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

A CORPUS-BASED STUDY ON AUTHOR STANCE IN ACADEMIC ENGLISH

Reyhan AĞÇAM

A Ph.D. DISSERTATION

ADANA / 2014

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

A CORPUS-BASED STUDY ON AUTHOR STANCE IN ACADEMIC ENGLISH

Reyhan AĞÇAM

Supervisor: Assoc. Prof. Dr. Cem CAN

A PHD DISSERTATION

ADANA / 2014

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

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

Supervisor: Assoc. Prof. Dr. Cem CAN

Member of Examining Committee: Prof. Dr. Hatice SOFU

Member of Examining Committee: Prof. Dr. Birsen TÜTÜNİŞ

Member of Examining Committee: Assoc. Prof. Ahmet DOĞANAY

Member of Examining Committee: Asst. Prof. Dr. Abdurrahman KİLİMCİ

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

…./… / 2014

Prof. Dr. Yıldırım Beyazıt ÖNAL
Director of Institute

P.S: The uncited usage of the reports, charts, figures and photographs in this thesis, whether original or quoted for mother sources, is subject to the Law of Works and Thought No:5846

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı fikir ve Sanat Eserleri Kanunu'ndaki hükümlere tabidir. .

ÖZET

AKADEMİK İNGİLİZCEDEKİ YAZAR DURUŞU ÜZERİNE DERLEM TEMELLİ BİR ÇALIŞMA

Reyhan AĞÇAM

Doktora Tezi, İngiliz Dili Eğitimi Anabilim Dalı
Danışman: Doç. Dr. Cem CAN
Haziran 2014, 208 sayfa

Duruş; tutum, duygu, yargı ya da bir iletinin içeriği ile ilgili yüklenimin sözcüklerle ifadesidir (Biber & Finegan, 1989). Duruş, akademisyenlerin bir iddianın doğruluğu ya da güvenirliği ile ilgili yorum yaparken konularını açıklamada veya tutumlarını karşı tarafa aktarmada izledikleri yoldur (Hyland, 2005). Biber (2006)'a göre duruş; duygu ifadesi içeren sözcüklerle (*sevmek, nefret etmek* vb.), dilbilimsel olmayan yollarla (el hareketleri vb.) ve dilbilgisel yollarla (kiplik, yarı-kiplik, duruş zarfları ve tamlayan tümcecikleri) aktarılabilir. Bu çalışmada, Akademik İngilizcede yazar duruşunu aktarmada kullanılan bilgisel duruş araçları ele alınmıştır. Anadili İngilizce, İspanyolca ve Türkçe olan doktora öğrencileri tarafından İngiliz dilinde yazılmış tezler arasında söz konusu araçların kullanımı açısından anlamlı bir fark olup olmadığının ortaya çıkarılması amaçlanmıştır. Derlem temelli olan çalışma, söz konusu öğrenciler tarafından yazılmış 136 doktora tezinin Karşılaştırmalı Ara Dil Analizi (Granger, 1996) sonuçlarını içermektedir. Tezlerdeki duruş araçları, WordSmith Tools Programı (Scott, 2001) kullanılarak belirlenmiş ve *bilgisel zarflar*, *bilgisel fiiller*, *bilgisel sıfatlar* ve *bilgisel adlar* (Biber, 2006) olmak üzere dört ana grupta incelenmiştir. Daha sonra, bu araçlar *kesinlik* ve *olasılık* araçları olmak üzere her kategoride iki alt gruba daha ayrılmış ve her bir derlem içindeki sıklıkları hesaplanmıştır. Derlemler arasında söz konusu araçların kullanımı bakımından anlamlı bir farkın olup olmadığı Log Likelihood Test ile belirlenmiştir. Bulgular, üç derlemde de kesinlik araçlarının olasılık araçlarından daha sık kullanıldığını ancak anadili İngilizce olmayan doktora öğrencileri tarafından yazılan tezlerden oluşan iki derlemde iki grup arasında ölçülen farkın anadili İngilizce olan doktora öğrencileri tarafından yazılan tezlerin yer aldığı derlemde bulunandan çok daha yüksek olduğunu ortaya koymuştur. Sonuç olarak, anadili

İngilizce olan akademik yazarların düşünce ve öngörülerini aktarmada anadili İngilizce olmayan akademik yazarlara oranla daha temkinli olduğu düşünülmektedir.

Anahtar kelimeler: Yazar duruşu, bilgisel duruş araçları, aradil, karşılaştırmalı aradil analizi.

ABSTRACT

A CORPUS-BASED STUDY ON AUTHOR STANCE IN ACADEMIC ENGLISH

Reyhan AĞÇAM

Ph.D. Dissertation, Division of English Language Teaching
Supervisor: Assoc. Prof. Dr. Cem CAN
June 2014, 208 pages

Stance refers to the lexical and grammatical expression of attitudes, feelings, judgments, or commitment concerning the propositional content of a message (Biber and Finegan, 1989). It is the way academics annotate their texts to comment on the possible accuracy or creditability of a claim, the extent they want to commit themselves to it, or the attitude they want to convey to an entity, a proposition or the reader (Hyland, 2005). According to Biber (2006), it can be conveyed through value-laden word choice (*love, hate*, and etc.), paralinguistic devices (e.g. *gestures*) and grammatical stance devices (modal and semi-modals, stance adverbials and stance complement clauses). The present study was intended to cover the epistemic stance devices used in conveying author stance in Academic English. It investigated whether there is a statistically significant difference across the doctoral dissertations authored in English by native, Spanish-speaking and Turkish-speaking candidates with respect to the use of these devices. Being corpus-based in design, the study included the results elicited through the Contrastive Interlanguage Analysis (Granger, 1996) of a total number of 136 doctoral dissertations written by native and non-native speakers of English (Native: 45; Spanish: 43 and Turkish: 48). Following Biber (2006), epistemic stance devices found in the dissertations were identified through Wordsmith Tools (Scott, 2001) and examined in four groups (*epistemic adjectives, epistemic adverbs, epistemic nouns* and *epistemic verbs*). Subsequently, these devices were subcategorized into two groups as *certainty* and *likelihood* devices in each category and their frequencies were separately measured for each corpus. Finally, Log Likelihood Test was administered to see whether there is a statistically significant difference across the groups as regards the use of these devices. Findings of the study have indicated that certainty devices were used

more frequently than likelihood devices in three corpora and that the difference between the two categories in non-native corpora is higher than the one measured in the native corpus, which leads us to the conclusion that native academic authors are more cautious with their arguments and predictions while non-native academic authors are relatively more confident in their writing.

Keywords: Author stance, epistemic stance device, interlanguage, contrastive interlanguage analysis.

ACKNOWLEDGMENTS

I would like to express my gratitude to some people without whose support this dissertation could not have been completed. First of all, I would like to thank my supervisor Associate Professor Doctor Cem CAN who has been a great source of inspiration since the very first day I met him during my MA study. He has always encouraged me to go further and I believe, thanks to his trust in me, I have achieved things most of which I thought I could not.

I am obliged to the member of my examining committee Professor Doctor Hatice SOFU, whose great knowledge about linguistics and language education, and dedication to her work have fascinated me during my higher education at Çukurova University. I must confess that I became more conscious of how privileged we are for being trained by an excellent instructor like her especially when I began to teach English as a foreign language. I believe her conscientiousness and encouragement have contributed a lot to my professional development. In one of my visits to her office, I was really nervous because I had to make data analysis for a third or fourth time in this study due to some points I missed during the process, which also made me feel exhausted. When I told her about that and was expecting a look implying "how could you be so careless?", she just said, "that is the way doctorate is done; the process of writing a dissertation resembles a war which inevitably entails blood, sweat and tear". Upon those words, I, once again, admired the way she motivates her students no matter how frustrated they feel.

Learning how to become a good teacher requires attending teacher training programmes, reading books or periodicals on educational approaches, principles and techniques, participating in seminars and conferences and more. Yet, one can achieve this simply by observing Professor Doctor Birsen TÜTÜNİŞ teaching in a class or just talking about a topic that does not have to do with language education and taking her as a model. In a short period of time I have known her, I recognised that she is a great orator as she perfectly captures the attention of her audience no matter what she is talking about. My deepest gratitude goes to her for accepting a member of my examining committee and for her insightful comments and questions that pushed me to look at matters from a totally different perspective.

I would like to extend my deepest gratitude to my former supervisor Associate Professor Doctor Yasemin KIRKGÖZ who made great contributions to my professional development during both my undergraduate and graduate studies at Çukurova

University. I am also thankful to her for never turning me down when I wanted to visit her for personal reasons no matter how busy she was.

They say there is nobody perfect on earth and I used to believe that before meeting another member of my examining committee Associate Professor Doctor Ahmet DOĞANAY. The people who got to know him in a way are likely to think I am not exaggerating when I say he is someone like a God-given gift especially for his students. He teaches statistical analysis course which is mostly found challenging particularly by students studying social sciences so efficiently that at the end of the course, they cannot help asking themselves, "was it so easy?". I also propose that if a statistical analysis happened to be conducted on the feelings of students who visited him in the office or just came across him in the street and had a few words with him, the results would reveal %100 "positive". I will always feel privileged for being one of those students. My heartfelt thanks go to him for accepting to be a member of my examining committee and for his precious contributions to this study.

I owe special thanks to Assistant Professor Dr Abdurrahman KİLİMCİ for accepting to be in my examining committee and providing me invaluable feedback on this dissertation.

My heartfelt thanks are extended to all the other instructors working at the Division of English Language Teaching, Çukurova University and all my teachers in primary and secondary schools for doing more than teaching courses stated in the curricula to us. They have always endeavoured to raise individuals who will do their best for our country and the rest of the world.

Neither last nor least, I am indebted to Assistant Professor Doctor Muzaffer Pınar BABANOĞLU for being a great advisor, a true friend and another sister to me. I will remain thankful to her for the contributions she has made to my life since we began to work together in January 2010.

I would like to express my appreciation to Associate Professor Doctor Ramazan ÇETİNTAŞ who is the Head of School of Foreign Languages, Kahramanmaraş Sütçü İmam University for his understanding and encouragement during the period I have written this dissertation. I also would like to thank him for being our "leader" rather than an ordinary administrator since the day he started to work with us.

My warmest thanks go to my friends Fatma AVCIOĞLU and Mahmut ÖZKAN, for whom I feel full of pride to say, "I am their friend". Both have always been honest, sensitive and kind to me. Thanks to their constructive feedback on my professional and

personal lives, I believe I have been able to make right decisions on various issues. Another source of pride for me is that I am trying to do the same job with them. I will look forward to the day they will start to train outstanding teachers and counsellors just like themselves.

I also would like to thank the writers of doctoral dissertations that are included in this study. As I was supposed to construct my corpora within a limited period of time, I just took the related part of each dissertation trying not to pay attention to its content, which was really hard to me. Once completed mine, I feel very happy to have time to read those wonderful works.

Not many people are so lucky as I am for having a great family. I would like to thank them for trusting and supporting me in all my decisions and feeling even much happier than me whenever I achieved a success.

As stated in a well-known postal card published in France, I strongly believe "Turkey is indebted to God for having a leader like Mustafa Kemal ATATÜRK and it is indebted to him for the rest". I am mostly grateful to him for everything he did for our country in his lifetime. Without him, it might have been hard even to dream of learning to read and write my own name in my first language let alone writing a dissertation in another language and training teachers who are going to teach that language to others.

TABLE OF CONTENTS

Page

ÖZET ..... **iii**
ABSTRACT ..... **v**
ACKNOWLEDGEMENTS ..... **vii**
LIST OF ABBREVIATIONS ..... **xiii**
LIST OF TABLES ..... **xiv**
LIST OF FIGURES ..... **xvii**
LIST OF APPENDICES ..... **xviii**

CHAPTER I
INTRODUCTION

1.1. Background to the Study 1
1.2. Purpose of the Study 5
1.3. Research Questions 5
1.4. Significance of the Study 6
1.5. Limitations of the Study 6
1.6. Operational Definitions 6

CHAPTER II
REVIEW OF LITERATURE

2.0. Introduction 10
2.1. Academic Writing 10
2.1.1. Teaching Academic Writing 13
2.1.2. Corpus-Based Academic Writing 21
2.1.2.1. Stance and Stance Devices 25

CHAPTER III
METHODOLOGY

3.0. Introduction 42

3.1. Contrastive Interlanguage Analysis 42
3.2. Methodology for Corpora Analysis 43
3.3. Data 44
3.4. Instruments 46
3.4.1. WordSmith Tools 6.0 46
3.4.2. Log-likelihood Statistics 49
3.4.3. WMatrix 52
3.5. Data Analysis 55

CHAPTER IV
RESULTS AND DISCUSSION

4.0. Introduction 56
4.1. Results 56
4.1.1. Epistemic Stance Devices in Three Corpora 56
4.1.2. Certainty and Likelihood Devices in Three Corpora 65
4.1.2.1. Certainty Devices in Three Corpora 69
4.1.2.2. Likelihood Devices in Three Corpora 74
4.1.3. Individual Epistemic Stance Devices 79
4.1.4. Lexical Categories in Three Corpora 91
4.1.5. Individual Lexical Categories in Three Corpora 96
4.1.5.1. Adjectives in Three Corpora 96
4.1.5.2. Adverbs in Three Corpora 112
4.1.5.3. Nouns in Three Corpora 135
4.1.5.4. Verbs in Three Corpora 150
4.2. Discussion 172
4.2.1. Epistemic Stance Devices 173
4.2.2. Lexical Categories 176

