positively to the items in this domain. They stated that they would consult corpus to practice speaking in the future, recommend it to their peers, and they would have had a better performance on speaking if they had known about corpora before. Surprisingly, the learners stated conflicting responses to the item questioning if they would consult corpus in order to practice for their other courses. All in all, it was found that the learners' attitude towards using concordance lines to learn epistemic markers in speech. Therefore, it can be claimed that the use of corpus-informed materials in the classroom may be a motivating resource for the learners as what they came across in corpus is not an invented language, but rather modern and widely used in real life (McCarhty, 2004).

5.5 Pedagogical Implications and Suggestions for ELT

The present study has revealed two main facts: 1) a learner corpus-based research has the potential to provide a better description to specific patterns in learner language and to provide useful information both for EFL learners and teachers; 2) the explicit teaching of epistemic stance markers through a corpus-informed approach could be effective to make EFL learners more proficient in expressing their epistemic commitment in speech.

Within the frameworks of CIA, the present study has first investigated the use epistemic stance markers in Turkish EFL learners' spoken production compared to the native spoken language, then explored the effectiveness of a corpus-informed teaching methodology in learning epistemic stance markers in speech. The findings of the current study have a number of significant implications for foreign or second language teaching. First of all, the study shows that epistemic stance in L1 and L2 differs and L2 learners use a very restricted number of epistemic stance markers in speaking. It suggests that the epistemic devices should be explicitly explained and taught to the learners in companion with their varying functions as the EFL learners seem to be unfamiliar with the linguistic properties of spoken language. Thus, from a pedagogical

point of view, it would be useful to consult corpus to uncover the deviances from the native language in order to enable learners to be able to express an appropriate degree of (un)certainty. As a result, language learners could be aware of the pragmatic aspects of the language with low salience to avoid miscommunications in the target language.

Second, the study has provided evidence that a corpus-informed approach had a significant impact on learning epistemic stance markers by lower level EFL learners. Considering that the participants of the study consisted of learners at A2 and B1 level, it was maintained that the lower level learners can effectively employ the epistemic stance markers in speech, which provides a contrast to the idea that the corpus-based instructional sources are most useful for learners of English at advanced level who received a lot of training (Boulton, 2009). The present study has shown that even lower level EFL learners were able to understand and appropriately employ the stance markers, which are harder to understand the epistemic aspects, such as the modal verbs *would* and *might.* Thus, it can be suggested that the corpus-informed materials can improve the learners in order to convey their epistemic commitment in speech, which can lead to a more native-like and fluent speech.

Third, it was maintained in the study that the explicit teaching of epistemic stance markers in speech was effective in improving the students' understanding and usage of the target forms in the short term. However, it is a fact that the loss of gains between the post-test and delayed post-test showed that the learners fail to move some of the target items to their long term memory. This points out that the gains can be improved and maintained if the explicit treatment is followed by richer amount of input (Ellis, 2008). Thus, the target forms need to be regularly revisited in the classroom to be fitted into the learners' oral production.

Furthermore, the present study has shown that the EFL learners have benefited from the corpus-informed teaching sessions. This, in turn, points that material developers and language teachers can make use of the information drawn from the learner corpora to provide more meaningful input which is enriched by the use of epistemic stance marker use in context. Both curriculum designers and language teachers should consider integrating findings obtained through corpora to be able to present more informed materials and activities to learners, particularly on speaking skill, which is a common trouble spot for the learners. Additionally, the use of corpus-based

materials in the classroom can provide the students with a more interesting atmosphere going beyond the routines of the traditional course book activities in instructional settings.

Last but not the least, use of a spoken corpus to improve the students' ability to express their level of (un)certainty in speaking can motivate L2 learners to build confidence in speaking in the target language since through corpus-informed materials, students will observe that the native speakers also hesitate, form shorter sentences, and may use a narrow range of vocabulary when speaking (Walsh, 2010). Taking into consideration the fact that the students hold a very positive attitude towards the use of corpus-based activities in the classroom, this suggests that this type of activities can motivate the students to confidently participate in speaking activities in the classroom.

5.6 Limitations of the Study and Suggestions for Further Research

In spite of the promising findings obtained from the current study, there are several limitations that needs consideration. Firstly, the study remained limited to only 39 participants at A2 and B1 level of proficiency in English. Regarding the spoken data, it is possible that the learners know the items but not use them in the tasks. Thus, further research needs to be conducted with a larger number of learners and with learners from a wider range of proficiency levels from beginner level to advanced level to be able to increase the generalizability of the findings and elicit more target forms in learner speech. Also, it is necessary to investigate the use of epistemic markers in the oral production of learners from different backgrounds in order to check whether the findings obtained from the present study apply to those students coming from different backgrounds.

Second, due to time limitations and fatigue factor of the students, it was not possible to elicit oral production from each student in the classroom during the intervention sessions. Providing a chance to perform their dialogue in the classroom to each student could have created a better impact on the scores of the students in the post-test and delayed post-test. Researchers may take that into account for future research.

Third, some technical problems were experienced during the intervention sessions. Due to limit constraints of BNC corpus after a certain number of searches, there was a need for a readjustment of the teaching sessions as some of the markers

could not be studied as planned in the schedule. Therefore, it is necessary that the issues related to technical and technological aspects need to be considered in more detail before the intervention.

Fourth, the study adopted a one group pre-test post-test design since it was not possible to find an equal control group in terms of proficiency level and number of the students in the research setting. It would provide better insights if the future studies conduct a study using a control group to detect the effects of the corpus-informed teaching better in comparison to the traditional classroom activities.

Finally, data from the delayed post-test was taken 4 weeks after the intervention due to time constraints. It would be useful to investigate whether the learners retain their knowledge of epistemic stance markers after a longer period of time or not, which could present more precise findings of learner development.

REFERENCES

Aarts, B. (1992). Small Clauses in English: The Nonverbal Types. Berlin and New York: Mouton de Gruyter.

Aarts, J. (2002) Does Corpus Linguistics exist? Some old and new issues. In L. Brevik and A. Hasselgren (eds.) From the COLT's mouth…and others. Amsterdam: Rodopi. 1–17.

Aas, H. L. (2011). Recurrent word-combinations in spoken learner English: A study of corpus data from Swedish and Norwegian advanced learners. (Unpublished Doctoral Dissertation) University of Oslo, Sweden.

Adams, H., & Quintana-toledo, E. (2013). Adverbial stance marking in the introduction and conclusion sections of legal research articles. Revista de Lingüística Y Lenguas Aplicadas, 8(1), 13–22.

Ağçam, R. (2014). A Corpus-Based Study on Epistemic Adjectives in Academic English. American Journal of Educational Research, 2(12), 1230–1236.

Aijmer, K. (1997) 'I think – an English modal particle.' In: T. Swan, O. J. Westvik, (Eds), *Modality in Germanic Languages. Historical and Comparative Perspectives*. Berlin: Mouton de Gruyter. 1-47.

Aijmer, K. (2001). I think as a marker of discourse style in argumentative Swedish student writing. In K. Aijmer (Ed.), A Wealth of English.Studies in Honour of Go¨ran Kjellmer. (pp. 247–258) Acta Universitatis Gothoburgensis, Go¨teborg,

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

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. Nordic Journal of English Studies, 3(1), 173–190.

Aijmer, K. (2009). "So er I just sort I dunno I think it's just because…": A corpus study of I don't know and dunno in learners' spoken English. In A. H. Jucker, D. Schreier & M. Hundt (Eds.), Corpora:Pragmatics and discourse (pp. 151–168). The Netherlands: Rodopi.

Almeida, S. & Ferrari, L. (2012). Subjectivity, Intersubjectivity and Epistemic Complementation Constructions. Selected Papers from UK-CLA Meetings, 1, 110–127. Retrieved from http://www.ukcla.org.uk/proceedings/volume_1%5Cnpapers3://publication/uuid/9D1D884D-D067-4C24-B416-AFD608A61478

Alonso Almeida, F., López Brito, B., & Quintana Toledo, E. (2014). Teaching and learning stance in specialized English for computing. A corpus linguistics approach. Actas de las I Jornadas Iberoamericanas de Innovación Educativa en el Ámbito de las TICs, 27 – 28.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Aşık, A., Vural, A. Ş., & Akpınar, K. D. (2015). Lexical Awareness and Development through Data Driven Learning: Attitudes and Beliefs of EFL Learners. Journal of Education and Training Studies, 4(3), 87–96.