CHAPTER V
CONCLUSION

5.0. Introduction 179
5.1. Evaluation of Research Questions 179

5.2. Implications of the Study 181
5.3. Suggestions for Further Research 184

REFERENCES 186
APPENDICES 193
CURRICULUM VITAE 206

LIST OF ABBREVIATIONS

BoE : Bank of English
BNC : British National Corpus
CIA : Contrastive Interlanguage Analysis
CIC : Cambridge International Corpus
CL : Corpus Linguistics
COBUILD : Collins Birmingham University International Language Database
DDL : Data-Driven Learning
EA : Error Analysis
EAP : English for Academic Purposes
EFL : English as a Foreign Language
ELT : English Language Teaching
ESD : Epistemic Stance Device
ESL : English as a Second Language
ESP : English for Specific Purposes
EPP : English for Professional Purposes
ICLE : International Corpus of Learner English
IL : Interlanguage
LGSWE : Longman Grammar of Spoken and Written English
L1 : First Language
L2 : Second Language
LL : Log-likelihood
LOCNESS : Louvain Corpus of Native English Essays
NAE : Native Academic Author of English
NACE : Native Academic Corpus of English
SAE : Spanish Academic Author of English
SACE : Spanish Academic Corpus of English
TAE : Turkish Academic Author of English
TACE : Turkish Academic Corpus of English
SLA : Second Language Acquisition

LIST OF TABLES

Page

Table 1. Traditional Topics in Academic Writing 21
Table 2. Model of Interaction in Academic Discourse 25
Table 3. Epistemic (Certainty and Likelihood) Stance Devices 27
Table 4. Attitude and Evaluation Stance Devices 27
Table 5. Common Grammatical Stance Devices 28
Table 6. Grammatical Stance Devices: Modal Verbs (and semi-modals), Stance Adverbs and Stance Complement Clauses 31
Table 7. Coding Doctoral Dissertations in Corpus 44
Table 8. Corpus Size 45
Table 9. Epistemic Stance Devices 45
Table 10. Contingency Table of LL Calculation 50
Table 11. ESD Types not Found in Three Corpora 57
Table 12. Overall Comparative Frequency Distribution of ESDs in Three Corpora 58
Table 13. LL Ratio of Overall ESDs in TACE and NACE 60
Table 14. LL Ratio of Overall ESDs in SACE and NACE 61
Table 15. LL Ratio of Overall ESDs in TACE and SACE 61
Table 16. Overused/underused ESD Categories in TACE in comparison with NACE .62
Table 17. LL Ratio of ESD Categories in TACE and NACE 63
Table 18. Overused/underused ESD Categories in SACE in comparison with NACE .63
Table 19. LL Ratio of ESD Categories in SACE and NACE 64
Table 20. Overused/underused Certainty and Likelihood Devices in TACE in comparison with NACE 66
Table 21. Overused/underused Certainty and Likelihood Devices in SACE in comparison with NACE 66
Table 22. Overused/underused Certainty and Likelihood Devices in TACE in comparison with SACE 67
Table 23. LL Ratio of Certainty and Likelihood Devices in TACE and NACE 67
Table 24. LL Ratio of Certainty and Likelihood Devices in SACE and NACE 68
Table 25. LL Ratio of Certainty and Likelihood Devices in TACE and SACE 68
Table 26. ESDs in Certainty Categories 69
Table 27. LL Ratio of Certainty Categories in TACE and NACE 70

Table 28. LL Ratio of Certainty Categories in SACE and NACE....72
Table 29. LL Ratio of Certainty Categories in TACE and SACE....73
Table 30. ESDs in Likelihood Categories....75
Table 31. LL Ratio of Likelihood Categories in TACE and NACE....76
Table 32. LL Ratio of Likelihood Categories in SACE and NACE....77
Table 33. LL Ratio of Likelihood Categories in TACE and SACE....78
Table 34. Most Frequently Used 10 ESDs in NACE....79
Table 35. Most Frequently Used 10 ESDs in TACE....80
Table 36. LL ratio of Overused ESDs in TACE....81
Table 37. LL Ratio of Underused ESDs in TACE....84
Table 38. Most Frequently Used 10 ESDs in SACE....86
Table 39. LL ratio of Overused ESDs in SACE....87
Table 40. LL ratio of Underused ESDs in SACE....89
Table 41. Frequency Distribution of Lexical Categories in Three Corpora....92
Table 42. Overused/underused Lexical Categories in TACE in comparison with NACE....93
Table 43. LL Ratio of Lexical Categories in TACE and NACE....93
Table 44. Overused/underused Lexical Categories in SACE in comparison with NACE....94
Table 45. LL Ratio of Lexical Categories in SACE and NACE....95
Table 46. LL Ratio of Identical Adjectives in SACE and NACE....97
Table 47. Most Frequently Used 50 Adjectives in TACE....98
Table 48. Most Frequently Used 50 Adjectives in SACE....100
Table 49. LL Ratio of Adjectives in TACE and NACE (Top 50)....102
Table 50. LL Ratio of Adjectives in SACE and NACE (Top 50)....106
Table 51. Most Frequently Used 50 Adverbs in NACE....113
Table 52. Most Frequently Used 50 Adverbs in TACE....114
Table 53. Most Frequently Used 50 Adverbs in SACE....116
Table 54. LL Ratio of Identical Adverbs in TACE and NACE (Top 50)....117
Table 55. LL Ratio of Identical Adverbs in SACE and NACE (Top 50)....124
Table 56. Most Frequently Used 50 Nouns in NACE....136
Table 57. Most Frequently Used 50 Nouns in TACE....137
Table 58. Most Frequently Used 50 Nouns in SACE....139
Table 59. LL Ratio of Identical Nouns in TACE and NACE (Top 50)....141

Table 60. LL Ratio of Identical Nouns in SACE and NACE (Top 50) 143
Table 61. Most Frequently Used 50 Verbs in NACE 151
Table 62. Most Frequently Used 50 Verbs in TACE 153
Table 63. Most Frequently Used 50 Verbs in SACE 154
Table 64. LL Ratio of Identical Verbs in TACE and NACE (Top 50) 156
Table 65. LL Ratio of Identical Verbs in SACE and NACE (Top 50) 159

LIST OF FIGURES

Page

Figure 1. A Screenshot of the Main Screen of WS Tools....46
Figure 2. A Screenshot of Concordance Menu in WS Tools....47
Figure 3. A Screenshot of a Wordlist in WS Tools....48
Figure 4. A Screenshot of Keyword Page in WS Tools....49
Figure 5. A Screenshot of Log-likelihood Calculator....50
Figure 6. A Screenshot of Log-likelihood Calculator Results....51
Figure 7. A Screenshot of the Main Page of Wmatrix....52
Figure 8. A Screenshot of a Sample Folder uploaded to Wmatrix....53
Figure 9. A Screenshot of the Tag Wizard in Wmatrix....54
Figure 10. A Screenshot of Word POS Frequencies....54
Figure 11. Distribution of ESD Types in Three Corpora....58
Figure 12. Epistemic Categories in Three Corpora....59
Figure 13. Certainty and Likelihood Devices in Three Corpora....65
Figure 14. Certainty Categories in Three Corpora....70
Figure 15. Likelihood Categories in Three Corpora....75
Figure 16. Lexical Categories in Three Corpora....91
Figure 17. Epistemic Adjectives and All Adjectives in Three Corpora....96
Figure 18. Identical adjectives in three corpora (1-10)....109
Figure 19. Epistemic Adverbs and All Adverbs in Three Corpora....112
Figure 20. Identical Adverbs Three Corpora Significantly Differ in Use....132
Figure 21. Identical Adverbs in SACE and NACE against TACE....133
Figure 22. Epistemic Nouns and All Nouns in Three Corpora....135
Figure 23. Identical Nouns Three Corpora Significantly Differ in Use....145
Figure 24. Epistemic Verbs and All Verbs in Three Corpora....151
Figure 25. Identical Verbs of Similar Frequencies over Three Corpora....163
Figure 26. Identical Verbs Three Corpora Significantly Differ in Use....165
Figure 27. Identical Verbs in SACE and NACE against TACE....166
Figure 28. Identical verbs in TACE and NACE against SACE....169
Figure 29. Identical Verbs in TACE and NACE against SACE....171

LIST OF APPENDICES

Page

Appendix A. Dissertations in NACE 193
Appendix B. Dissertations in TACE 197
Appendix C. Dissertations in SACE 202

CHAPTER I

INTRODUCTION

1.1. Background to the Study

A corpus is a principled collection of texts available for qualitative and quantitative analysis (Biber et al, 1998). McEnery and Wilson (2001) propose that any collection of more than one text may be identified as a corpus when taken as *body*, its counterpart in Latin and that it entails more specific connotations such as sampling and representativeness, finite size, machine-readable form and a standard reference when considered in the context of modern linguistics. It is defined by Meyer (2002) as a collection of texts or parts of texts upon which some general linguistic analysis can be conducted. Stubbs (2004) maintains that a language corpus refers to a text collection, which is large, computer readable and designed for linguistic analysis. According to O'Keeffe et al (2007), it is a collection of texts, written or spoken, which is stored on a computer.

Granger (1998) holds that corpus linguistics is a new way of thinking about language, which is challenging some of the deeply-rooted ideas about language and that it may be considered to contrast with the Chomskyan approaches to language learning due to its focus on performance (rather than competence), description (rather than universals) and quantitative (rather than qualitative) analysis. In a similar vein, McEnery and Wilson (2001) do not count it as a branch of linguistics in the same sense as syntax, semantics and sociolinguistic but a methodology that may be utilized in various areas of linguistics, enabling us to discriminate between methodological approaches employed in the study of language. Confirming the belief that corpus is not (necessarily) applied linguistics or a branch of linguistics, Stubbs (2004) identifies it as "an empirical approach to studying language, which uses observation attested data in order to make generalizations about lexis, grammar and semantics" (2004, p.107). He suggests that it aims to improve language description and theory and that applied linguistics is supposed to be in charge with the assessment of this work to practical applications. He also advocates that corpus data are essential for the accurate description of language use clarifying the interaction among lexis, grammar and semantics.

Corpus Linguistics can be used for syllabus design, materials developments and classroom activities (Krieger, 2003). O'Keeffe et al (2007) consider language corpora as standard tools, which use word corpora to investigate word frequency, patterning and semantics in the compilation of dictionaries. They suggest that Samuel Johnson compiled the first corpus-based dictionary of the English language building a corpus which included samples of usage from the period 1560 to 1660. Besides, they state that Cambridge International Corpus (CIC), which consists of over 27 million words of written productions of learners, provides valuable information about the types of lexical and grammatical errors made by learners enabling dictionary writers and other materials producers to emphasize typical problems. They cited the Collins Birmingham University International Language Database (COBUILD) project which was launched by John Sinclair at the University of Birmingham in 1980, previously identified by Granger (1998) as a pioneering study which gave rise to a variety of EFL tools based on authentic data. Römer (2010) notes that the project in concern was developed with the aim of providing English language learners more functional dictionaries and teaching materials which present real language and focus on the items and meanings they are more likely to encounter in actual communicative situations. She states that the project yielded the Bank of English (BoE, hereafter), an increasing multimillion word corpus of spoken and written English including the productions of English language learners with various mother tongue backgrounds and that BoE-based dictionaries, grammars and usage guides contain genuine examples rather than invented ones. She concludes that over the past 25 years, corpora, corpus tools and corpus evidence have not only constituted a basis for linguistic research but they have also been used in the teaching and learning of languages such as Tim Johns' data-driven learning (DDL) and Dieter Mindt's empirical grammar research and John Sinclair's aforementioned work with COBUILD.

Stubbs (2004) asserts that studies conducted within the field of CL have applications in language teaching, translation, forensic linguistics, and broader cultural analysis in a direct or indirect manner. He also claims that advanced language learners can benefit from a corpus as long as they have an access to it; namely, they can see how the items they have newly learned are used in various contexts. He is of the opinion that corpora provide observable evidence about language use leading to new descriptions, which are exemplified in dictionaries, grammars, and teaching materials. According to him, corpus-based dictionaries such as CIDE (1995), COBUILD (1995), LDOCE (1995),

OALD (1995) and grammars target mass of which are advanced language learners and which include authentic examples of current usage such as COBUILD (1990) have been created since the late 1980s. He holds that corpora are sources of evidence available both to theoretical and applied linguists; namely, a corpus of audio-recorded conversations might simultaneously be utilized by a sociolinguist who investigated relations between social class and accent, by a psycholinguist who studies slips of the tongue and by a lexicographer who examines the frequency of different phrases.

Granger (1998) argues that the main goal of second language acquisition research is to uncover the principles that govern the foreign/ second language (SLA, hereafter) learning process which needs to be accessed via learner performance data for its being mental and therefore not directly observable. In this sense, she advocates that learner corpora are of vital importance in SLA research underlining the following facts:

1. People incline not to use natural language for various reasons,
2. Variables affecting language use can not be investigated systematically for being not controlled,
3. Natural language use data fail to reveal the entire linguistic repertoire of learners because they tend to use only aspects in which they feel confident most.

She notes that learner corpus flourished in the period when Error Analysis (EA) was popular stressing most of them were very small (less than 2000 words) while some were too large to estimate such as the one used in the Danish Project in Foreign Language Pedagogy (PIF) conducted under the direction of Faerch et al (1984), which she attributes to the fact that counting the number of words was an impossible task due to the lack of computerizes corpora. According to her, heterogeneity of the learner data constituted a limitation to the early learner corpora because of the factors, as put by Odlin (1989, p 151), such as “considerable variation in the number of subjects, in the backgrounds of the subjects, and in the empirical data, which come from tape-recorded samples of speech, from student writing, from various types of texts and from other sources”. She also maintains that early learner corpora were only considered as depositories of errors rather than as corpora in their own right, which is another weakness of most of those corpora. She highlights three foremost points of the today’s learner corpora:

1. They are much larger making it possible to analyze most language features
2. They have much stricter design criteria requiring investigations of the different variables affecting learner performance.
3. They are computerized.

All in all, she advocates that computer learner corpora provide us with access to errors produced by learners as well as their whole interlanguage (IL, henceforth). Granger and Tribble (1998) contend that learner data enable us to estimate where learners or a particular group of learners (i.e. those with the same mother tongue background or with the same level of proficiency) have encounter difficulty in their attempt to learn a foreign/ second language and that especially error-coded learner corpora can be employed in order to raise their consciousness of grammatical, lexical and discourse features which distinguish their IL from the target language. For example, as she cites, language corpora including productions of non-native speakers can be used in a foreign/ second language classroom in handling overuse or underuse of certain vocabulary.

Campoy-Cubillo et al (2010) suggest that corpus-based approaches lead to the elaboration of better quality learner input and provide teachers and researchers with a broader horizon in language use enabling them to understand how language works in specific contexts.

Discriminating between general corpora and specialized corpora, Connor and Upton (2002) argue that the former are most effective for the study of language structure and use while the latter are effectively used in exploring language in academic and professional settings since they focus on specific genres. Likewise, Flowerdew (2004) is of the opinion that specialized corpora offer more contextual information than larger corpora. Campoy-Cubillo et al (2010) advocate that genre analysis could benefit from the use of specialized corpora enabling one to understand the function and use of language more accurately. They also suggest that teachers offering the course of English for Professional Purposes (EPP, henceforth) are gradually becoming more eager to have their students produce texts based on language uses, which are required for their profession in the future. English for Specific Purposes (ESP, henceforth) is destined to play an increasingly vital role in language teaching due to the growing demand for professionals with specialized knowledge and skills (Crawford-Cemiciottoli, 2010).

1.2. Purpose of the Study

The present study scrutinised author stance through the sections *Findings, Discussion, Conclusion, Pedagogical Implications (Implications to English Language Teaching)* and *Suggestions for Further Research* of the doctoral dissertations produced by native and non-native academic authors of English within the fields of English Language Teaching, English Language and Literature, Applied Linguistics and Modern Languages between 2005 and 2012. Initially, the dissertations written by Turkish-speaking academic authors of English (TAE, henceforth) and those by native academic authors of English (NAE, hereafter) was analysed through WS Tools (6.0) to identify epistemic stance devices (ESDs, henceforth) across the abovementioned sections and the related type/ token ratio was calculated for each corpus. Subsequently, the findings were compared using Log Likelihood Test to see whether there is a statistically significant difference between the groups with regard to the use of these devices. The same procedure was followed to find out whether there is a difference of such kind across the related sections of the dissertations produced by Spanish-speaking academic authors of English (SAE, henceforth) and those by the native authors. Lastly, findings elicited from the analyses of the corpora including dissertations of non-native authors (TAEs and SAEs) were examined for Contrastive Interlanguage Analysis (CIA) to see whether they significantly differ in terms of ESDs.

1.3. Research Questions

Our study sought for answers to the following questions:

1. Do native academic authors of English and Turkish-speaking academic authors of English significantly differ in the use of English ESDs?
2. Do native academic authors of English and Spanish-speaking academic authors of English significantly differ in the use of English ESDs?
3. Do Turkish-speaking academic authors of English and Spanish-speaking academic authors of English significantly differ in the use of ESDs?

1.4. Significance of the Study

This study focused on native and non-native English-speaking academic authors' use of epistemic stance devices in their doctoral dissertations and it aimed to contribute to the studies within the scope of author stance in academic writing. It investigated whether two groups of non-native academic authors with high proficiency in English significantly differ from the native speakers of English in terms of author stance. It also attempted to reveal whether the non-native groups significantly differ from each other in this concern. The study is hoped to shed light into the discipline of teaching academic writing especially to ESL learners at graduate level offering some reasons and solutions to the problems these learners might encounter while using the academic language in their writing.

1.5. Limitations of the Study

The present study is limited to the size of three corpora; NACE as a corpus native academic authors of English, TACE as a corpus of Turkish-speaking academic authors of English and SACE as a corpus of Spanish-speaking academic authors of English. In addition, its findings are limited to the certain sections of the works produced in the fields of English Language Teaching, English Language and Literature, Applied Linguistics and Modern Languages between 2005 and 2012. Hence, the results of the study cannot be generalized to academic writers with Turkish, English and Spanish as mother tongue and even to the other works of the writers whose doctoral dissertations are included in this study. Finally, misuse of the epistemic stance devices has been excluded from this study owing to the fact that its scope is limited to overuse and underuse of epistemic stance devices across the dissertations of non-native academic authors in comparison to those of native academic authors of English and to comparison between the dissertations produced by two groups of non-native academic writers of English in terms of these devices.

1.6. Operational Definitions

Academic English: A variety of English that refers to the language used by the educated and is needed to function at the university level and beyond (Uribe, 2008). The

language of school-based learning and extended, reasoned discourse (Gersten et al., 2007).

Academic Writing: 'Structured research' written by 'scholars' for other scholars (with all university writers being 'scholars' in this context). Academic writing addresses topic-based 'research questions' of interest to anyone who is seeking factually-based, objectively-presented information on a particular topic (retrieved from http://www15.uta.fi/FAST/FIN/RESEARCH/acadwrit.html).

Bank of English (BoE): Both a reference corpus and a monitor corpus of general English launched in 1991 by COBUILD and the University of Birmingham. It is reported to consist of 525 million words of spoken and written language in April 2005. The texts are in British English, North American English and other varieties of English. (Baker et al., 2006)

Cambridge International Corpus (CIC): A corpus based on a 100.000 word sample of newspaper and magazines from McCarthy (1998, p. 122-123)

Certainty Stance Device: Vocabulary that signals the degree of confidence (certainty) that writers have towards claim with his research (retrieved from http://sana.tkk.fi/awe/style/vocabulary/).

Corpus: A principled collection of electronic texts usually stored on a computer and available for qualitative and quantitative analysis (O'Keffee et al, 2007).

Corpus-Based Academic Writing: Scientific writing structured on data obtained from a particular corpus.

Corpus Linguistics: Study of language based on examples of real life language use and methodology rather than an aspect of language requiring explanation or description. (McEnery and Wilson, 1996 p.231).

Concordance: Referred as key word in context (KWIC), a list of all occurrences of a particular search term in a corpus, presented within the context in which they occur – usually a few words to the left and right of the search term. (Baker et al., 2006)

Contrastive Interlanguage Analysis (CIA): An approach to corpus linguistics that compares native and learner varieties of language in order to shed light on non-native features of learner writing and speech through detailed comparisons of linguistic features in native and non-native corpora. It also involves comparing non-native varieties of language by eliciting data from different mother tongue backgrounds to differentiate between features which are shared by several learner populations and are more likely to be developmental and those which are peculiar to one national group and

therefore possibly L1-dependent (Granger, 1996).

Data-driven Learning (DDL): A technique proposed by Johns (1997), used in language teaching whereby the student takes a pro-active role in their own learning process by carrying out and analysing a series of concordances from a corpus. (Baker et al., 2006)

English as a Foreign Language (EFL): A variety of English that is not used in the learner's home country.

English as a Second Language (ESL): A variety of English that is spoken in the learner's home country.

English for Academic Purposes (EAP): A programme that entails training students, usually in a higher education setting, to use language appropriately for study. It is a challenging and multi-faceted area within the wider field of English language learning and teaching (ELT), and is one of the most common forms of English for specific purposes (ESP) (Retrieved from http://en.wikipedia.org/).

English for Specific Purposes (ESP): An approach focusing on studying and teaching specialized varieties of English, most often to non-native speakers of English, in advanced academic and professional settings (Bawarshi & Reiff, 2010).

First Language: The language that a child learns, also known as primary language, native language, or mother tongue. (Gass and Selinker, 2008).

Interlanguage: A term devised by Selinker (1972), referring to the linguistic rules and patterns that learners of a second language build for themselves. (Baker et al., 2006)

LGSWE: A descriptive grammar of English describing the actual grammatical features in different varieties of English: mainly conversation, fiction, newspaper language, and academic prose (Biber et al, 1999).

Log-Likelihood: A test for statistical difference, similar to the chi-squared measure that is often used in corpus analysis such as collocation or keyword analysis. (Baker et al., 2006)

Michigan Corpus of Academic Spoken English (MICASE): A corpus composing academic speech sampled from a range of contexts at the University of Michigan, USA. (Baker et al., 2006)

Second Language Acquisition (SLA): The term referring to the process of learning another language after the native language has been learned. (Gass and Selinker, 2008).

Native Academic Corpus of English (NACE): A set of data including the sections

of results and discussion, conclusion, pedagogical implications and suggestions for further research doctoral dissertations written by native academic authors of English between 2005-2012.

Spanish Academic Corpus of English (SACE): A set of data including the sections of results and discussion, conclusion, pedagogical implications and suggestions for further research doctoral dissertations written by Spanish-speaking academic authors of English between 2005-2012.

***That*-clause:** A clause with a nominal function which may serve virtually all of the functions served by NPs (Brinton, L.J., 2000).

***To*-clause:** A clause that precedes many catenative verbs functioning as a nominal, an adverbial clause, a post-modifier, or an adjective complement (Chalker, S. & Weiner, E., 1994)

The TOEFL 2000 Spoken and Written Academic Language (T2k-SWAL) Corpus: Corpus was designed to represent the full range of spoken and written registers used at U.S. universities (e.g., classroom teaching, office hours, study groups, textbooks) as well as in the major academic disciplines (e.g., humanities, natural sciences) and academic levels (lower division, upper division, and graduate) (Biber et al., 2002).

Turkish Academic Corpus of English (TACE): A set of data including the sections of results and discussion, conclusion, pedagogical implications and suggestions for further research doctoral dissertations written by Turkish-speaking academic authors of English between 2005-2012.

Wordsmith (WS): A software package developed by Scott (2001) for analysing the lexis of texts and corpora which can be used to produce frequency lists, to run concordance searches and calculate collocations for particular words, and to find keywords in a text and examine their distribution. (Baker et al., 2006)

Wmatrix: A software tool for corpus analysis and comparison. It provides a web interface to the USAS and CLAWS corpus annotation tools, and standard corpus linguistic methodologies such as frequency lists and concordances. It also extends the keywords method to key grammatical categories and key semantic domains (Retrieved from http://ucrel.lancs.ac.uk/wmatrix/).

CHAPTER II

REVIEW OF LITERATURE

2.0. Introduction

This chapter, first of all, provides information about academic writing and academic writing instruction (e.g., definition and theoretical assumptions about the process). Then, corpus-based academic writing, one of the most well known approaches to academic writing, is introduced. Finally, stance and stance devices, which are investigated throughout the study, are described and outlined.

2.1. Academic Writing

Academic writing is defined by Oshima and Hague (1994) as the kind of writing that requires to be applied in college or university. Likewise, Irvin (2010) describes it as "a form of evaluation that requires demonstrating knowledge and showing proficiency with certain disciplinary". She maintains that style of writing investigates the state of an issue and presents one's position based on the evidence of his/her research.

Oshima and Hague (1994) hold that academic writing differs from other kinds of writing (personal, literary, journalistic, business and etc.) in audience, tone and purpose. According to them, academic writers, first of all, should be concerned with their audience who are primarily their professors. In a similar vein, it is anticipated that they should show their audience that they have a through knowledge of the material and the subject. Secondly, academic writers should be concerned with tone of their writing since it reveals their attitude towards their subject by their choice of words, grammatical structures as well as length of sentences. They also advocate that a letter to a friend may contain some slang expressions and active verb forms, which is not acceptable in academic writing. Instead, a scientific paper is expected to contain more technical vocabulary and passive verb forms; therefore, its tone would probably be highly formal and impersonal. Likewise, it is argued that academic tone is formal, objective and factual. Oshima and Hague (1994) contend that it is the target audience rather than the subject matter that is more likely to determine the tone of writing and that academic writing is always formal in tone. In a similar vein, Irvin (2010) claims that writing resembles having a blindfold over eyes and hands tied behind backs making it

impossible for authors to know exactly about their audience and position. Accordingly, it becomes essential to use punctuation and to choose words meticulously in order to communicate their tone. She notifies that setting the right tone depends on such issues as the nature and purpose of the task, the lecturer's requirements, the assignment type, the social distance between the writer and reader and the writing's being factual, objective and supported by evidence.

As for the matter of purpose, Oshima and Hague (1994) emphasize that academic writers should have a specific and clear purpose in their writing. They discriminate the purpose of academic writing from those of other kinds of writing and assert that the purpose of literary writing is probably to entertain and that of journalistic writing is most likely to seek to inform or persuade whereas the main purpose of academic writing is to explain no matter how it may also be to persuade or convince the audience of the correctness of a particular point of view on a particular issue.

Hyland (2006) proposes that the formality in academic texts is achieved through the use of specialist vocabulary, impersonal voice and the ways express ideas with relatively few words. He classifies these features into three main areas as "high lexical density", "high nominal style" and "impersonal constructions". What he means by high lexical density is that content words occur in higher proportions when compared to grammar words such as prepositions, articles and pronouns, which he believes make "more tightly packed with information". High nominal style requires actions and events to be presented as nouns rather than verbs "to package complex phenomena as a single element of a clause". Finally, he maintains that academic prose should be kept as impersonal as possible and the use of first person pronouns should be replaced with passive constructions.

Features of academic writing are described as complexity, formality, objectivity, explicitness, hedging and responsibility. Academic English is relatively more complex than spoken English (Biber, 1988; Halliday and Hasan, 1989).

> Grammatically, Academic English has more subordinate clauses, more that/ to complement clauses, more long sequences of prepositional phrases, more attributive adjectives and more passives than spoken language. Lexically, academic writing often requires longer, more complex words and phrases. This kind of writing is lexically dense compared to spoken language –it has

> proportionately more lexical words than grammatical words (Nga, 2009, p. 114).

What is referred to formality in academic writing is avoiding certain words and phrases as well as grammatical structures in the production process of scientific writing. Nga (2009) stresses that colloquial words and expressions (e.g., *stuff, thing, sort of*), abbreviated forms (e.g., *can't, doesn't, shouldn't*), two word verbs (e.g., *put off, bring up*) and sub-headings, numbering and bullet points in formal writings.

Academic writing, particularly in scientific fields, has been characterized as objective, written in the third person, and distanced from personal feelings and experiences (Burke, 2010). In a similar vein, Nga (2009) maintains that it, therefore, includes fewer words that refer to the writer or the reader, which means the emphasis should be on the information that is intended to provide and the arguments to be made rather than the writer himself/ herself.