Aull, L. L., & Lancaster, Z. (2014). Linguistic Markers of Stance in Early and Advanced Academic Writing: A Corpus-based Comparison. Written Communication (Vol. 31).

Baumgarten, N., & House, J. (2010). I think and I don't know in English as lingua franca and native English discourse. Journal of Pragmatics, 42(5), 1184–1200.

Bennet, G. (2010). Using corpora in the language learning classroom: Corpus linguistics for teachers. Michigan: University of Michigan Press.

Berman, R., Ragnarsdóttir, H., & Strömqvist, S. (2002). Discourse stance: Written and spoken language. Written Language & Literacy, 5(2), 253–287.

Biber, D. (2006). Stance in spoken and written university registers. Journal of English for Academic Purposes, 5, 97-116.

Biber, D. (2010). Corpus-based and corpus-driven analyses of language variation and use. In B. Heine & H. Narrog (Eds.), The Oxford handbook of linguistic analysis (pp. 193–223). Oxford: Oxford University Press.

Biber, D. & Finegan, E. (1989) Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. Text 9 (1), 93-124.

Biber, D., & Finegan, E. (1988). Adverbial Stance Types in English. Discourse Processes, 11(1), 1–34.

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). *The Longman grammar of spoken and written English*. London: Longman.

Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL*, 21(1), 37-54.

Brezina V. (2012). *Epistemic markers in university advisory sessions: Towards a local grammar of epistemicity*. (Unpublished PhD dissertation). University of Auckland, Auckland, New Zealand.

Brezina, V. (2009). "We only say we are certain when we are not": a corpus-based study of epistemic stance. In *The Proceedings of the Fifth International Conference SLOVKO 2009.* Bratislava: Tribun.

Brezina, V. (2013). Certainty and uncertainty in spoken language: In search of epistemic socio- lect and idiolect' in M. Reif-Hu¨ lser, J.A. Robinson and M. Pu¨ tz (eds): Variation in Language and Language Use: Linguistic, Socio- Cultural and Cognitive Perspectives. Peter Lang, 97–130.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies, 7*(4–5), 585–614.

Bybee, J. (2007). Frequency of use and the organization of language. Oxford: Oxford University Press.

Campoy, M., Bellés-Fortuno, B., & Gea-Valor, M. (2010). Corpus-Based Approach to English Language Teaching. *Journal of Chemical Information and Modeling* (Vol. 53).

Can, C. (2009). İkinci dil edinimi çalışmalarında bilgisayar destekli bir Türk öğrenici İngilizcesi derlemi: ICLE'nin bir altderlemi olarak TICLE. *Dil Dergisi*, 144 (Nisan-Mayıs-Haziran), 16–34.

Can, C. (2012). Uluslararası Türk öğrenici İngilizcesi derleminde tutum belirteçleri. *Dilbilim Araştırmaları*, 1, 39–53.

Chafe, W. L. (1986). Evidentiality in English conversation and academic writing. In W. L. Chafe, & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 261–272). Norwood, NJ: Ablex.

Chandrasegaran, A., & Kong, K. M. C. (2007). Stance-taking and stance-support in students' online forum discussion. *Linguistics and Education*, *17*(4), 374–390.

Chang, P., & Schleppegrell, M. (2011). Taking an effective authorial stance in academic writing: Making the linguistic resources explicit for L2 writers in the social sciences. *Journal of English for Academic Purposes*, *10*(3), 140–151.

Chen, H. (2010). Contrastive learner corpus analysis of epistemic modality and interlanguage pragmatic competence in L2 writing. *Arizona Working Papers in SLA &Teaching*, *51*, 27–51.

Chen, Z. (2012). Expression of Epistemic Stance in EFL Chinese University Students' Writing. *English Language Teaching*, *5*(10), 173–179.

Cheong, E. (2012). Stance Features in ESL Students' Oral Presentations. Retrieved, October, 2016, from https://nau.edu/uploadedFiles/Academic/CAL/PIE/2012-2013_Research/Stance_features_ESL_students'_oral_presentations.pdf.

Chomsky, N. (1962), Paper given at Third Texas Conference on Problems of Linguistic Analysis in English, 1958. Austin: University of Texas.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge MA: MIT Press.

Cohen, L., Manion, L., & Morrison K. (2000). Research Methods in Education (5th Edition). London: Routledge Falmer.

Conrad, S. and Biber, D. (1999). "Adverbial marking of stance in speech and writing", in S. Hunston and G. Thompson (eds.) Evaluation in text. Authorial stance and the construction of discourse. New York: Oxford University Press, 56-73.

Cook, T. D. (2015). Quasi-Experimental Design. *Wiley Encyclopedia of Management*, 11, 1–2.

Crawford Camiciottoli, B. (2010). A corpus-informed approach to teaching lecture comprehension skills in English for Business Studies. In M.C. Campoy-Cubillo, B. Bellés Fortuño & L. Gea-Valor (Eds.), *Corpus-Based Approaches to ELT* (pp. 95-106). London and New York: Continuum.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literatures (BELL), New Series* (2), 225– 246.

De Cock, S. (2010). Spoken learner corpora and EFL teaching. In M. C. Campoy, B. Bellés-Fortuño and M. L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching* (pp. 123-137). London: Continuum.

De Cock, S. (2011). Preferred patterns of use of positive and negative evaluative adjectives in native and learner speech: an ELT perspective. In A. Frankenberg Garcia, L. Flowerdew & G. Aston (Eds.), New Trends in Corpora and Language Learning, (pp. 198-212). London & New York: Continuum,

Du Bois, J.W. (2007) 'The Stance Triangle', in R. Englebretson (ed.) *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction,* pp. 139–82. Philadelphia, PA: John Benjamins.

Efstathiadi, L. (2010). The use of epistemic markers as a means of hedging and boosting in the discourse of L1 and L2 speakers of Modern Greek: A corpus-based study in informal letter-writing. *Themes in Science and Technology Education: Special Issue on ICT in Language Learning. Computer-Aided Language Analysis, Teaching and Learning: Approaches, Perspectives and Applications*, *3*(1-2), 181–206.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford: Oxford University Press.

Englebretson, R. (ed.) (2007). *Stancetaking in discourse*. Amsterdam: Benjamins.

Fitzmaurice, S. (2004). Subjectivity, intersubjectivity and the historical construction of interlocutor stance: From stance markers to discourse markers. *Discourse Studies*, *6*(4), 427–448.

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

Fordyce, K. (2009). A comparative study of learner corpora of spoken and written discursive language: Focusing on the use of epistemic forms by Japanese EFL learners. *Hiroshima Studies in Language and Language Education,12,* 135–50.

Fordyce, K. (2014). The differential effects of explicit and implicit instruction on EFL learners' Use of Epistemic Stance. *Applied Linguistics*, *35*(1), 6–28.

Gablasova, D. & Brezina V. (2015). Does speaker role affect the choice of epistemic adverbials in L2 speech? Evidence from the Trinity Lancaster Corpus. In J. Romero-Trillo (Ed), *Yearbook of Corpus Linguistics and Pragmatics 2015* (pp. 117–36). Springer.

Gablasova, D., Brezina, V., Mcenery, T., & Boyd, E. (2015). Epistemic Stance in Spoken L2 English: The Effect of Task and Speaker Style. *Applied Linguistics*, 1 – 26.

Gass, S. M., & Selinker, L. (2001). *Second language acquisition: An introductory course*. Mahwah, N.J: L. Erlbaum Associates.

Gibbs, D. (1990). Second language acquisition of the English modal auxiliaries can, could, may, and might. *Applied Linguistics 11*(3), 297-314.

Gilquin, G. (2012). LINDSEI. Retrieved on August, 2016 from www.uclovain.be/en-cecl-lindsei.html

Gilquin, G., De Cock, S., & Granger, S. (Eds.). (2010). *LINDSEI Louvain international database of spoken English interchange. Handbook and CD-ROM*. Louvain-la-Neuve: Presses universitaires de Louvain.

Girgin, U. (2011). *Corpus-Based Activities at Lower Levels of EFL Proficiency: The Effectiveness of Using Concordance Lines on Grammar Learning*. (Unpublished masters' thesis). Bilkent University, Ankara, Turkey.