Academic writing is explicit in its signposting of the organisation of the ideas in the text (Biber et al, 1999). The lack of explicitness causes a few (if any) problems, because the expert reader anticipates the expected readings that will occur in this context and thus, a compact (and inexplicit) style actually helps readers keep up with the increasing volume of information produced by scientific researchers (Biber and Gray, 2010).

Hedging is a collective term used to refer to those linguistic elements that express possibility, probability, and uncertainty, thus reducing the degree of the author's commitment to the content of a statement (Pisanski-Peterlin, 2010).

Nga (2009) states that academic writers should be responsible for providing evidence and justification for the claims they make and for demonstrating an understanding of any source text used they use in their writing.

Thaiss and Zawacki (2006) have determined three characteristics of academic writing via the study through which they investigated professors' perceptions about academic writing and its standards. The following are the characteristics in concern.

1. There should be clear evidence in writing that the writer(s) have been persistent, open-minded and disciplined in the study.
2. Reason should be dominant over emotions or sensual perception.
3. In the writing process, the writer should imagine readers who are rational, reading for information and intending to formulate a reasoned response.

Maroko (2013) posits that postgraduate students and their lecturers must gain fluency in the conventions of academic writing in English to understand their disciplines, to establish their careers or to successfully develop their learning. The following are the universal conventions of academic discourse that Thonney (2011) find beneficial for the students in their first year at university. According to her, academic writers;

- respond to what others have written about the topic they are concerned,
- state the value of their work and announce the plan for their papers,
- acknowledge that others might disagree with the position they have taken,
- adopt a voice of authority,
- use academic and discipline-specific vocabulary,
- emphasize evidence, often in tables, graphs, and images.

Maroko (2013) examined frequency and usage of stance-marking features in academic writing across selected dissertations submitted to Kenyan Public Universities. He found that personal pronouns and the third person pronoun were extensively used in dissertations produced within the field of humanities while the 'faceless' agentless passive voice was mostly preferred in those written in hard sciences. According to him, this suggests that choices for such features in dissertations are a function of the epistemology and ideology of the disciplines. In conclusion, he proposes a genre-based approach to teaching those preparing to write their dissertations.

2.1.1. Teaching Academic Writing

Academic Writing is arguably the most important language skill at university because students' grades are largely determined by their performance in written assignments, tests and examinations or graduation thesis (Leki and Carson, 2009). Kaplan (2001) emphasizes foreign students attending universities in the USA might be enable to compose adequate themes, term papers, theses and dissertations despite having a good knowledge of syntactic structures in English. Namely, their written production can be out of focus as the rhetoric and sequence of thought they use do not fulfil the expectations of the native reader. Hence, he posits that a student can write a good essay in his/ her native language does not necessarily mean that particular student

can achieve this in a second language. He analysed six hundred compositions written by foreign students studying at US universities. He found that paragraphs of Arabic-speaking students developed on the basis of a complex series of parallel constructions. Namely, conjunctions and sentence connectors were extensively used in those productions, which is linguistically acceptable in their native language but not in English, which favours subordinate clauses rather than coordinators in writing. He believes that such kind of paragraphs would be striking for the native readers and they would be considered to violate the principles of clear communication when produced in modern English. He also examined professional translations of the texts originally written in French, Russian and Spanish. He found that digressions and extraneous materials are more favoured in these languages than English. He suggests some difficulties in paragraphs written in Russian are linguistic since the structure of the sentences in Russian greatly differs from that in English and some others are rhetorical as the paragraphs was composed of "a series of presumably parallel constructions and a number of subordinate structures" and half of the sentences were not related to the central idea, which is again unlikely to meet the expectations of the native speaker of English. In order to prevent or overcome such problems, he offers;

> In teaching of paragraph structure to foreign students, whether in terms of reading or in terms of composition, the teacher must be himself aware of these differences, and he must take these differences overtly apparent to his students. In short, contrastive rhetoric must be taught in the same sense that contrastive grammar is presently taught… The patterns of paragraphs in other languages are not so well established, or perhaps only not so well known to speakers of English. these patterns need to be discovered or uncovered and compared with the patterns of English in order to arrive at a practical means for the teaching of such structures to non-native users of the language (Kaplan, 2001, p. 20).

He suggests that an authentic paragraph written in English might be arbitrarily scrambled, the sentences numbered, and the students asked to unscramble the sentences in an order they find "normal". He believes this kind of task is likely to reveal diversity of views or cultures represented in the classroom and that, at the end of the task, the

instructor should be able to explain and justify the right order by presenting students the original version of the paragraph.

As a result of contrastive analyses of various languages, Clyne (1994) proposes five cultural parameters: (1) English cultures more strongly foreground form while other cultures are more content-oriented (p. 186); (2) English cultures stress the written language as the main medium of effective communication while other cultures stress oral language (p. 189); (3) English cultures tend to stress symmetry and do not stress positive politeness while other cultures function differently in both contexts (p. 190); (4) English cultures tend to be unique in their emphasis on linearity (p. 190); (5) English cultures tend to stress concreteness and reasoning (p. 191). Kaplan (197) highlights that those parameters should not be considered as polar oppositions and that "culturally-defined writing practices tend to move in one or the other direction along the posited continua" (p. 31). He maintains that adopting such an approach in analysing students' texts can help them understand how to organize a text and how to avoid some of the significant faults in their text organization, some of which he attributes to L1 interference and some to lack of experience in text organization. For classroom application, he suggests instructors to form small groups of students ask them to analyse texts they have read. Subsequently, the groups can be asked to compare their analyses and discuss differences among them. He believes this technique improves students' sense of text organization.

Likewise, Nasiri (2012) points out that a number of studies conducted on cross-cultural differences between academic research articles revealed the challenges non-native speakers confront in their attempt to learn how to write academically in a foreign language. He suggests many non-native speakers are unable to understand the requirements of academic writing in the target language due to lack of training and knowledge of academic writing. He believes that language, identity and culture of the authors are inextricably intertwined and that they inevitably provide insights into the identity of an author of a written text. Thus, native language, identity and cultural values of a writer are reflected on what s/he writes. Therefore, he suggests EAP instructors should be aware of cultural differences and try to teach in a way that the intercultural differences are minimized and students get familiar with the style of academic writing in the English culture.

Yue (2009) Chinese and English display distinct rhetoric styles due to different cultural backgrounds and traditions of the communities speak the languages. Namely,

attitudes of "self" greatly differ in the compositions written in Chinese and those in English. According to Connor (1996), directness and explicitness are emphasized in the western style writing. Gregg (1983) notes that writers from western cultures endeavoured to support their own ideas or inventions and the readers raised in this culture need more supportive information that enables them to continue their own inquiries. Chinese concept of self makes it difficult for Chinese writers to express their own opinion so directly in their composition (Yue, 2009). Connor (1996) holds that writers with Chinese as native language tend to pay special attention to history, tradition, authority and so, "they often cite references from historical and religious classical texts or proverbs instead of expressing themselves" (p. 95). The way Matalene (1985) explain approximately two decades ago account for Connor's point. He argues that Chinese writers extensively use maxims, proverbs and analogies in their writing due to the fact that all great religions in China de-emphasize the importance of the individual against collectivism. Yue (2009) confirms the idea noting that "Confucianism ordinates the group to individual; Taoism subordinates the group to the nature; Buddhism denies the self-exist and Marxism rejects bourgeois individualism" (p. 95).

Hinds (1987) holds that languages differ in the responsibility assigned to readers and writers. He proposes, for example, readers are responsible for understanding the writer's position or intention in Japanese whereas it is the writer who is responsible for making "clear "and "well-organized" statements for effective communication in English. He maintains that transition statements appear less frequently in Chinese compositions than those in English and that English writers extensively use them so as to make their points explicitly. According to Yue (2009), Chinese-speaking learners of English use inappropriate rhetorical features in their compositions in English, which might be contributed to L1 transfer. Reminding that the biggest problem stems from the differences the way people think and their culture, he suggests educators to cultivate the students' logical and critical thinking skills by integrating cultural awareness into the teaching and learning process (e.g., using conventional English paragraphs and composition structure in L2 writing instruction).

McCrostie (2008) investigated the degree of writer/ reader visibility in argumentative academic essays produced by university-level Japanese learners of English and found that they include a much higher degree of spoken language characteristics than the ones written by native speakers of English. In the meanwhile, he

reports that those characteristics were significantly decreased in the essays of second-year students. Namely, the second-year students were observed to avoid overt author presence in their writing when compared to the novice students; first and second person pronouns, for instance, were found much less frequently in the essays written by the second-year students than the ones written by the novice students, indicating one-year instruction appears to play a prominent role in their writing.

Hyland (2002) explored the notion of writer identity as the way it is expressed through self-reference in the written productions of L2 undergraduate students in Hong Kong. He mainly investigated the frequency and role of first person pronouns and the determiners in the final year reports of these students and compared the findings with a corpus consisting of journal articles written by native speakers of English. He revealed that L2 students significantly underused authorial pronouns and determiners and they were not inclined to accept clear connotations of authority conveyed through authorial pronouns. Rhetorical identity is influenced by the writer's background and this becomes more intricate for students familiar with intellectual traditions which may be very different from those practised in English academic contexts (Hyland, 2002, p. 1111). He drew to the conclusion that a conscious usage of authorial identity is encouraged in Anglo-American academic conventions with the aim of managing the reader's awareness of the author's role and viewpoint whereas L2 writers with different cultural background "might be reluctant to promote an individual self". All in all, he suggests teachers need to raise consciousness in their students by presenting rhetorical options and assisting them understand the effects of manipulating these options for interactional purposes so that they could gain control their writing and effectively use academic genres in a second language.

Johns and Paz (1997) argue that students often fail to write effective summaries is caused either by lack of appropriate strategies for writing such kind of summaries or "they are taught relatively inflexible strategies inappropriate to the genre they are reading (p. 33). They believe that the interaction of language, text structure and function should be taught in all reading and writing classes. Inspired by the theories about science course books by Johns and Davies (1983), they selected and analysed texts for student reading and summarization, which enabled them to gain a better understanding of "the relationships between text function, content and macrostructure and to select appropriate texts for summarizing" (p. 41). Subsequently, they divided students into groups and provided each group with a text from a particular function type, with a

diagram or illustration to complete, and with a chart including a list of information structure constituents of that function. They report the students performed better in relating words to visual elements than in paraphrasing or summarizing. At this point, they criticize that innovative theories and approaches to text analysis and writing are not adequately incorporated into curricula or course books in ESL/EFL classes. Thus, students are likely to gain more competence in writing when the theories and approaches in concern are put into practice in their writing classes.

In her study based on macro- and micro-level analysis of student summaries of a fiction story, Kramsch (1997) examined the way teacher and students discuss a text and they talk about each others' interpretations of the text. By doing so, she aimed to make students, who are from East-Asian and Latin American countries and who attend a university in the USA, personally interested in the relationship of words to meaning and to elicit multiple perspectives on the original text. She stresses that the primary source of meaning was not authorial intention and the quality of student summaries was not focused in the process. What was inquired through the study was the richness of the interpretation readers construct; namely, the purpose was identified as "to jointly construct a meaning for the story, based on the individual versions of the original, and to make sense of the conscious or unconscious "choices" made by the students" (p. 51). She maintains that the teacher is not expected to correct or provide positive feedback for students' summaries; his/ her role is, rather, "to probe, draw possible analogies and contrasts, suggest possible interpretations, reflect choice on words" (p. 61). She suggests students may be asked to find titles to each other's summaries or to rewrite a given summary with the purpose of encouraging them to relate the words to the meaning proposed in it. She concludes both teachers and students can become conscious of the way they co-construct models of understanding through such kind of task.

As cited in Salager-Meyer (1997), Stubbs argues that all sentences encode a point of view and that academic texts are no different in containing the author's presence. Along with this line, Salager-Meyer contends that "academic writing cannot be considered as a series of impersonal statements of facts which add up to the truth" (p. 105). She (1993, 1994) revealed medical editorials and review articles contain more hedging devices than research papers and case reports. She concludes that discourse become more hedged when generalizations become stronger and more universal. She is critical of the fact that hedges in academic writing are heavily ignored in instructional

materials for non-native speakers of English. She suggests reading and writing exercises that "enable learners to use their limited linguistic resources to achieve greater delicacy of meaning" should be employed in the classroom in order to "empower these academics to express (and recognize) doubt where there is no certainty, to know how and when to mediate their claims, and to use these techniques successfully" (pp. 114-115). She also recommends "using authentic and challenging materials from their own field of study", which she describes "one of the best ways of developing students' academic meaning potential" (p. 115).

Nga (2009) anticipates that English-major students in Vietnam display poor competence in using academic style in their writing might be partly stemmed from "teachers' improper emphasis on writing Academic English" and "students' vague knowledge about it". She maintains most of second-year students of English are unable to clarify the distinction between spoken and written academic style or identify the features of these styles. She, therefore, presumes that they usually fail to produce a formal piece of writing such as a paragraph or an essay in their second and third years at the university. She also assumes this might be triggered by the fact that the students in concern are not presented the knowledge of academic style earlier than their seventh semester at university. Accordingly, she postulates it should be included in the curricula of the second semester so that the students could extensively practice and improve their academic writing skills.

Spack (1988) asserts that writing teachers should teach the writing process to teach writing and that what they need in order to accomplish is to know how to write. She believes that academic writing can be learned by observing the process through which professional academic writers produce texts or by attending programmes with an objective to teach non-native graduate students how to write a research paper. She emphasizes that lecturers should avoid assigning students writing tasks they themselves can nor master and that they should not confine their assignments to "prescribed and rule-governed" tasks. Instead, she offers them to draw on their own knowledge and abilities as they strengthen and expand the knowledge and abilities of their students. Finally, she argues that the most effective way of teaching academic writing is to adopt general strategies beyond disciplines. Thonney (2011) concurs with this view suggesting that general knowledge about academic writing should be provided to the students from the very start of their higher education disregarding the varieties across discourse communities so that they could improve their academic writing skills. She

contends that this might be facilitated when students read authentic academic texts from various disciplines, when they become aware of varieties in academic writing, when they infer and practice academic writing principles and when they see that academic writing is a dynamic phenomenon. So, she believes, the role of instructors is to assist them develop consciousness of these issues and create environments where they have the opportunity to practice their writing skills.

Hyland (2002) highlights that a great number of survey results have revealed writing tasks assigned to students at university tend to be specific to discipline and related to educational level. He draws attention to the study carried out by Bridgeman and Carlson (1984) who found that in postgraduate programmes, engineers are more expected to describe charts while students of business studies are required to compare ideas and take a position. Likewise, another study conducted by Wallace (1995) indicates that in undergraduate classes, lab reports are common in Chemistry while programme documentation and article surveys are paramount in Science and Maths, respectively. Clyne (1999), on the other hand, claims essay writing plays such a central role in evaluation of students' performance that it is virtually impossible to imagine assessment without it in the education system of Australia. He remarks this is the case more frequented in English-speaking countries where the way students provide answers in examinations and class exercises is of great significance than continental European countries where the focus is on the evaluation of specific and general knowledge students have attained in a subject. He maintains, therefore, techniques of essay writing are emphasized not only in the first-language classes but also in other subjects at schools in English-speaking countries.

Adel (2010) outlines the traditional topics taught in teaching academic writing, which are illustrated in Table 1.

Table 1

Traditional Topics in Academic Writing

Effective Persuasion	Higher-level rhetorical stages
Audience Analysis/ Awareness	
Using Sources and Citation	How to use and refer to other sources, including avoiding plagiarism
Text Types	e.g. argumentative, expository
Genres	e.g. research papers, business letters
Coherence and Cohesion	e.g. development of ideas, structuring of information, paragraph writing, use of connectives
Organization	e.g. types of information to put in various sections; strategies for planning the text such as mind-mapping
Level of Formality/ Style	
Rhetorical Actions	e.g. paraphrasing, concluding, introducing topic
Drafting, Proofreading and Revising	
Vocabulary	
Lexico-grammar and Phraseology	
Formatting	e.g. how to write a bibliography following The Chicago Manual of Style

Souece: Retrieved from Adel, 2010, p. 41.

She explains the topics at the bottom of the table such as lexico-grammar, collocation and academic vocabulary are essential features of academic writing and that they are mostly covered in the corpus classroom. However, it is a fact that these are not the only features of writing and that many other ones cannot be taught using direct corpus-based approaches during writing classes. Nonetheless, she believes that it is possible to cover them in such a classroom and that the topics in concern can be presented in a more systematic way.

2.1.2. Corpus-based Academic Writing

Adel (2010) advocates that corpus-based courses on academic writing are strongly needed due to the fact that non-native speakers lack native-speaker intuitions and have a predisposition to transfer writing patterns in their L1 to a rhetorically different L2 in an inappropriate manner. However, she asserts, they are still in their infancies and there is an insufficient number of examples of teaching academic writing through corpora, particularly above the level of vocabulary and collocation. She proposes seven challenges to teaching writing in the corpus classroom as *lack of corpus availability, the*

corpus as a maze, drowning in data, interpretation, evaluation, decontextualized data and *focus on surface form.* What she means by *lack of corpus availability* is that corpora of academic writing, which are generally not available, frequently consists of text fragments rather than full texts, making it impossible to study many patterns in writing. She is of the opinion that this might be overcome by making corpora generally available and with the resolution of copyright issues. As a second challenge, she maintains we might have difficulty in deciding what to look for in a corpus, which could be avoided in teacher-guided settings and clearly defined tasks related to corpus searches. The third challenge is drowning in data, which might occur when the corpus includes a large amount of data. In order to meet the challenge in concern, she offers building random sampling techniques into search tools, which she believes would help students "access linguistic data in manageable quantities and in a statistically sound way". She suggests that teachers could also assign different samples to different students if the above-mentioned technical solution is not available or it does not work. Thereby, she believes, the total number of analysed examples increase. The fourth challenge has to do more with interpretation of the corpus. She argues it is hard to figure out why a linguistic feature is used in the way it is since many factors such as grammatical and sociolinguistic ones influence language use. However, she contends that this could be turned into an advantage especially for advanced students showing them that data can be interpreted in different ways. The first measure she suggests for the solution to this problem is that students should be made familiar with the texts included in the corpus. Thus, they should be informed about the authors of the texts as well as when and how those texts were produced. Comparison is the other method she offers to make it easier for the students to understand linguistic features found in a corpus. Evaluation is the fifth concern that could make it hard to use corpora when it does not present models or patterns for students to follow. She explains this with the example of data including the first person pronoun many times. In such a case, students might consider it as an acceptable feature in academic writing and produce their texts accordingly. As a solution to this problem, she argues that instructors should be both descriptive linguists and language policy makers. Namely, they should be able to provide students with recommendations as well as good arguments in the writing sessions and they should use complementary materials in these sessions. Decontextualized data refers to the fact that data concordance lines considerably lack the social context, which involves elements such as *communicative context, typical reader-writer roles, cultural values* and

intertextual knowledge, which does not constitute a drawback in teaching writing since a particular corpus can not be expected to convey the communicative purpose of a text or to socialize students into the academic world. To overcome the problem in concern, she suggests that students should be gained other perspectives on the target writing and that corpus-based writing should be considered as a method in academic writing that should be accompanied by other methods and techniques. She informs that focus on surface form is the last and most severe challenge owing to the fact that computer-assisted methods confine us to work exclusively at the word or phrase level. In her opinion, annotation of functional features might help students understand the data beyond the surface level.

She conducted a small-scale experiment in the writing instruction where she offered corpus consultation to improve writing skills of non-native graduate students with high proficiency level in English. The course was based on rhetorical functions and phraseology in academic writing. The students were asked to examine MICUSP, a corpus consisting of highly proficient student writing, via WS Tools (Scott, 2004) to find out what academic writers use when they (a) give an example, (b) refer to other texts or researchers, (c) introduce the topic and (d) start the Conclusion section. At the end of the analyses, which were realized in a controlled setting, students revealed to be enthusiastic about inductive learning about the use of computer. She also reports that the corpus session was nicely supported by instruction from the regular writing instructors, during which note-taking on hand-outs was encouraged.

In another study conducted by Chang and Swales (1999), informal elements in texts were investigated through handbooks as well as corpus and interview data from manual writers, journal editors and authors of research articles. In the first place, they obtained a list of rules of appropriateness for English scholarly writing most frequently mentioned in style manuals and writing guidebooks. They examined ten papers written in the disciplines of Statistics, Linguistics and Philosophy and published in the same journal with respect to ten informal grammatical features in accordance with the rules in the list. The results have shown that the mostly employed features are *first person pronoun*, *unattended this* and *sentence-initial conjunctions* and that *final prepositions, sentence fragments* and *exclamations* occurred with relatively low frequencies. It also reveals that philosophers are the ones whose works include approximately all the informal features. That is, they display a more informal and interactive writing style when compared to those of the linguists and statisticians. The researchers conclude that

linguistic features might be accounted for different contemporary stylistic preferences in the disciplines in concern. Throughout the study, L2 graduate students were asked whether informal features influence academic writing in English in a positive or negative way. The majority report that greater informality entails greater flexibility, and that they do not want to be concerned with how to include these features into their writing stating that they already find it challenging to learn the rules of formal academic English.

Hyland and Milton (1997) investigated the degree of caution and certainty in EAP writers' works in a corpus of one million words including 1700 UK and Hong Kong students' exam scripts. They also aimed to explore the ways students attending Hong Kong University present assertions in their writing. The results of the study have indicated striking similarities between two sets of data with respect to the overall frequencies of caution and certainty items. The groups seem identical in total frequencies of epistemic *will, may, would* and *always*; however, they differ in the frequencies of the individual items in concern. Namely, *epistemic would* and *epistemic will*, both of which can be referred to express present and future probabilities, were favoured by different groups; that is, it is observed that the former appeared as twice often in the works of non-native group whereas the latter was as twice frequent in those of the native group. According to the researchers, the result in concern indicates conceptual differences between two groups; thus, L2 writers preferred confident prediction while native writers tend to use more tentative expressions. Another finding of their study is that non-native writers with higher proficiency approximate native speaker usage more closely than those with lower proficiency. Namely, certainty markers were more frequented in the scripts of the latter while those of former include probability and possibility devices in higher proportions, which leads the researchers to the conclusion that the higher the writers' proficiency, the more tentative expressions they use in modifying their statements. On the other hand, essays written by non-native writers revealed relatively fewer epistemic clusters to emphasize the strength of the accompanying proposition. The data obtained from the essays indicate that most of the writers in concern are not proficient enough to combining epistemic clusters correctly. Overall, the study shows that both groups native and non-native groups of writers rely on modal verbs and adverbs in presenting their assertion. Finally, it points out that non-native writers have a tendency to employ syntactically simpler constructions, a more limited range of devices and statements with stronger commitment.

2.1.2.1. Stance and Stance Devices

Stance is generally understood to have to do with the methods, linguistic and other, by which interactants create and signal relationship with the propositions they utter and with the people they interact with (Jaffe, 2009). It is the point of intersection in discourse between the ideational and interpersonal functions of language (Faircough 1992, p. 160). According to Biber and Finegan (1989), it refers to "the lexical and grammatical expression of attitudes, feelings, judgments, or commitment concerning the propositional content of a message" (p. 124). It is defined as "a socially recognized disposition" (Ochs, 1990). Hyland (2005b) describes it as 'the ways academics annotate their texts to comment on the possible accuracy or creditability of a claim, the extent they want to commit themselves to it, or the attitude they want to convey to an entity, a proposition or the reader. ' He proposes the Model of Interaction in Academic Discourse consisting four major elements as *hedges, boosters, attitude markers* and *self mentions*, which are outlined in.

Table 2

Model of Interaction in Academic Discourse

Element	**Definition**
Hedges	Devices that indicate the writer's decision to withhold complete commitment to a proposition, allowing information to be presented as an opinion rather than accredited fact
Boosters	Devices that allow writers to express their certainty in what they say and to mark involvement with the topic and solidarity with their audience
Attitude Markers	Devices that indicate the writer's affective attitude to propositions, conveying surprise, agreement, importance, frustration, and so on, rather than commitment
Self- Mentions	Devices that refer to the use of first person pronouns and possessive adjectives to present propositional, affective and interpersonal information

Source: Adapted from Hyland, 2005b, p. 5.

Hyland (1995) states "hedging essentially represents an absence of certainty and is used to describe any linguistic item or strategy employed to indicate either a lack of commitment to the truth value of an accompanying proposition or a desire not to express that commitment categorically". He defines hedges as "the means by which writers can present a proposition as an opinion rather than a fact: items are only hedges

in their epistemic sense, and only when they mark uncertainty" (1998, p. 5). According to Hyland (1998), boosters are the items that "allow writers to express conviction and assert a proposition with confidence" and they "mark involvement and solidarity with an audience, stressing shared information, group membership and direct engagement with readers". He remarks that hedges and boosters are communicative strategies which are used to increase or reduce the force of statements and that they contribute to an appropriate rhetorical and interactive tenor in academic discourse. As for attitude markers, Hyland (2005b) suggests that they represent writer's affective attitude to propositions and convey "surprise, agreement, importance, obligation and frustration" rather than comment on its status. Self-mention refers to the degree of explicit author presence in the text measured by the frequency of first- person pronouns and possessive adjectives (Hyland, 2001). He advocates that "writers cannot avoid projecting an impression of themselves and how they stand in relation to their arguments, their community and their readers".

Using the term "evaluation" for "stance", Hunston and Thompson (2000) assign three basic functions to it: "to express the writer's opinion, and in doing so to reflect the value system of that person and their community", "to construct and maintain relations between the writer and the reader", and "to organize the discourse". Martin (2000), on the other hand, uses "appraisal" referring to the semantic resources that negotiate how authorial opinion is expressed and similarly. Conrad and Biber (2000) classify stance into three semantic categories as *epistemic, attitudinal* and *style.* Epistemic stance, in general, represents how certain or reliable the writer's proposition is, attitudinal stance reports personal attitudes or feelings and style stance describes how the information is being presented (MacIntyre, 1999). In the following sections, these categories will be covered in detail.

Biber (2006) suggests that epistemic meanings such as *certainty* and *likelihood* can be expressed by *stance adverbs, adjectives* and *nouns controlling that*-clauses and *verbs* and *adjectives controlling to*-clauses, as exemplified in Table 3.

Table 3

Epistemic (Certainty and Likelihood) Stance Devices

Adverb	That's **certainly** a major cause of absenteeism. Now these schools might **possibly** be able to afford this special fund
Verb + *that*-clause	Lashly **found** [that retention of a maze habit is directly proportional to the amount of brain tissue remaining after extirpation]. Therefore there is no reason to **suspect** [that the two means are different].
Adjective + *that*-clause	It is **evident** [that by 1925 Gandhi had fully worked out his approach to, and explanation of, the text]. If there is inflation, and the cost of the automobile goes up, then it's quite **likely** [that the interest rate will also go up].
Noun + *that*-clause	Supporting the experimenters' quantitative results was their general **observation** [that the town dwellers were noticeably more friendly and less suspicious than the city dwellers]. There's some **indication** [that prenatal development has an influence on lifespan development].
Verb + *to*-clause	Our best evidence **seems** [to suggest that if you control for size, the more international you are, the less risk you have].
Adjective + *to*-clause	These "control failures" are **certain** [to happen occasionally].

Source: Biber, 2006, p. 94

He remarks that stance devices can also be used to express attitude and evaluations. Table 4 illustrates the devices of such kind.

Table 4

Attitude and Evaluation Stance Devices

Adverb	**Amazingly** the archivists at the national archives are under an extreme amount of pressure.
Verb + *that*-clause	I really **hope** [that it doesn't take that long to put the whole thing together].
Adjective + *that*-clause	We are **pleased** [that you are considering graduate study at Northern Arizona University].
Adjective + *to*-clause	Jane's question is not **appropriate** [to discuss in marketing class].