Götz, S., & Schilk, M. (2011). Formulaic sequences in spoken ENL, ESL, and EFL. In M. Hundt & J. Mukherjee (Eds.), *Exploring second-language varieties of English and learner Englishes: Bridging a paradigm gap* (pp. 79-100). Amsterdam: John Benjamins.

Granger S. (1993). The International Corpus of Learner English, in: J. Aarts, P. De Haan & N. Oostdıjk (eds.) English Language Corpora: Design, Analysis and Exploitation. Rodopi. Amsterdam & Atlanta: 57-71.

Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In Sylviane Granger (Ed.), *Learner English on computer* (pp. 1–19). U.S.A.: Addison Wesley Longman Limited.

Granger, S. (2002). A bird's eye view of learner corpus research. In Sylviane Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–36). Amsterdam & Philadelphia: John Benjamins.

Granger, S. (2003). A Promising Synergy. *CALICO Journal*, *20*(3), 465–480. Retrieved from https://calico.org/memberBrowse.php?action=article&id=289

Granger, S. (2003). The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, *37*(3), 538–546.

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13–32). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Granger, S. (2012). How to use foreign and second language learner corpora. In A. Mackey & S. Gass (eds.) *Research Methods in Second Language Acquisition: A Practical Guide.* Malden: Blackwell, 7-29.

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. (2004). Computer learner corpus research : Current status and future prospects. In U. Connor & T. A. Uptom (Eds.), *Applied Corpus Linguistics. A multidimensional perspective* (pp. 123–146). Amsterdam: Rodopi.

Gries, S. T. (2009). What is Corpus Linguistics? *Language and Linguistics Compass*, *3*(5), 1225–1241.

Gries, S. T. (2010). Useful statistics for Corpus Linguistics. In A. Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics: selected approaches* (pp. 269– 291). Frankfurt: Peter Lang.

Gries, S.T. (2010), Corpus linguistics and theoretical linguistics: A love-hate relationship? Not necessarily.., *International Journal of Corpus Linguistics*, (*15*), 327-343.

Haddington, P. (2013) Pragmatics of stance. In Chapelle, C.A. (ed.) *The Encyclopedia of Applied Linguistics*. Oxford, UK: Wiley-Blackwell.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar* (2nd ed.). London: Edward Arnold.

Hasko, V. (2013). Qualitative Corpus Analysis. *The Encyclopaedia of Applied Linguistics*, *13*, 1–10.

Heritage, J. (2012). The Epistemic Engine: Sequence Organization and Territories of Knowledge. *Research on Language & Social Interaction*, *45*(1), 30–52.

Hita, J. A. (2008). Teaching modality in context: A sample lesson. In N. Norgaard (Ed.). *Systemic functional linguistics in use: Odense working papers in language and communication* Vol. 29 (pp. 365-380). Odense, Denmark: University of Southern Denmark.

Hu, C., & Li, X. (2015). Epistemic modality in the argumentative essays of Chinese EFL learners. *English Language Teaching*, *8*(6), 20–31.

Huang, L. F. (2014). "I think" in NS and Chinese NNS spoken English. *International Journal of Computer-Assisted Language Learning and Teaching*, *4*(1), 84–100.

Hunston, S. (2006). Corpus Linguistics. *Encyclopedia of Language and Linguistics*, (2001), 234–248.

Hunston, S. (2007). 'Using a corpus to investigate stance quantitatively and qualitatively'. in Englebretson (ed.) *Stancetaking in Discourse*. Benjamins. 27-48.

Hunston, S., & Thompson, G. (Eds.). (2000). *Evaluation in text: Authorial stance and the construction of discourse*. New York, NY: Oxford University Press.

Hyland, K. (1996). "Writing without conviction? Hedging in science research articles", *Applied Linguistics*, *17*(4), 433-454. doi:10.1093/applin/17.4.433

Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.

Hyland, K. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173–192.

Hyland, K. (2010). Metadiscourse: Mapping Interactions in Academic Writing. *Nordic Journal of English Studies*, *9*(2), 125–143.

IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Jaffe, A. (2009). *Stance: Sociolinguistic Perspectives*. Oxford: Oxford University Press.

Jiajin, X., & Manying, L. (2008). Comparing Stance in Chinese EFL learners ' English and Chinese argumentative essays of a shared topic, 1–39.

Johansson, S. (2008). Some aspects of the development of corpus linguistics in the 1970s and 1980s. In A. Lüdeling & M. Kytö (Eds.), Corpus linguistics: An international handbook (Vol. 1, Berlin, Germany: De Gruyter.

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

Johnstone B. (2009). Stance, style, and the linguistic individual. In A. Jaffe (Ed.), *Stance: Sociolinguistic perspectives* (pp. 29-52). New York, NY: Oxford University Press.

Jones, C., & Carter, R. (2014). Teaching spoken discourse markers explicitly: A comparison of III and PPP. *International Journal of English Studies*, *14*(1), 37–54. Retrieved September, 2016, from http://search.proquest.com/docview/1552718638/fulltextPDF/4E9AFC9F60A4436BPQ/12?accountid=62689

Kärkkäinen, E. (1992). Modality as a strategy in interaction: Epistemic modality in the language of native and nonnative speakers of English. *Pragmatics and Language Learning*, *3*, 197-216. Retrieved from http://www.eric.ed.gov/PDFS/ED395530.pdf

Kärkkäinen, E. (2003). *Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think*. Amsterdam: John Benjamins.

Kärkkäinen, E. (2006). Stance-taking in conversation: From subjectivity to intersubjectivity. *Text & Talk*, 26(6), 699–731.

Kärkkäinen, E. (2010). Position and scope of epistemic phrases in planned and unplanned american ¨ english. In: G. Kaltenböck, W. Mihatsch & S. Schneider, (Eds.), *New approaches to hedging* (pp. 203–236). Amsterdam: Elsevier.

Kärkkäinen, E. (2012). I thought it was very interesting. Conversational formats for taking a stance. *Journal of Pragmatics*, *44*(15), 2194–2210.

Keisanen, T. (2006). *Patterns of stance taking: Negative yes/no interrogatives and tag questions in American English conversation*. Oulu, Finland: Oulu University Press.

Keisanen, T. (2007). Stancetaking as an interactional activity: Challenging the prior speaker. In R. Englebretson (Ed.), Stancetaking in discourse: Subjectivity, evaluation, interaction (pp. 253–81). Amsterdam, Netherlands: John Benjamins.

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

Kiesling, S. F. (2009). Style as stance: Stance as the explanation for patterns of sociolinguistic variation. In A. M. Jaffe (Ed.), *Stance. Sociolinguistic perspectives* (pp. 171–194). Oxford: Oxford University Press.

Kilimci, A. (2003). Stance and attitude in advanced Turkish learners' written discourse. In *The 38th Linguistics Colloquium: Language and Language-Processing*. Piliscsaba, Hungary.

Kilimci, A. (2009). *Negotiation of meaning in L2 academic writing*. In G. Socarras (Ed) Proceedings of 1st International Conference on Literature, Languages and Linguistics. Athens: ATINER.

Kilimci, A. (2014). LINDSEI-TR: A new spoken corpus of advanced learners of English. *International Journal of Social Sciences and Education*, *4*(2), 401–410.

Kirkham, S. (2011). Personal style and epistemic stance in classroom discussion. *Language and Literature*, *20*(3), 201–217.

Ko, C. (2014). English language teaching : Teaching of hedges. *Journal of Education and Learning*, *8*(2), 106–114.

Krieger, D. (2003). Corpus Linguistics: What it is and how it can be applied to teaching. *The Internet TESL Journal*, 9(3). Retrieved from iteslj.org/Articles/Krieger-Corpus.html at July 3rd in 2016.

Lancaster, Z. (2014). Exploring Valued Patterns of Stance in Upper-Level Student Writing in the Disciplines. *Written Communication*, *31*(1), 27–57.

Larsen-Freeman, D. & Long, M.H. (1991). *An introduction to second language acquisition research*. New York, NY.: Longman.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Leech, G. (1991). The State of the Art in Corpus Linguistics. In: K. Aijmer, /B. Altenberg, (eds.), *English Corpus Linguistics: In honour of Jan Svartvik*. London: Longman, 8–29 .