Source: Adapted from Biber, 2006, p. 94

Stance can be expressed to differing extents through grammatical devices, value-laden word choice and paralinguistic devices (Biber 2006, p. 87-88). He contends that it can be conveyed via grammatical devices whereby a particular grammatical structure is used to express stance in terms of some other proposition. He (2006) analysed

grammatical stance devices in three major groups: modal and semi-modal verbs (*should, might, can* and etc.), stance adverbs (*roughly, certainly, indeed* and etc.) and complement clauses controlled by stance verbs (*demonstrate, notice, reckon* etc.), by stance adjectives (*confident, unlikely, essential* and etc.) and stance nouns (*conclusion, opinion, reason* and etc.). Table 5 is intended to exemplify these devices.

Table 5

Common Grammatical Stance Devices

Modal verbs and semi-modals	We *should* eat something. We *had better* go now. We *must* hand in our assignments by Thursday.
Stance adverbial	They are *actually* expected to practice the strategies that are indispensable interpersonal communication. *In fact*, the narrator shares only a little about the house's or the garden's organisation or size as well as about the family relationship. (Dontcheva, 2009)
CCs controlled by stance adjectives and nouns	It is *essential* that we all be at the meeting at 2 pm. I am of the *opinion* that unemployment is the biggest problem of the country.

Source: Adapted from Biber, 2006, p. 92

Biber (2006) states that stance can also be conveyed through value-laden word choice such as *hate*, *love* and *dislike*.

e.g. I *love* this time of the year.
I *hate* the way he talks to me.

He advocates that evaluative word choice and grammatical stance marking are fundamentally different in that the former requires one particular proposition while the latter involves expression of stance relevant to some other proposition.

1a. I *love* chocolate.
1b. I am *afraid* we have run out of sugar.

Using an evaluative word (*love* in this case), the speaker/author intends to convey only one proposition, as illustrated in 1a whereas s/he communicates a proposition as well as his/her personal stance towards the proposition in concern.

According to Biber (2006), another way of conveying stance is to use paralinguistic devices, which includes pitch, intensity and duration in speech, and italics, bold face and underlining or using speaking adverbs such as *angrily* and *bitterly* in writing. He asserts that value-laden word choice and paralinguistic devices are not so explicit as grammatical stance devices since they do not overtly express an evaluative frame for other propositions and that grammatical stance devices can be perceived as direct expression of speaker/writer stance. He highlights that they can be attributed to the speaker/author himself/herself and to a 2nd or 3rd person. The following exemplify the stance structures with a 1st person subject.

2a.	I	*know*	that she is going to be here by 5.
	1st person pronoun	stance verb	*that*-clause

2b.	We	are *afraid*	that you have failed the test.
	1st person pronoun	stance adj.	*that*-clause

Biber (2006) notes that stance structures including a 1st person object or a 1st person possessive pronoun can also be attributed to the speaker/author, as in 3a, 3b and 3c:

3a.	It could be	*beneficial* for	me	to attend the annual meeting.
		Stance adj.	1st person object pronoun	*to*-clause

3b.	It would be	*surprising* to	me	that he arrive on time.
		Stance adj.	1st person object pronoun	*that*-clause

3c.	It results from	our	failure	to read the instructions carefully.
		1st person possessive pronoun	stance	*to*-clause

Biber (2006) holds that stance expressions attributed to a 2nd or 3rd person express some sort of attitude and evaluation; nonetheless, they do not reflect the personal stance of the speaker/author in all cases, as in the following:

e.g. Do *you think* we will get this on the test?
Yeah I know *you want* to condense it down to a title.
She didn't realize that you are a stranger.
They needed to rebuild the entire government system.

(Biber, 2006, p. 91)

He proposes that grammatical stance devices typically consist of two major elements: the stance marker and a proposition framed by the marker in concern and that they can differ in structure; namely, they can be words (e.g. *undeniably*), phrases (e.g. *most probably*), or clauses (e.g. *as mentioned previously*).

Table 6 illustrates three major structural categories suggested by Biber (2006) as well as common controlling words each category contains.

Table 6

Grammatical Stance Devices: Modal Verbs (and semi-modals), Stance Adverbs and Stance Complement Clauses

1. Modal and semi-modal verbs (See LGSWE, pp. 483ff.)
   - -possibility/ permission/ ability: *can, could, may, might*
   - -necessity/ obligation: *must, should, (had) better, have to, got to, ought to*
   - -prediction/ volition: *will, would, shall, be going to*

2. Stance Adverbs (See LGSWE, pp. 557–558; 853–874)
   - -Epistemic:
     - Certainty: *actually, always, certainly, definitely, indeed, inevitably, in fact, never, of course, obviously, really, undoubtedly, without doubt, no doubt*
     - Likelihood: *apparently, evidently, kind of, in most cases/instances, perhaps, possibly, predictably, probably, roughly, sort of, maybe*
   - -Attitude: *amazingly, astonishingly, conveniently, curiously, hopefully, even worse, fortunately, importantly, ironically, rightly, sadly, surprisingly, unfortunately*
   - -Style: *according to, confidentially, frankly, generally, honestly, mainly, technically, truthfully, typically, reportedly, primarily, usually*

3. Complement clauses controlled by stance verbs, adjectives, or nouns
   - 3.1 Stance complement clauses controlled by verbs
     - 3.1.a Stance verb + *that*-clause (See LGSWE, pp. 661–670)
       - -Epistemic verbs:
         - Certainty: *conclude, demonstrate, determine, discover, find, know, learn, mean, notice, observe, prove, realize, recognize, remember, see, show, understand*
         - Likelihood: *assume, believe, doubt, gather, guess, hypothesize, imagine, predict, presuppose, presume, reckon, seem, speculate, suppose, suspect, think*
       - -Attitude verbs: *agree, anticipate, complain, concede, ensure, expect, fear, feel, forget, hope, mind, prefer, pretend, require, wish, worry*
       - -Speech act and other com. verbs: *announce, argue, assert, claim, contend, declare, tell, emphasize, explain, imply, insist, mention, promise,*
         propose, recommend, remark, respond, say, state, suggest

(Table 6. Continued)

	3.1.b	Stance verb + *to*-clause (See LGSWE, pp. 693–715)	
		-Probability verbs:	*appear, happen, seem, tend*
		-Cognition/ perception verbs:	*assume, believe, consider, expect, find, forget, imagine, judge, know, learn, presume, pretend, remember, suppose*
		-Desire/ intention/ decision verbs:	*agree, choose, decide, hate, hesitate, hope, intend, like, love, mean, need, plan, prefer, prepare, refuse, want, wish*
		-Verbs of causation/ modality/ effort:	*allow, attempt, enable, encourage, fail, help, instruct, manage, oblige, order, permit, persuade, prompt, require, seek, try*
		-Speech act and other com. verbs:	*ask, claim, invite, promise, remind, request, be said, show, teach, tell, urge, warn*
3.2.		Stance complement clauses controlled by stance adjectives	
	3.2.a	Stance adjective + *that*-clause (See LGSWE, pp. 671–674)	
		-Epistemic adjectives:	
		Certainty:	*apparent, certain, clear, confident, convinced, correct, evident, false, impossible, inevitable, obvious, positive, right, sure, true, well-known*
		Likelihood:	*doubtful, likely, possible, probable*
		-Attitude/ emotion adjectives:	*afraid, amazed, aware, concerned, disappointed, encouraged, glad, happy, hopeful, pleased, shocked, surprised, worried*
		-Evaluation adjectives:	*amazing, appropriate, conceivable, crucial, essential, fortunate, imperative, inconceivable, incredible, interesting, lucky, necessary, nice, noteworthy, odd, ridiculous, strange, surprising, unacceptable, unfortunate*
	3.2.b	Stance adjective + *to*-clause (See LGSWE, pp. 716–721)	
		Epistemic adjectives:	
		- Certainty/ likelihood:	*apt, certain, due, guaranteed, liable, likely, prone, unlikely, sure*
		-Attitude/ emotion adjectives:	*afraid, ashamed, disappointed, embarrassed, glad, happy, pleased, proud, puzzled, relieved, sorry, surprised, worried*
		-Evaluation adjectives:	(*in*)*appropriate, bad/worse, good/better/best, convenient, essential, important, interesting, necessary, nice, reasonable, silly, smart, stupid, surprising, useful, useless, unreasonable, wise, wrong*

(Table 6. Continued)

		-Ability/ willingness adjectives:	(*un*)*able, anxious, careful, determined, eager, eligible, hesitant, inclined, obliged, prepared, ready, reluctant,* (*un*)*willing*
		- Ease/difficulty adjectives:	*difficult, easier, easy, hard*, (*im*)*possible, tough*
3.3.	Stance complement clauses controlled by nouns		
	3.3.a	Stance noun + *that*-clause (See LGSWE, pp. 648–651)	
		-Epistemic nouns:	
		Certainty:	*assertion, conclusion, conviction, discovery, doubt, fact, knowledge, observation, principle, realization, result, statement*
		Likelihood:	*assumption, belief , claim, contention, feeling, hypothesis, idea, implication, impression, notion, opinion, possibility, presumption, suggestion*
		-Attitude/ perspective nouns:	*grounds, hope, reason, view, thought*
		-Communication (non-factual) nouns:	*comment, news, proposal, proposition, remark, report, requirement*
	3.3.b.	Stance noun + *to*-clause (See LGSWE, pp. 652–653)	*agreement, decision, desire, failure, inclination, intention, obligation, opportunity, plan, promise, proposal, reluctance, responsibility, right, tendency, threat, wish, willingness*

Source: Biber, 2006, p. 92-93. (Cont'd)

Several studies have been conducted on stance and stance devices in both spoken and written registers. In a study carried out by Ventola and Maureanen (1990), Finnish-speaking learners of English and English native speakers have been compared in terms of ESDs and it has been revealed that the former group display a tendency to use relatively a small number of safe expressions of epistemic possibility. Hyland and Tse (2005) investigated the evaluative *that* in published articles and dissertations authored by ESL students and they have concluded that it is extensively used in the abstracts. In a study conducted by Hyland and Milton (1997), it has been revealed that Chinese-speaking secondary school students are more assertive and committed with their statements than native speakers of English. Likewise, Gabrielatos and McEnery (2005) analysed MA dissertations written by students at the Department of Linguistics and English Language, Lancaster University. They compared native and advanced non-native speakers of English with respect to the use of epistemic modality and they report that native speakers of English employ epistemic modals, epistemic adjectives and epistemic adverbs much more frequently than non-native speakers. McEnery and Kifle

(2002), on the other hand, have found that weak epistemic modals appear more frequently in Eritrean secondary school students than native speaker students. Sultan (2011) investigated discussion sections of research articles written by Arabic-speaking linguists and native speakers of English regarding interactive and interactional metadiscourse markers, which were previously proposed and defined by Hyland (2000) as "the expressions used to organize propositional information in ways that a projected target audience is likely to find coherent and convincing" and "the ways writers conduct interaction by intruding and commenting on their message", respectively. He has revealed that Arabic-speaking linguists use both interactive and interactional metadiscourse markers more intensively than native speakers of English and that both groups have a predisposition to employ interactive resources more widely than interactional ones.

Biber (2006) conducted a corpus-based study on grammatical stance devices. He mainly analysed these devices in three major structural categories as *modal verbs* (and *semi-modals*), *stance adverbs* and *stance complement clauses*. The data of his study is reported to have included all controlling words that commonly occurred in the LSWE Corpus (taken from the LGSWE) and any other words in the T2k-SWAL Corpus. Grammatical stance devices analysed in the study were taken from both spoken registers (e.g. classroom teaching, class management, office hours and etc.), and written registers (e.g. textbooks, course packs, institutional writing and etc.). He has found that modal verbs are used more frequently than stance adverbs and stance complement clauses both in the spoken and written registers and that stance adverbs and stance complement clauses appear more frequently in the spoken registers than in the written registers. He has reported several differences in the use of particular modals across registers. Namely, he has revealed that prediction/ volition modals (e.g. *will, would*) are the most common modals approximately in all spoken and written registers and that the necessity/ obligation modals (e.g. *must, should*) constitute the modal class found at the least frequency. In addition, the possibility modal *can*, the obligation modal *have to*, the prediction modals *will* and *shall,* and the prediction semi-modal *be going to* appear much more frequently in the spoken registers than in the written registers. On the other hand, the possibility modal *may*, the obligation modal *must* and *should* are used more frequently in the written registers than in the spoken registers. He reports that the prediction modal *would* used in the classroom management appear with a first person subject (*I*) and have a particular directive function, as illustrated below:

e.g. The instructor: I'*d* like you to bring some food.
I'*d* like for you to do problem 1-A.
I'*d* like you to review your quizzes, your mid-terms immediately, and let me know if we did something wrong.
(Biber, 2006, p. 98)

He calls attention to the fact that the prediction modal *will* can also be used with a directive function and that the underlying force is less polite and more direct in such a case, as in the following:~

e.g. The instructor: Um, on the tenth you *will* do chapters nine and ten. Um. After, well, let me do the days here. So, Monday, on the first of March, these people *will* be here. And the first five people *will* speak. And in this order. Andy *will* be speaking first.
(Biber, 2006, p. 98)

What he has revealed with respect to the distribution of the modal classes across registers is that the possibility modals (e.g. *can, might*) and the prediction modals (e.g. *will, would*) are almost equally common in the service encounters. He has founded that the most common class occurring in the textbooks are the possibility/ permission/ ability modals and that can and may display a tendency to be used with inanimate subjects and they differ in their functions; namely, former conveys both ability and possibility meanings whereas the latter, in general, communicates merely possibility meanings, as exemplified in the following:

e.g. International negotiation *can* also help to avoid a trade war. It *may* therefore be in the interest of a politician to adopt positions that are against the interest of the typical voter.
(Biber, 2006, p. 98)

He reports that the possibility modal *may* is used along with the possibility/ ability meanings of *can* in some institutional writing which he considers to be prepared in order to educate and inform students rather than set rules for them, as in:

e.g. A person who has HIV *may* be susceptible to diseases most healthy

> people can resist. People infected with HIV *can* look and feel healthy and *may* not even know that they are infected. Even though they don't look or feel sick, they *can* infect others…
>
> (Biber, 2006, p. 101)

The obligation modal must, which appears more frequently in the written registers than in the spoken registers, is reported to carry meanings of personal obligation and logical necessity, as in:

> e.g. We *must* be careful to avoid several logical pitfalls. (textbook; per. obligation)
> It *must* have something to do with the government. (study group; log. nec.)
>
> (Biber, 2006, p. 101)

He has analysed stance adverbs with regard to their semantic classes (certainty, likelihood, attitude and style) and to the registers in which they appear and concluded that certainty adverbs (e.g. *actually, indeed*) are widely used in all spoken registers such as classroom teaching, class managements and labs and underlined that they are used mostly by teachers (rather than students) to identify information as factual and beyond dispute. The adverbs in concern are reported to be used for the purpose of emphasizing the activities students and the instructor are expected to attend, as in:

> e.g. But, what I *really* have to do though is to keep the total time frame for each person, almost exact, to fifteen minutes.
> Well, we need to equal sign here so I can *actually* look that up in the table.
>
> (Biber, 2006, p. 105)

As for likelihood adverbs (e.g., *probably* and *maybe*), he maintains that they are used to indicate events and actions that are likely to occur (or should be done) and that they are especially employed in office hours to suggest actions and events that would be desirable or likely, as in:

> e.g. OK what I'll *probably* have to do then is just have you drop this one.