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), Directions in corpus linguistics (pp. 105–122). Berlin: Mouton de Gruyter.

Leech, G. (1997). Teaching and Language Corpora: A Convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 1-23) Harlow: Addison Wesley Longman.

Léon, J. (2005). Claimed and Unclaimed Sources of Corpus Linguistics. *The Henry Sweet Society Bulletin*, (44), 34–48.

Letica, S. (2009). Use of Epistemic Modality by Non-Native Speakers of English, 119–134.

Li, S. (2012). The Effects of Input-Based Practice on Pragmatic Development of Requests in L2 Chinese. *Language Learning*, *62*(2), 403–438.

Martin, J.R. (2000). Beyond Exchange: Appraisal Systems in English. In S. Hunston & G. Thompson (Eds.), *Evaluation in Text: Authorial Stance and the Construction of Discourse* (pp. 142-176). Oxford: Oxford University Press.

McCarthy, M. (2004). From Corpus to Course Book. *Touchstone From Corpus to Course Book*, 1–20.

McCarthy, M. J. (1998). *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.

McCarthy, M. J., & Carter, R. (1995). Spoken grammar: What is it and how can we teach it? *ELT Journal*, 49(3), 207–218.

McCarthy, M., & O'Keeffe, A. (2010). Historical Perspective: What are corpora and how they evolved? In M. McCarthy & A. O'Keeffe (Eds.), *The Routledge handbook of Corpus Linguistics* (pp. 3–13). Abingdon, UK: Routledge.

McEnery, T. & Hardie, A. (2012) *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T., & Wilson, A. (1996). *Textbooks in Applied Linguistics*. Edinburgh: Edinburgh.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

McEnery, T., & Xiao, R. (2010). *Corpus-based contrastive studies of English and Chinese*. Routledge.

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

Meunier, F. (2002). The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung and S. Petch-Tyson (Eds), *Computer learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 119-142) Philadelphia: John Benjamins.

Meunier, F. (2011). Corpus linguistics and second / foreign language learning : exploring multiple paths Linguística de corpus e aprendizagem de segunda língua / língua estrangeira : explorando caminhos múltiplos. *RBLA, Belo Horizante*, *11*(2), 459–477.

Meyer, C. (2002), *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Morek, M. (2015). Show that you know - Explanations, interactional identities and epistemic stance-taking in family talk and peer talk. *Linguistics and Education*, *31*, 238–259.

Mukherjee, J., & Rohrbach, J.-M. (2006). Rethinking applied corpus linguistics from a language-pedagogical perspective: New departures in learner corpus research. In

B. Kettemann & G. Marko (Eds.), *Planning, gluing and painting corpora: Inside the applied corpus linguist's workshop* (pp. 205-232). Frankfurt, France: Peter Lang.

Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, *21*(4), 373-391.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 125–157). Amsterdam: John Benjamins Publishing Company.

Nesselhauf, N. (2005). Corpus linguistics: A practical introduction. Retrieved July, 2016, from http://www.as.uni-heidelberg.de/personen/Nesselhauf

Nozawa, Y. (2015). The use of hedges in polite disagreement by Japanese advanced learners of English. *Waseda Daigaku Daigakuin Kyoikugaku Kenkyuka Kiyo Betsu*, *22*(2), 175–188.

Ochs, E. (1996). Linguistic resources for socializing humanity. In J. Gumperz & S. Levinson (Eds.), *Rethinking Linguistic Relativity*, (pp. 407 – 437). New York: Cambridge University Press.

Ochs, E. (Ed.) (1989). The pragmatics of affect. Special issue of Text, Vol. 9.

Oh, S. Y., & Kang, S. J. (2013). The effect of English proficiency on Korean undergraduates' expression of epistemic modality in English argumentative writing. *Journal of Asia TEFL*, *10*(4), 97–132.

Olinger, A. R. (2011). Constructing identities through "discourse": Stance and interaction in collaborative college writing. *Linguistics and Education*, *22*(3), 273–286.

Orfano, B. M., Oliveira, A. L. A. M. & Tenuta, A. M. (2014). Epistemic modality through the use of adverbs: a corpus-based study on learners' written discourse. *Letras & Letras*, *30*(2), 104–121.

Palmer, F. R. (2001). Mood and Modality. *Summer Institute of Linguistics Inc*, (1986), 258.

Paquot, M., & Bestgen, Y. (2009). Distinctive words: A comparison of three statistical tests. In M. Hundt, D. Schreier, & A. H. Jucker (Eds.), *Corpora: Pragmatics and discourse* (pp. 243–265). Amsterdam: Rodopi.

Park, J. S., & Takanashi, H. (2011). Introduction Reframing Framing : Interaction and the Constitution of Culture and Society, 185–190.

Pravec, N. A. (2002). Survey of learner corpora. *ICAME Journal*, 26, 81–114.

Precht, K. (2003). Stance moods in spoken English : Evidentiality and affect in British and American conversation. *Text*, *23*(2), 239–257.

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

Recsky, L. (2006). Epistemic modality and spoken discourse: an EnglishPortuguese cross-linguistic investigation. *Linguagem & Ensino*, 9, 1, 159-185.

Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik, 54*(2), 121-134.

Römer, U. (2008). Corpora and language teaching. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 112–30). Berlin, Germany: De Gruyter.

Roseano, P., González, M., Borràs-Comes, J., & Prieto, P. (2015). Communicating epistemic stance: How speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Processes, 53*(3), 135-174.

Şahin Kızıl, A. (2013). *Recurrent Phrases in Learner English: A Corpus Driven Approach*. (Unpublished doctoral dissertation). Çukurova Üniversitesi Sosyal Bilimler Enstitüsü, Adana, Turkey.

Şahin Kızıl, A. & Kilimci, A. (2014). Recurrent phrases in Turkish EFL learners' spoken interlanguage: A corpus-driven structural and functional analysis. *Journal of Language and Linguistic Studies, 10*(1), 195-210.

Saville-Troike, M. (2005). Introducing second language acquisition. Cambridge: Cambridge University Press.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209–240.

Silver, M. (2003). The stance of stance: A critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes*, *2*(4), 359–374.

Simon-Vandenbergen, A.M. & Aijmer, K. (2007). Functionality of adjectival and adverbial epistemic expressions. In C. Butler, H. T. Downing, J. Lavid & A. Downing (Eds.), *Functional perspectives on grammar and discourse*. (pp. 419-445). Amsterdam: John Benjamins.

Sinclair, J. (1996). EAGLES. Preliminary recommendations on Corpus Typology. Retrieved July 12, 2016, from www.ilc.pi.it/EAGLES96/corpustyp/corpustyp.html.

Sinclair, J. (2005). "Corpus and Text - Basic Principles" in *Developing Linguistic Corpora: a Guide to Good Practice*, ed. M. Wynne. Oxford: Oxbow Books: 1-16. Available online from http://ota.ox.ac.uk/documents/creating/dlc/[Accessed 2016-06-30].

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

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

Sripicharn, P. (2004). Examining native speakers' and learners' investigation of the same concordance data and its implications for classroom concordancing with ELF learners. In G. Aston, S. Bernardini&D. Stewart (Eds.),*Corpora and language learners* (pp. 233-345). Amsterdam: John Benjamins.

Stone, L. D., & Thompson, G. A. (2014). Classroom mood and the dance of stance: The role of affective and epistemic stancetaking in the development of a classroom mood. *Learning, Culture and Social Interaction*, *3*(4), 309–322.

Stubbs, M., & Halbe, D. (2013). Corpus linguistics: Overview. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1377–1379). Oxford: Blackwell.

Takahashi, S. (2001). The role of input enhancement in developing pragmatic competence. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp.171- 199). Cambridge, UK: Cambridge University Press.

Taylor, C. (2008). What is corpus linguistics? What the data says. *ICAME Journal*, 179–200.

Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics*, *10*(1), 1–13.

Tognini-Bonelli, E. (2010). Theoretical overview of the evolution of Corpus Linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of Corpus Linguistics* (pp. 14–27). Abingdon, UK: Routledge.

Tono, Y. (2003). Learner corpora: Design , development and applications. In D. Archer, P. Rayson, A. Wilson, & A. McEnery (Eds.), Proceedings of Corpus Linguistics 2003 (pp. 800–809). England: Lancaster University.