> And uh um I guess *probably* the best thing to do would have been for me to go over and meet everybody over there over Christmas time but I'm not going to get a chance to do that so, in case that doesn't happen what I was going to say *maybe* if we can't all travel let's *maybe* try a conference call. (Biber, 2006, p. 105)

He concludes that style adverbs and adverbial phrases (e.g. *reportedly, according to*) are employed to identify the source of a claim, as in:

e.g. *According to* Wallerstein, this system works to the advantage of core nations and the disadvantage of periphery nations.
(Biber, 2006, p. 105)

Another finding of his study is that the stance verbs controlling *that*-clause differ across registers; namely, they appear more frequently in the spoken registers than in the written registers. The study also reveals that certainty verbs controlling *that*-clauses occur more common in the teacher-centred academic registers while the likelihood verbs controlling the clauses in concern are seen more frequently in the student-centred academic registers, as in:

e.g. Instructor: We *recognize* that it is a real error [unclear words] because he pursues the ideal out of this world...
Student: …. I *guess* we *thought* that you still had to have them even though like we're only looking at four bits, we *thought* you had to fill like the eight bits.
(Biber, 2006, p. 107)

He reveals that students use the communication verbs in order to identify the source of information, as in:

e.g. Actually he *said* they can be either one but they prefer to be called justices.

> And then, to figure out our T. score, our actual T. score, you go to the table, it *says* the sample size is nine.
>
> (Biber, 2006, p. 108)

To-clauses framed by stance verbs are reported to have occurred at almost equal frequencies in the spoken and the written registers. The study has shown, for instance, that verbs of desire (e.g. *want, like*) controlling *to*-clauses are the most common class, particularly in the spoken registers such as the class management, office hours and service encounters. Biber (2006) notifies that they have a predisposition to be used with an indirect directive function to communicate the actual desire of the speaker or to direct the listener to carry out a specific action, as in:

e.g. There may be a few things I want, I *want* to emphasize as important.
I *need* to talk to you sometime about when to footnote, how to footnote.
You *need* to be here at eight oh clock, oh, you *need* to be here pretty frequently, the class is based generally on lectures. Most of the test will be over lectures and so you will find that if you are not here, it will hurt your grade.

(Biber, 2006, p. 109-110)

He also reports that causation/effort verbs are approximately as frequent as desire verbs in the written registers, where they are typically employed with directive functions, as in:

e.g. If you *fail* to file your application or pay the graduation fee by the time specified, we may have to schedule you for graduation at a later date.
We do *permit* you to use coffee makers, water warmers, and microwave ovens under 700 watts.

(Biber, 2006, p. 110)

He suggests that causation/effort verbs are used for more informative purposes particularly in the course packs and textbooks, as in:

e.g. Such studies increase our general knowledge of language acquisition because they *allow* us to gain a better understanding of the conditions upon which it relies.
Opening valves and stopcocks too fast can *cause* joins to blow apart.
(Biber, 2006, p. 111)

He states that mental verbs controlling *to*-clauses are commonly used with directive function especially in the course management, as in:

e.g. I *expect* you to be prepared and to actively participate in the classroom discussion if the opportunity arises.
Remember to spell check and grammar check!
Reminder: Do not *forget* to convert the original data to SI.
(Biber, 2006, p. 111)

For complement clauses controlled by stance adjectives/ nouns, he drew to the conclusion that *that*-clauses framed by adjectives tend to be used with first person pronouns (*I, we*) to convey personal stance, as exemplified below:

e.g. Yeah I was, *I was afraid* [that you know maybe the flowery type stuff wouldn't be what they wanted]
I wanted to make *certain* [that you know sort of got up to date]
(Biber, 2006, p. 112)

As for the written registers, it is seen that *that*-clauses controlled by adjectives are used to convey directive functions in syllabi and institutional writing, as in:

e.g. With regard to the two different views of art contrasted above, *it is not surprising* [that Navajo society is one of artists…] (course packs)
It is imperative [that you come prepared to discuss and debate these materials…] (syllabi)
(Biber, 2006, p. 112)

The study has demonstrated that *that*-clauses controlled by nouns are primarily employed in the academic registers such as classroom teaching, textbooks and course

packs, whereby nouns carry the information of *argument, assumption, knowledge, possibility* and etc.

e.g. He dispels the *notion* [that the government is this overwhelming entity that can simply control us all].

(Biber, 2006, p. 112)

It has been revealed that *to*-clauses controlled by adjectives and nouns are more common than *that*-clauses particularly in the written registers, as in:

e.g. A hunting and gathering people in this area are *likely* [to have used such resources at one time or another]
It is very *difficult* [to determine the beginning of the exploration in human ecology in the USSR].

(Biber, 2006, p. 113)

He has found that institutional writing intensely rely on *to*-clause constructions framed by nouns which are nominalised counterpart of verbs and adjectives, as in:

e.g. If you have *plans* [to leave the country for a short visit], you must have your I20 and IAP66 forms signed…
Your *ability* [to make smart decisions] is lessened when you are drunk or high.

(Biber, 2006, p. 113)

Finally, he notifies that *noun* + *to*-clause constructions frequently impersonal although they are used with directive functions, as in:

e.g. *Failure* [to do so] may result in cancellation of enrolment, credits earned, or both.

The course provides the *opportunity* [to learn and apply various systems].

Permission [to enrol for less than the normal number of units] should be obtained from the O.I.P.

(Biber, 2006, p. 114)

Research articles written by Turkish-speaking and Spanish- speaking learners of English and those written by native speakers of English were compared with regard to the use of stance modals in a study conducted by Kafes (2009). The researcher has found that non-native speakers of English do not significantly differ in the use of stance modals except *may* and that there is a statistically significant difference between the native and non-native groups concerning the use of *can, could, may, might, would* and *should.* He has also found that non-native speakers of English have a tendency to employ similar modals for similar functions. Namely, it has been observed that both groups use *can, could* and *may* for expressing possibility, *will* and *would* for prediction, and *should* for expectation. Another finding of the study is that both native and non-native speakers display a tendency to use *must* for expressing obligation/necessity and *should* for suggestions.

Can (2012) investigated stance adverbs across argumentative essays written by Turkish university students and those by American university students. He revealed Turkish university students use stance adverbs more frequently than native speaker university students and that the former group employ less variety of stance adverbs than the latter.

Kotrc (2012) analysed stance adverbials in the language of press across randomly selected 18 articles published in the newspapers during the General Election 2010 in UK. Among his main findings are epistemic stance adverbials constituted the dominant majority of the adverbials seen in the articles and certainty adverbials outnumber likelihood adverbials.

CHAPTER III

METHODOLOGY

3.0. Introduction

This chapter mainly describes and outlines the design and the data analysis procedure of the study. First of all, Contrastive Interlanguage Analysis (CIA) is covered as the methodology employed in the analysis of non-native corpora used in the present study. Subsequently, native and non-native corpora are described in detail. Finally, data analysis procedure is clarified; namely, software packages utilized in the study are introduced and the ways statistical figures related to the findings of the study have been computed are outlined.

3.1. Contrastive Interlanguage Analysis

Contrastive Interlanguage Analysis (CIA) is a method proposed by Granger (1996) which primarily involves comparison of written or spoken utterances made by two or more groups of non-native speakers of a particular language. As suggested by Gilquin (2001), the approach allows two types of comparisons: (a) native language and non-native variety of that language, (b) different non-native varieties of the same language. Granger (2002) postulates that the former is useful in exploring the distinguishing features of learner language and that the latter enables us to assess the degree of generalizability of interlanguage features across learner populations and language situations.

As previously mentioned, our study is corpus-based in nature and it involves investigating author stance across the doctoral dissertations of native and non-native academic writers of English. Since it aims to reveal whether Turkish-speaking, Spanish-speaking and native academic writers significantly differ from each other with regard to epistemic stance devices, CIA has inevitably been selected as the main approach in data analysis procedure. The following are the six stages of data analysis procedure involved in this particular part of the study, the last three of which are supposed to provide answers to the research questions.

1. The Analysis of TACE: The corpus including doctoral dissertations of Turkish-speaking academic writers of English was examined with respect to epistemic stance devices in four main groups as *epistemic adjectives, epistemic adverbs, epistemic nouns* and *epistemic verbs*.
2. The Analysis of NACE: The data including doctoral dissertations of native academic writers of English were searched in terms of the epistemic stance devices in concern.
3. The Analysis of SACE: The doctoral dissertations produced by Spanish-speaking academic writers of English were examined with respect to the devices in question.
4. TACE vs NACE: Findings obtained from the analyses applied to TACE and NACE were compared to find out whether they significantly differ in the use of ESDs. (Research Question 1)
5. SACE vs NACE: Results of the analysis administered to SACE and NACE were compared to reveal whether they significantly differ in the use of ESDs. (Research Question 2)
6. TACE and SACE: Findings obtained from the analyses run to TACE and SACE were compared to see whether they significantly differ concerning the use of ESDs. (Research Question 3)

3.2. Methodology for Corpora Analysis

Aiming to find out whether there is a significant difference among native and non-native authors of English with respect to author stance, the present study is corpus-based in design. It includes the analysis of the epistemic stance devices that are frequently reported to have occurred in the written academic registers in Biber's study (2006), as mentioned in the second chapter (See Table 2.6). Wordsmith Tools Version 6.0 (Scott, 2012) was administered to three sets of corpora in order to identify occurrences of the ESDs. The analysis was performed through the sections *Findings, Discussion, Conclusion, Pedagogical Implications (Implications to English Language Teaching)* and *Suggestions for Further Research* of the doctoral dissertations produced by native and non-native authors of English within the fields of English Language Teaching, English Language and Literature, Applied Linguistics and Modern Languages between 2005 and 2012.

3.3. Data

The researcher collected doctoral dissertations produced by Turkish-speaking Academic Authors of English (TAEs, henceforth), Spanish-Speaking Academic Authors of English (SAEs, hereafter) and Native Academic Authors of English (NAEs, henceforth) within the field of English Language Teaching, English Language and Literature, Applied Linguistics and Modern Languages. Hence, three sets of corpora were built with a total number of 133 dissertations penned between 2005 and 2012 [TACE (Turkish Academic Corpus of English): 48; SACE (Spanish Academic Corpus of English): 43; NACE (Native Academic Corpus of English): 45]. For practical reasons, the corpora were constructed with the collection of dissertations available online. Each work included in the study was identified with a particular code containing the name of the corpus it belongs to, the institution it was submitted to, capitals of the name of its author and the year it was written. Table 7 displays a part of the table where the information at stake was recorded.

Table 7

Coding Doctoral Dissertations in NACE

Author	Code	Title	University	Year
Sarah Collie	<NACE-UE-2007-AB>	English stress preservation and Stratal Optimality Theory	The University of Edinburgh	2007
Catherine (Kate) Wild	<NACE-UG-2010-CW>	Attitudes towards English Usage in the Late Modern Period: The Case of Phrasal Verbs	University of Glasgow	2010

The sections *Abstract, Introduction, Review of Literature, Methodology* and *References* of the doctoral dissertations were not included in the corpora. The analysis was performed through the sections *Findings, Discussion, Conclusion, Pedagogical Implications (Implications to English Language Teaching)* and *Suggestions for Further Research* due to the assumption that those who supervised the dissertations might have influenced the writing process relatively less in these sections than in the above-mentioned excluded sections. In addition, all titles, tables, subtitles, tables, figures, quotations and paraphrases appearing through the sections in concern were excluded from the data. Over 2 million words were analysed with respect to the use of ESDs, which were previously categorized into four main groups by Biber (2006) as *epistemic*

adverbs, epistemic verbs, epistemic nouns and *epistemic adjectives*. Table 8 and Table 9 illustrate the size of the corpora used in this study and the categorization of ESDs, respectively.