Tribble, C. (2012, May). Teaching and language corpora: Quo vadis? Paper presented at the 10th Teaching and Language Corpora Conference, Warsaw, Poland.

University of Cambridge: *ESOL Examinations (2004). Quick Placement Test* (Version 1 & 2). Oxford University Press.

Walsh, S. (2010). What features of spoken and written corpora can be exploited in creating language teaching materials and syllabuses? In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp.333-344). Abingdon: Routledge.

Wharton, S. (2012). Epistemological and interpersonal stance in a data description task: Findings from a discipline-specific learner corpus. *English for Specific Purposes*, *31*(4), 261–270.

White, P.R.R. (2001). *Appraisal: An overview.* Retrieved from http://www.grammatics.com/appraisal at July 3rd in 2016.

Wierzbicka, A. (2006). *English: Meaning and Culture*. Oxford: Oxford University Press.

Williams, G. (2006). “La linguistique de corpus: Une affaire préposition- nelle”. In F. Rastier & M. Ballabriga (Eds.), *Corpus en Lettres et Sciences Sociales: Des Documents Numériques à l’Interprétation*. Paris: Texto, 151–158.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing, *Journal of Second Language Writing 13*, 257–283.

APPENDICES

APPENDIX A

LINDSEI TRANSCRIPTION GUIDELINES

1. Interview identification

Each interview is preceded by a code of this type: <h nt="TR" nr="FR+three-figure number"> All interviews should end with the following tag (on a separate line):</h> 2. Speaker turns

Speaker turns are displayed in vertical format, i.e. one below the other. Whilst the letter "A" enclosed between angle brackets always signifies the interviewer's turn, the letter "B" between angle brackets indicates the interviewee's (learner's) turn. The end of each turn is indicated by either </B> or </B>.

3. Overlapping speech

The tag <overlap /> (with a space between "overlap" and the slash) is used to indicate the beginning of overlapping speech. It should be indicated in both turns. The end of overlapping speech is not indicated.

4. Punctuation

No punctuation marks are used to indicate sentence or clause boundaries.

5. Empty pauses

Empty pauses are defined as a blank on the tape, i.e. no sound, or when someone is just breathing. The following three-tier system is used: one dot for a "short" pause (< 1 second), two dots for a "medium" pause (1-3 seconds) and three dots for "long" pauses (> 3 seconds).

6. Filled pauses and backchanneling

Filled pauses and backchanneling are marked as (eh) [brief], (er), (em), (erm), (mm), (uhu) and (mhm). No other fillers should be used.

7. Unclear passages

A three-tier system is used to indicate the length of unclear passages: <X> represents an unclear syllable or sound up to one word, <XX> represents two unclear words, and <XXX>represents more than two words.

If transcribers are not entirely sure of a word or word ending, they should indicate this by having the word directly followed by the symbol <?>. Unclear names of towns or titles of films for example may be indicated as <name of city> or <title of film>.

8. Anonymisation

Data should be anonymised (names of famous people like singers or actors can be kept). Transcribers can use tags like <first name of interviewee>, <first name and full name of interviewer> or <name of professor> to replace names.

9. Truncated words

Truncated words are immediately followed by an equals sign.

10. Spelling and capitalisation

British spelling conventions should be followed. Capital letters are only kept when required by spelling conventions on certain specific words (proper names, I, Mrs, etc) – not at the beginning of turns.

11. Contracted forms

All standard contracted forms are retained as they are typical features of speech.

12. Non-standard forms

Non-standard forms that appear in the dictionary are transcribed orthographically in their dictionary accepted way:cos, dunno, gonna, gotta, kinda, wanna and yeah.

13. Acronyms

If acronyms are pronounced as sequences of letters, they are transcribed as a series of upper-case letters separated by spaces.

If, on the other hand, acronyms are pronounced as words, they are transcribed as a series of upper-case letters not separated by spaces.

14. Dates and numbers

Figures have to be written out in words. This avoids the ambiguity of, for example, "1901", which could be spoken in a number of different ways.

15. Foreign words and pronunciation

Foreign words are indicated by <foreign> (before the word) and </foreign> (after the word). As a rule, foreign pronunciation is not noted, except in the case where the foreign word and the English word are identical. If in this case the word is pronounced as a foreign word, this is also marked using the <foreign> tag.

16. Phonetic features

(a) Syllable lengthening A colon is added at the end of a word to indicate that the last syllable is lengthened. It is typically used with small words like to, so or or. Colons should not be inserted within words.

(b) Articles -when pronounced as [ei], the article a is transcribed as a[ei]; -when pronounced as [i:], the article the is transcribed as the[i:].

17. Prosodic information: voice quality

If a particular stretch of text is said laughing or whispering for instance, this is marked by inserting <starts laughing> or <starts whispering> immediately before the specific stretch of speech and <stops laughing> or <stops whispering> at the end of it.

18. Nonverbal vocal sounds

Nonverbal vocal sounds are enclosed between angle brackets.

19. Contextual comments

Non-linguistic events are indicated between angle brackets only if they are deemed relevant to the interaction (if one of the participants reacts to it, for example).

20. Tasks

The three tasks making up the interview (set topic, free discussion and picture description) should be separated from each other. This is done using the following tags: <S> (before the set topic), </S> (after

the set topic), <F> (before the free discussion), </F> (after the free discussion), <P> (before the picture description), </P> (after the picture description). These tags should occupy a separate line and should not interrupt a turn.

APPENDIX B

LINDSEI TASKS

I'd like to interview you informally on things of interest in your life for fifteen minutes. To get the conversation started could you please choose one of the following topics and think about what you are going to say. You should aim to be able to talk for 3-5 minutes. The conversation will then continue informally.

Topic 1: An experience you've had which has taught you an important lesson.Youshould describe the experience and say what you have learnt fromit.

Topic 2: A country you have visited which has impressed you. Describe yourvisit and say why you found the country particularly impressive.

Topic 3: A film/play you've seen which you thought was particularly good/bad.Describe the film/play and say why you thought it was good/bad. Please don't take any notes as I would like it to be a spontaneous talk. Story for retelling:The four pictures below tell a story. Study the pictures and then make up a story around them.

APPENDIX C

PRE-TEST

Part 1 (2–3 minutes)

Interlocutor

(to both students)	Good morning/afternoon/evening. I'm _____ .
(to student A)	Now, what's your name? _____ Thank you.
(to student B)	And what's your name? _____ Thank you.

Interlocutor asks the following questions. Use students' names throughout. Ask student A first.

Where do you come from?

Do you think English will be useful for you in the future?

Thank you.
(Repeat for student B)

(Introduction to Part 2)

In the next part, you are going to talk to the researcher on a pre-identified topic.

Part 2 (2–3 minutes)

Interlocutor
(to student A first)

These are job applicants for a company. Who do you think will get the job? Why?

Thank you.

(Now repeat for student B)

Introduction to Part 3

In the next part, you are going to talk to each other.

Part 3 (2–3 minutes)

Examiner
Say to both students:

I'm going to describe a situation to you.

A young man is going to visit a city for the weekend, but he doesn't enjoy sightseeing. Talk together about the different things he could do in the city and say which would be most fun for him.

Here is a picture with some ideas to help you.

Ask both students to look at the picture and repeat the frame.

I'll say that again.

A young man is going to visit a city for the weekend, but he doesn't enjoy sightseeing. Talk together about the different things he could do in the city and say which would be most fun for him.

All right? Talk together.

Allow the students enough time to complete the task without intervention. Prompt only if necessary.

Thank you. This is the end of the test.

Pre-Test on the usage of stance markers in spoken learner English

Part 1 (2–3 minutes)

Interlocutor

(to both students)	Good morning/afternoon/evening. I'm _____ .
(to student A)	Now, what's your name?
	_____
	Thank you.
(to student B)	And what's your name?
	_____
	Thank you.

Interlocutor asks the following questions. Use students' names throughout. Ask student A first.

Where do you come from?

Do you think English will be useful for you in the future?

Thank you.
(Repeat for student B)

(Introduction to Part 2)

In the next part, you are going to talk to the researcher on a pre-identified topic.

Part 2 (2–3 minutes)

Interlocutor
(to student A first)

Who will say yes? Who will say no? What do you think? Why?

Thank you.

(Now repeat for student B)

Introduction to Part 3

In the next part, you are going to talk to each other.

Part 3 (2–3 minutes)

Examiner
Say to both students:

I'm going to describe a situation to you.
You're both going on a three-day school walking trip in the countryside.
Talk together about the things you will need, and decide which are the most important things to take with you.
Here is a picture with some ideas to help you.

Ask both students to look at the picture and repeat the frame.

I'll say that again.

You're both going on a three-day school walking trip in the countryside.
Talk together about the things you will need, and decide which are the most important things to take with you.

All right? Talk together.

Allow the students enough time to complete the task without intervention.
Prompt only if necessary.

Thank you. This is the end of the test.

Part 1 (2–3 minutes)

Interlocutor

(to both students)	Good morning/afternoon/evening. I'm _____ .
(to student A)	Now, what's your name? _____ Thank you.
(to student B)	And what's your name? _____ Thank you.

Interlocutor asks the following questions. Use students' names throughout. Ask student A first.

Where do you come from?

Do you think English will be useful for you in the future?

Thank you.
(Repeat for student B)

(Introduction to Part 2)

In the next part, you are going to talk to the researcher on a pre-identified topic.

Part 2 (2–3 minutes)

Interlocutor
(to student A first)

What do you think the people are doing? Where are they? What is happening?

Thank you.

(Now repeat for student B)

Introduction to Part 3

In the next part, you are going to talk to each other.

Part 3 (2–3 minutes)

Examiner
Say to both students:

I'm going to describe a situation to you.

A young man on holiday in **North America** wants to buy a **present** to take home to his **parents**. Talk together about the different presents he could **buy**, and say which would be **best**.

Here is a picture with some ideas to help you.

Ask both students to look at the picture and repeat the frame.

I'll say that again.

A young man on holiday in **North America** wants to buy a **present** to take home to his **parents**. Talk together about the different presents he could **buy**, and say which would be **best**.

All right? Talk together.

Allow the students enough time to complete the task without intervention. Prompt only if necessary.

Thank you. This is the end of the test.

APPENDIX D

SAMPLE STUDENT HANDOUT

LEARNING EPISTEMIC STANCE MARKERS VIA CONCORDANCE LINES

- Search the concordance to find the marker "I think".
- Generate the concordance list for the marker
- Look through the sample sentences and write them down in the following box.

SAMPLE SENTENCES "I THINK"
1.
2.
3.
4.
5.

- Looking at the sample sentences, in pairs or groups, discuss features of the particular sentence within the given context.

- Search the concordance to find the marker "Maybe".
- Generate the concordance list for the marker
- Look through the sample sentences and write them down in the following box.

SAMPLE SENTENCES "MAYBE"
1.
2.
3.
4.
5.

- Looking at the sample sentences, in pairs or groups, discuss features of the particular sentence within the given context.

- In pairs, write a short dialogue in which 'I think and Maybe' occur.

A SHORT DIALOGUE

- Can you find out general functional patterns in which 'I THINK and MAYBE' occur?

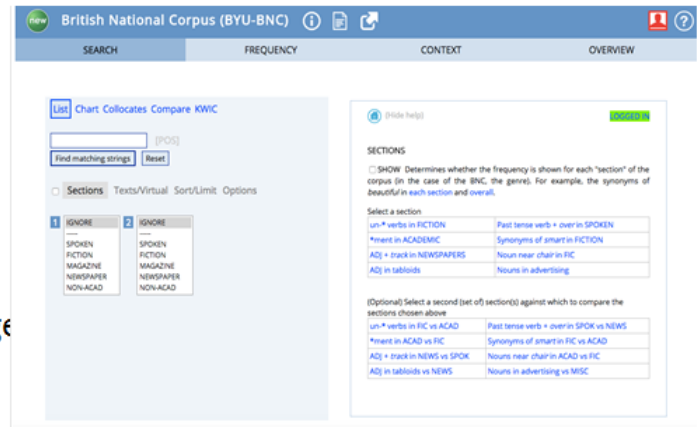
_____
_____
_____
_____
_____
_____

APPENDIX E

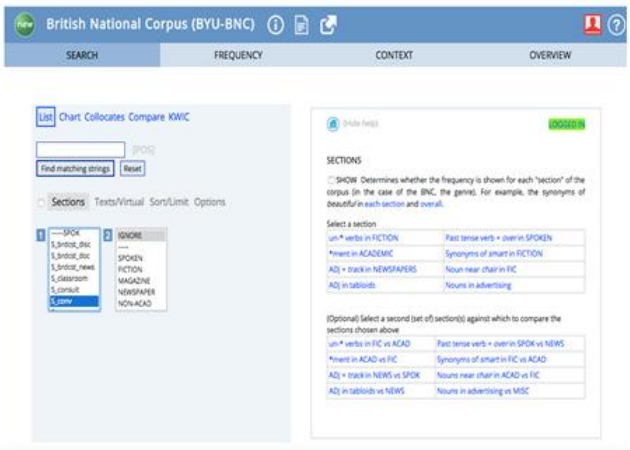
SAMPLE INSTRUCTIONAL MATERIAL

TO SEARCH FOR EPISTEMIC MARKERS

- 1. Type the word you want to search in the box under the "LIST" option
- Then click on the "SECTIONS" button
- On the right side of the page please click on "show" to activate it



- Type 'I think' in the search box
- Then, scroll down the options in SECTION 1 box and select "conv" under the "------SPOK" option.



- Now click on "FIND MATCHING STRINGS"
- Look at the search results and click on 'I think' again.
- You will see the concordance lines for 'I think'
- Now, write 5 of the example sentences on your worksheet.

I think…

- **1)** Starting point function
- Along with topic shifts
- Presentation of new perspectives on the same topic
- Indicating increased involvement and commitment
- **2)** on-line planning
- 3) Turn completion

Tentative & Deliberative Use of *I think*

- Tentative use = Used to express uncertainty
- Deliberative use = Used to increase commitment (Usually followed by that)

Examples

- *And ˆ**I think** it's contributing to my problems. I really do.*
- (indicating a degree of certainty & hightened speaker commitment)
- **Deliberative use**

- *Rebecca: Now the `bag that he was carrying around ˆthis time, what was it.*
  Rickie: .. let me `see,
  ***I ˆthink**,*
  ... a duffel bag?
- **Online planning & Tentative use**

  - *RICKIE: (SNIFF) `That was ˆthat, and then **I ˆthink**,*
    uh=,
    .. `couple of days ˆafter, .. `one of the `uh,
    .. ˆdetectives,
    `called me and `then I made a ˆreport. **(Online planning)**

  - Rebecca: ***I think** that's,*
    ***I `think** he he `finds somebody that's ˆisolated,*
    Rickie: [Mhm].
    *Rebecca: and he .. goes for a certain age group, **I `think**.*
    Rickie: Yeah.
    (Signaling completion of opinion sequences)

APPENDIX F

QUESTIONNAIRE (ENGLISH)

Using Concordance Lines To Learn Epistemic Markers In Speech

A.1. Background information

1. Name: _____
2. In general, do you like to use computer? Yes ☐ No ☐
3. How often do you use computer for personal purposes? (e.g. email) Check one.
   Several times a day ☐ About once a day ☐ About 5 times a week ☐
   About once a week ☐ About once a month ☐ Seldom ☐
   Others (list)_____

4. How often do you use computer for school work? (e.g. writing a paper) Check one.
   Several times a day ☐ About once a day ☐ About 5 times a week ☐
   About once a week ☐ About once a month ☐ Seldom ☐
   Others (list)_____

5. When you use computer for personal purposes, do you use English or your native language?
   English ☐ Native language ☐ Both ☐ Others(list) _____
6. How much of your total computer time is in your native language?
   About 75% ☐ About 50% ☐ About 25% ☐ Almost none ☐ Others(list) _____
7. Do you have internet access at home? Yes ☐ No ☐
8. Do you use Internet to improve your speaking skill?
9. Do you think the Internet is useful to improve your speaking skill?
10. Had you heard about *corpora* before you took this class? Yes ☐ No ☐
11. Had you used *corpora* before you took this class? Yes ☐ No ☐
12. If yes, which corpora did you use? _____
13. Do you use corpora at home and/or school?
    Home ☐ School ☐ Both ☐

A.2. Reactions to using concordance lines to learn EM's (epistemic markers) in spoken language

Scale Items	Strongly disagree	Disagree	Partly disagree	No opinion	Partly agree	Agree	Strongly agree
1. The corpus is more helpful than a dictionary for my English speaking ability	☐	☐	☐	☐	☐	☐	☐
2. Using the corpus is helpful for learning the meaning of EM's	☐	☐	☐	☐	☐	☐	☐
3. Using the corpus is helpful for learning the usage of EM's	☐	☐	☐	☐	☐	☐	☐
4. Using the corpus is helpful for learning the function of EM's	☐	☐	☐	☐	☐	☐	☐
5. Using the corpus has improved my understanding of certainty in spoken language	☐	☐	☐	☐	☐	☐	☐
6. Using the corpus has improved my understanding of uncertainty in spoken language	☐	☐	☐	☐	☐	☐	☐
7. I believe that I can express my stance appropriately when speaking after this instruction	☐	☐	☐	☐	☐	☐	☐

8. The use of concordance lists challenged me to actively make generalizations about the function of a marker.	☐	☐	☐	☐	☐	☐	☐
9. Concordance was useful for learning the stance markers in spoken language.	☐	☐	☐	☐	☐	☐	☐
10. I have some difficulty in using the corpus due to time and effort spent on analyzing the data	☐	☐	☐	☐	☐	☐	☐
11. I have some difficulty in using the corpus due to unfamiliar vocabulary on concordance/collocate output	☐	☐	☐	☐	☐	☐	☐
12. I have some difficulty in using the corpus due to cut-off sentences in concordance output	☐	☐	☐	☐	☐	☐	☐
13. I have some difficulty in using the corpus due to too many sentences in concordance output	☐	☐	☐	☐	☐	☐	☐
14. I have some difficulty in using the corpus due to the limited number of sentences in concordance output	☐	☐	☐	☐	☐	☐	☐
15. I have some difficulty in analyzing concordance output	☐	☐	☐	☐	☐	☐	☐
16. I have some difficulty in analyzing output for EM's	☐	☐	☐	☐	☐	☐	☐
17. I have some difficulty in performing the search technique	☐	☐	☐	☐	☐	☐	☐
18. The real texts in the corpus are too difficult to understand	☐	☐	☐	☐	☐	☐	☐
19. The searching technique was easy to learn to use as a reference when I practice speaking	☐	☐	☐	☐	☐	☐	☐
20. I understand the purpose of using the corpus in this activity	☐	☐	☐	☐	☐	☐	☐
21. When I need to get prepared for a spoken performance, I search for help in the corpus	☐	☐	☐	☐	☐	☐	☐
22. When I search for information in the corpus, I usually get the information that I need	☐	☐	☐	☐	☐	☐	☐
23. I use the corpus when practicing for other courses	☐	☐	☐	☐	☐	☐	☐
24. As I have learned more about the corpus, I have come to like them more	☐	☐	☐	☐	☐	☐	☐
25. I will use the corpus for my English speaking skill in the future	☐	☐	☐	☐	☐	☐	☐
26. Learning about the corpus has increased my confidence about speaking in English	☐	☐	☐	☐	☐	☐	☐
27. If I had used the corpus earlier, I would have had a better performance on speaking	☐	☐	☐	☐	☐	☐	☐
28. Overall, the corpus is a very useful resource for my English speaking.	☐	☐	☐	☐	☐	☐	☐
29. The corpus should be introduced in all EFL departments	☐	☐	☐	☐	☐	☐	☐
30. I will recommend the corpus to other students at Fırat University or elsewhere.	☐	☐	☐	☐	☐	☐	☐

APPENDIX G

QUESTIONNAIRE (TURKISH)

Konuşma Dilinde Tutum Belirteçlerini Öğrenmek için Derlemin (Corpus) Kullanılması

A.1. Kişisel bilgiler

1. Ad& Soyad: _____
2. Genel olarak bilgisayar kullanmayı sever misiniz? Evet □ Hayır □
3. Kişisel amaçlar için ne sıklıkta bilgisayar kullanırsınız?
   Günde bir kaç kez □ Ortalama günde bir kez □ Ortalama haftada 5 kez □
   Ortalama haftada bir kez □ Ortalama ayda bir kez □ Nadiren □
   Diğer _____

4. Okuldaki çalışmalarınız ve ödevleriniz için ne sıklıkta bilgisayar kullanırsınız?
   Günde bir kaç kez □ Günde bir kez □ Haftada yaklaşık 5 kez □
   Haftada bir kez □ Ayda bir kez □ Nadiren □ Diğer _____

5. Kişisel amaçlar için bilgisayar/internet kullandığınızda hangi dili kullanırsınız?
   İngilizce □ Ana dil □ Her ikisi □ Diğer _____

6. Toplam bilgisayar/internet kullanma zamanınızın ne kadarı ana dildedir?
   Yaklaşık 75% □ Yaklaşık 50% □ Yaklaşık 25% □ Hemen hemen hiç □ Diğer _____

7. Evde/yurtta internete erişiminiz var mı? Evet □ Hayır □
8. Konuşma becerinizi geliştirmek için internet kullanır mısınız? Evet □ Hayır □
9. Konuşma becerisini geliştirmede internetin faydalı olduğunu düşünüyor musunuz?
   Evet □ Hayır □
10. Bu uygulamaya katılmadan önce 'corpus' uygulamaları hakkında bilginiz var mıydı?
Evet □ Hayır □
11. Bu uygulamaya katılmadan önce 'corpus' uygulamalarını kullandınız mı?
    Evet □ Hayır □
12. Evet ise, hangi 'corpus'u kullandınız? _____
13. Corpus uygulamalarını evde ve/veya okulda kullanır mısınız?
    Evde □ Okulda □ Her ikisi □

A.2. Konuşma dilinde kullanılan tutum belirteçlerinin öğrenilmesi amacıyla derlemin (corpus) kullanılmasına dair tutumlar

Tutum belirteçleri (Epistemic Markers): Konuşmada yer alan kişisel değerlendirme, duygu ve görüş aktarımında kesinlik / olabilirlik (emin olma/ olmama) derecesini gösteren belirteçler (ör. Possibly, I don't know, I'm sure, I think, vs.)

	Scale Items	Kesinlikle katılmıyorum	Katılmıyorum	Kısmen katılmıyorum	Kısmen katılıyorum	Katılıyorum	Kesinlikle katılıyorum	Fikrim yok
	1. Corpus İngilizce konuşma becerisi açısından sözlükten daha yararlıdır.	☐	☐	☐	☐	☐	☐	☐
	2. Corpus'u kullanmak tutum belirteçlerinin anlamlarını öğrenmeme fayda sağlar.	☐	☐	☐	☐	☐	☐	☐
	3. Corpus'u kullanmak tutum belirteçlerinin cümle içinde kullanımını öğrenmeme fayda sağlar.	☐	☐	☐	☐	☐	☐	☐
	4. Corpus'u kullanmak tutum belirteçlerinin cümle içindeki fonksiyonunu öğrenmeme fayda sağlar.	☐	☐	☐	☐	☐	☐	☐
	5. Corpus'u kullanmak İngilizce konuşurken yargılarımın kesinliğini (emin olma durumu) nasıl belirteceğimi anlamama katkıda bulunur.	☐	☐	☐	☐	☐	☐	☐
	6. Corpus'u kullanmak İngilizce konuşurken yarıglarımın belirsizliğini(emin olmama durumu) nasıl belirteceğimi anlamama katkıda bulunur.	☐	☐	☐	☐	☐	☐	☐
	Scale Items	**Kesinlikle katılmıyorum**	**Katılmıyorum**	**Kısmen katılmıyorum**	**Kısmen katılıyorum**	**Katılıyorum**	**Kesinlikle katılıyorum**	**Fikrim yok**
	7. Bu uygulamadan sonra İngilizce konuşurken duruşumu (emin olma/olmama) gereğine uygun şekilde ifade edebileceğime inanıyorum.	☐	☐	☐	☐	☐	☐	☐
	8. Corpus'ta arama yaptığımda karşıma çıkan örnek cümleler listesi, bir belirtecin fonksiyonu hakkında genellemeler yapmak için beni teşvik etti.	☐	☐	☐	☐	☐	☐	☐
	9. Corpus'u kullanmak İngilizce konuşma dilindeki tutum belirteçlerini öğrenmemde faydalı oldu.	☐	☐	☐	☐	☐	☐	☐
	10. Verilerin analizinde harcanan zaman ve emek nedeniyle Corpus'u kullanmakta biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	11.Corpus'ta arama sonuçlarında karşıma çıkan örnek cümlelerdeki bilmediğim kelimeler nedeniyle corpus'u kullanmakta biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	12. Corpus'ta arama sonuçlarında karşıma çıkan örnek cümlelerin kesik/tamamının görünmüyor olması nedeniyle corpus'u kullanmakta biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	13. Corpus'ta arama sonuçlarında karşıma çıkan çok fazla sayıda örnek cümle olması nedeniyle corpus'u kullanmakta biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	14. Corpus'ta arama sonuçlarında karşıma çıkan örnek cümlelerin sınırlı sayıda olması nedeniyle corpus'u kullanmakta biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	15. Corpus'ta arama sonuçlarındaki veriyi analiz etmekte biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	16. Corpus'ta tutum belirteçleri için arama yaptığımda elde ettiğim veriyi analiz etmekte biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	17.Corpus'ta arama yaparken izlenen adımları uygulamada biraz zorlanıyorum.	☐	☐	☐	☐	☐	☐	☐
	18. Corpus'taki gerçek metinlerin anlaşılması çok zor.	☐	☐	☐	☐	☐	☐	☐

	19. Corpus'u İngilizce konuşma pratiği için bir referans olarak kullanmayı öğrenmek açısından; corpus'ta arama yapmak için izlenen adımlar kolaydı.	☐	☐	☐	☐	☐	☐	☐
	20. Bu uygulamada corpus'un kullanılış amacını anlıyorum.	☐	☐	☐	☐	☐	☐	☐
	21. İngilizce konuşmam gereken bir performansa(ödev, proje, sunum) hazırlanmam gerektiğinde, corpus'tan yardım alırım.	☐	☐	☐	☐	☐	☐	☐
	22. Corpus'ta bir arama yaptığımda, genellikle ihtiyacım olan bilgiye ulaşırım.	☐	☐	☐	☐	☐	☐	☐
	23. Diğer derslerime çalışırken corpus'u kullanırım.	☐	☐	☐	☐	☐	☐	☐
	24. Corpus hakkında daha fazla şey öğrendikçe, corpus'u daha çok sevmeye başladım.	☐	☐	☐	☐	☐	☐	☐
	25. Konuşma becerimi geliştirmek için corpus'u gelecekte de kullanacağım.	☐	☐	☐	☐	☐	☐	☐
	26. Corpus'u öğrenmek, İngilizce konuşmada kendime güvenimi artırdı.	☐	☐	☐	☐	☐	☐	☐
	27. Corpus'u daha önce kullanmış olsaydım, İngilizce konuşmada daha iyi bir performans gösterirdim.	☐	☐	☐	☐	☐	☐	☐
	28. Genel olarak, corpus İngilizce konuşma becerim için çok yararlı bir kaynaktır.	☐	☐	☐	☐	☐	☐	☐
	29. Corpus, İngilizce öğrenim görülen tüm bölümlerdeki öğrencilere tanıtılmalıdır.	☐	☐	☐	☐	☐	☐	☐
	30. Fırat Üniversitesi'ndeki veya diğer yerlerdeki öğrencilere Corpus'u tavsiye ederim.	☐	☐	☐	☐	☐	☐	☐

APPENDIX H

CERTIFICATE OF ORIGINALITY

SOSYAL BİLİMLER ENSTİTÜSÜ

YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

ÖĞRENCİ BİLGİLERİ	
Adı-Soyadı	Zehra SAVRAN
Öğrenci Numarası	141201102
Enstitü Anabilim Dalı	Batı Dilleri ve Edebiyatları
Programı	Yüksek Lisans
Danışmanının Unvanı, Adı-Soyadı	Yrd. Doç. Dr. Aysel ŞAHİN KIZIL
Tez Başlığı (Türkçe)	Öğrenici İnglizcesindeki Turtum Belirteçlerinin Karşılaştırmalı Analizine Dayalı Derlem Temelli Bir Çalışma: Derlemden Sınıfa

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

Yukarıda başlığı belirtilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 174 sayfalık kısmına ilişkin, 02/03/2017 tarihinde Sosyal Bilimler Enstitüsü tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 14 'tür.

Uygulanan filtrelemeler:

1- Kabul/Onay ve Bildirim sayfaları hariç,
2- Kaynakça hariç
3- Alıntılar hariç/dâhil
4- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Yukarıda bilgileri verilen öğrencinin yüksek lisans tezi Sosyal Bilimler Enstitüsü Yönetim Kurulu tarafından belirlenen azami benzerlik oranlarını aşmadığını ve tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim. Gereğini saygılarımla arz ederim.

Yrd. Doç. Dr. Aysel ŞAHİN KIZIL
.....
Danışmanın Adı-Soyadı
(İmzası)

Prof. Dr. Mehmet AYGÜN
.....
Anabilim Dalı Başkanı
(İmzası)

F.Ü.LİSANSÜSTÜ EĞİTİM ÖĞRETİM YÖNETMELİĞİ

Madde 41- Lisansüstü tezleri ile birlikte teslim edilmesi gereken belgeler şunlardır:

a) Lisansüstü tezler, savunma öncesinde **intihal program raporu** ve ilgili makale şartını sağladığına dair belgeleri ile birlikte enstitüye teslim edilir.
b) İntihal raporu ile ilgili olarak etik kurallar dâhilindeki benzerlik oranları ilgili Enstitü Yönetim Kurulu tarafından belirlenir. (Enstitü Yönetim Kurulu tarafından tezin, intihal kapsamı dışında değerlendirilmesi

APPENDIX I

ETHICAL COMMITTEE APROVAL

T.C.
FIRAT ÜNİVERSİTESİ
Girişimsel Olmayan Araştırmalar Etik Kurulu Başkanlığı

ETİK KURUL KARARI

TOPLANTI TARİHİ	TOPLANTI SAYISI	KARAR NO	ÇALIŞMACININ ADI SOYADI
14.06.2016	11	26	Yrd. Doç. Dr. Aysel ŞAHİN KİZİL

KARAR

"Öğrenici İngilizcesindeki Tutum Belirteçlerinin Karşılaştırmalı Bir Analizi: Derlemden Sınıfa" konulu çalışma etik kurulumuzda görüşülmüş olup; çalışmanın etik kurallara uygun olduğuna oybirliğiyle karar verilmiştir.

Prof. Dr. Mustafa KAPLAN (Başkan)			
Prof. Dr. Engin ŞAHNA (Üye)	İmza	Prof. Dr. Neriman ÇOLAKOĞLU (Üye)	İmza
Prof. Dr. Süleyman Serdar KOCA (Üye)	İmza	Prof. Dr. Demet ÇİÇEK (Üye)	Bulunmadı
Prof. Dr. Sefa KAZANÇ (Üye)	İmza	Prof. Dr. Ertan EVİN (Üye)	İmza
Prof. Dr. Erdal TAŞKIN (Üye)	İmza	Doç. Dr. Fatih FIRDOLAŞ (Üye)	Bulunmadı
Doç. Dr. Yalın Kılıç TÜREL (Üye)	İmza	Doç. Dr. Alper Osman ÖĞRENMİŞ (Üye)	İmza
Doç. Dr. Murat SUNKAR (Üye)	İmza	Doç. Dr. Yüksel SAVUCU (Üye)	Bulunmadı
Doç. Dr. Funda GÜLCÜ BULMUŞ (Üye)	İmza	Yrd. Doç. Dr. Nurhan HALİSDEMİR (Üye)	Bulunmadı

CURRICULUM VITAE

Zehra SAVRAN was born in 1990, in Denizli. She completed her primary and high school education in Denizli and received her bachelor's degree from the Department of English Language Teaching at Middle East Technical University in Ankara, Turkey, in 2012. She has been working as an English lecturer at Fırat University since 2012. Her research interests include corpus linguistics, computer-assisted language learning, self-regulated learning. Her research can be found in publications such as International Journal of Applied Linguistics and Novitas-ROYAL.