Table 8

Corpus Size

Corpus	Dissertation (n)	Word (n)
TACE	48	675.072
SACE	43	668.256
NACE	45	671.475
Total	136	2.014.873

Table 9

Epistemic Stance Devices

Epistemic Adverbs	
Certainty:	*actually, always, certainly, definitely, indeed, inevitably, in fact, never, of course, obviously, really, undoubtedly, without doubt, no doubt*
Likelihood:	*apparently, evidently, kind of, in most cases/instances, perhaps, possibly, predictably, probably, roughly, sort of, maybe*
Epistemic verbs controlling *that*- clauses	
Certainty:	*conclude, demonstrate, determine, discover, find, know, learn, mean, notice, observe, prove, realize, recognize, remember, see, show, understand*
Likelihood:	*assume, believe, doubt, gather, guess, hypothesize, imagine, predict, presuppose, presume, reckon, seem, speculate, suppose, suspect, think*
Epistemic verbs controlling *to*- clauses	
Likelihood:	*appear, happen, seem, tend*
Epistemic adjectives controlling *that*- clauses	
Certainty:	*apparent, certain, clear, confident, convinced, correct, evident, false, impossible, inevitable, obvious, positive, right, sure, true, well-known*
Likelihood:	*doubtful, likely, possible, probable*
Epistemic adjectives controlling *to*- clauses	
Certainty/ likelihood:	*apt, certain, due, guaranteed, liable, likely, prone, unlikely, sure*
Certainty:	*assertion, conclusion, conviction, discovery, doubt, fact, knowledge, observation, principle, realization, result, statement*
Likelihood:	*assumption, belief, claim, contention, feeling, hypothesis, idea, implication, impression, notion, opinion, possibility, presumption, suggestion*

Source: Biber, 2006, p. 92-93

3.4. Instruments

3.4.1. WordSmith Tools 6.0

WordSmith Tools provides concordance programme, which lists the occurrence of given word or phrase in a corpus (Scott, 2001). Baker et al (2006) describe it as "a software package for analysing the lexis of texts and corpora, developed by Scott (1998)". They suggest that it can be used to produce frequency lists, to run concordance searches and calculate collocations for particular words, and to find keywords in a text examine their distribution. They maintain that it can be used with both plain text and texts with mark-up tags (i.e. Standard Generalised Mark-up Language) and that, unlike some other analysis packages (SARA, WordCruncher), WS does not require the corpus to be indexed in advance. Figure 1 shows the main screen of the programme in question.

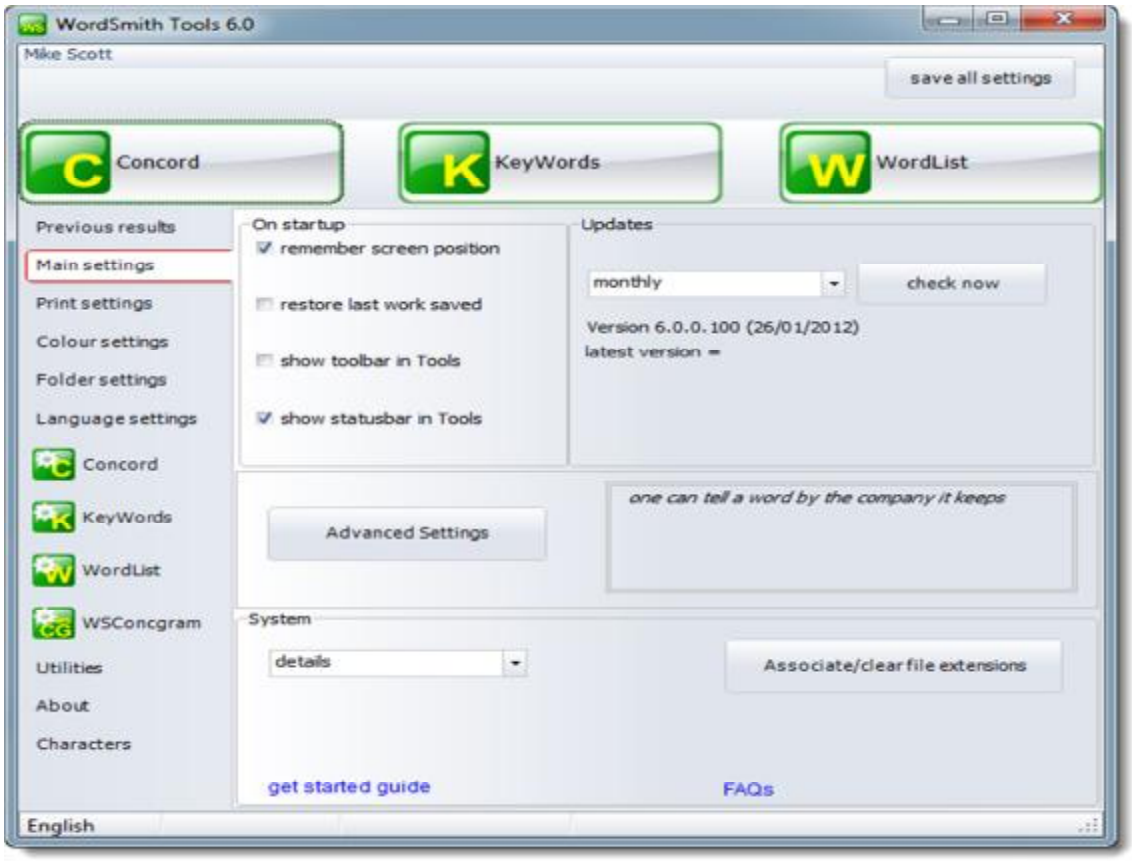


Figure 1. A screenshot of the main screen of WS Tools

According to Gavioli (2005), it has three main components. Firstly, it has a concordancer with the possibility of editing and classifying the list of instances, as illustrated in Figure 2.



Figure 2. A screenshot of concordance menu in WS tools

Each set of corpus was uploaded to the programme and Concordance menu was selected from the main screen of WS Tools. Subsequently, the epistemic stance devices were individually searched across each corpus. It is noteworthy that certain items were paid special attention in the process. They are,

1. Epistemic nouns and verbs of which spelling change across language forms (e.g. *analyse* in BrE and *analyze* in AmE)
2. Epistemic verbs (searched in present, progressive, past and participle forms, e.g. *see/ seeing/ sees/ saw/ seen*)
3. Epistemic nouns (searched in singular and plural forms, e.g. *conclusion/s*)
4. Epistemic nouns and epistemic verbs identical in writing (e.g. *doubt… that*)
5. Epistemic noun likely to be confused with a verb (e.g. *claim ….. that*)

Another facility offered in WS Tools is a wordlist to calculate the frequency of the words in a corpus. It helps the language researcher identify the common words in a

corpus, information which is useful for example when determining which lexical items to teach and which to ignore, or when the materials writer is attempting to ensure that new vocabulary is met more than once in a textbook (Scott, 2001). Figure 3 shows a screenshot of a wordlist in WS Tools.

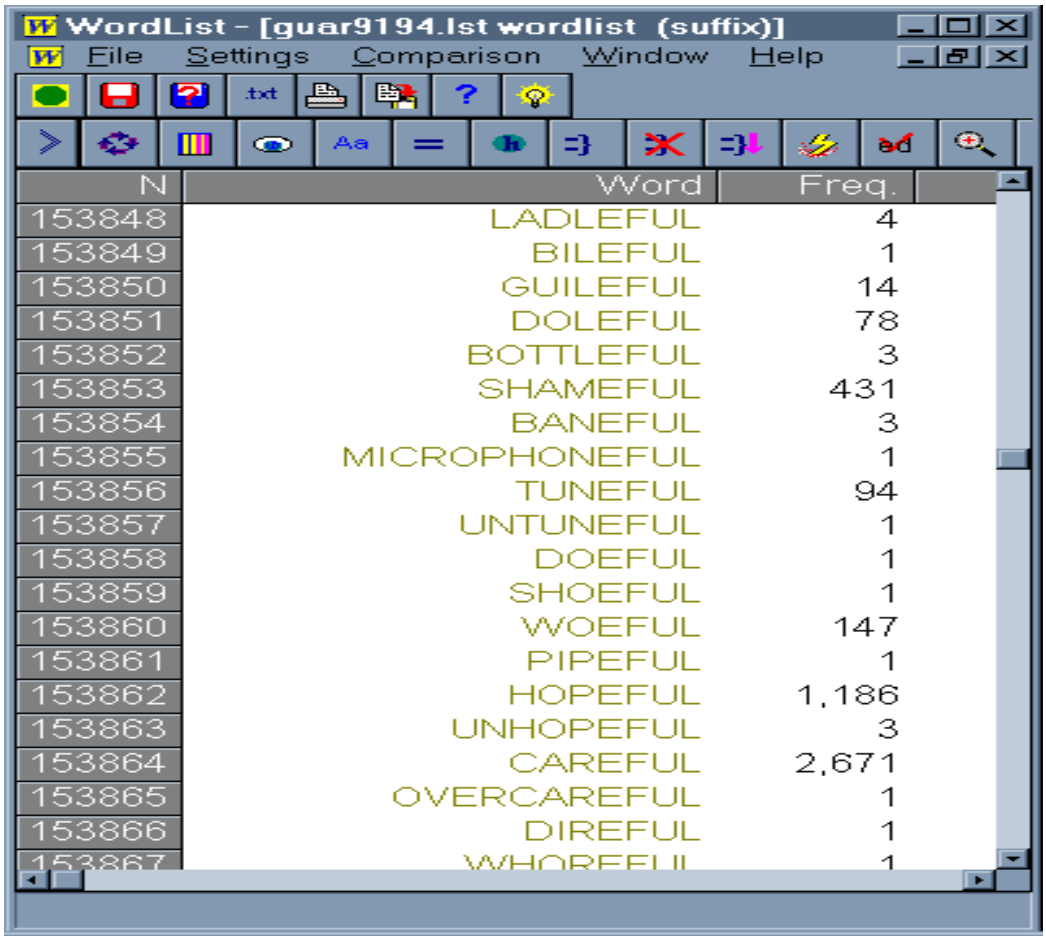
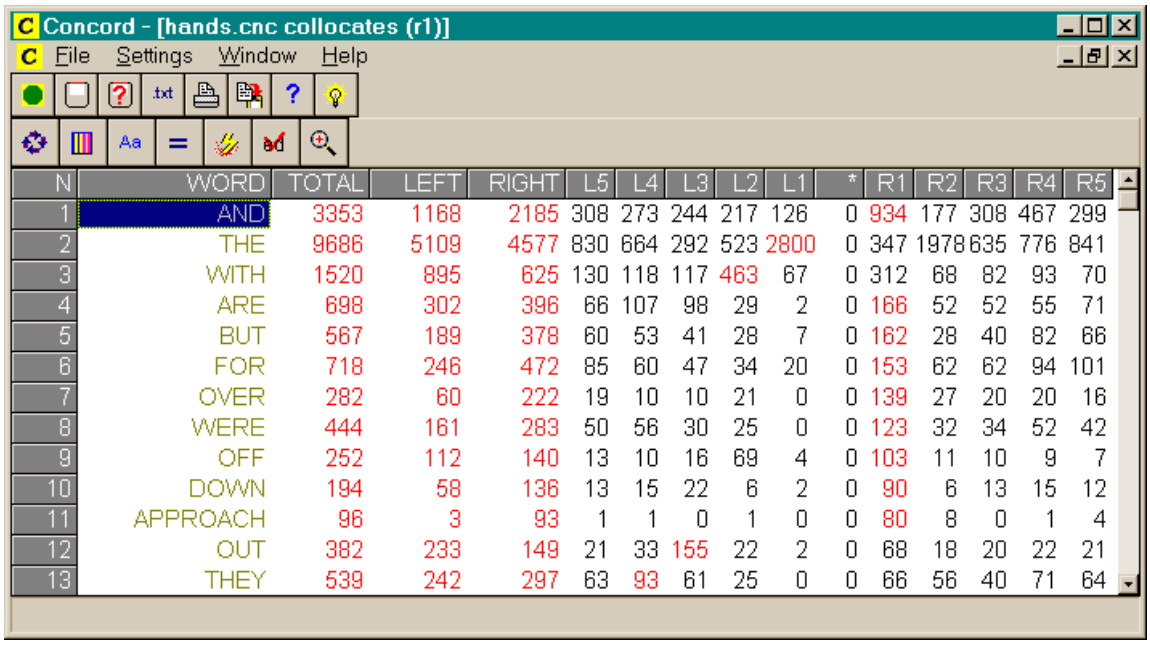


Figure 3. A screenshot of a wordlist in WS tools

Finally, it has a keyword to compare lists from different corpora, sub-corpora or texts and calculate those words, which are significantly more or less frequent in one set (Gavioli, 2005), as illustrated in Figure 4.



N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	AND	3353	1168	2185	308	273	244	217	126	0	934	177	308	467	299
2	THE	9686	5109	4577	830	664	292	523	2800	0	347	1978	635	776	841
3	WITH	1520	895	625	130	118	117	463	67	0	312	68	82	93	70
4	ARE	698	302	396	66	107	98	29	2	0	166	52	52	55	71
5	BUT	567	189	378	60	53	41	28	7	0	162	28	40	82	66
6	FOR	718	246	472	85	60	47	34	20	0	153	62	62	94	101
7	OVER	282	60	222	19	10	10	21	0	0	139	27	20	20	16
8	WERE	444	161	283	50	56	30	25	0	0	123	32	34	52	42
9	OFF	252	112	140	13	10	16	69	4	0	103	11	10	9	7
10	DOWN	194	58	136	13	15	22	6	2	0	90	6	13	15	12
11	APPROACH	96	3	93	1	1	0	1	0	0	80	8	0	1	4
12	OUT	382	233	149	21	33	155	22	2	0	68	18	20	22	21
13	THEY	539	242	297	63	93	61	25	0	0	66	56	40	71	64

Figure 4. A screenshot of keyword page in WS tools

WordSmith Tools has been used in several studies by means of describing and analyzing various issues in corpus research (Scott, 1997; Granger and Tribble, 1998; Sardinha, 1999; Barbara and Scott, 1999; Xiao and McEnery, 2005, Astrid and Johnson, 2006; Gilquin and Paquot, 2007, 2008, cited in http://lexically.net/wordsmith/corpus linguistics_links/papers_using_wordsmith.htm).

WordSmith Tools 6.0 was employed in the present study to investigate epistemic stance devices in four major groups such as epistemic adjectives, epistemic adverbs, epistemic nouns and epistemic verbs.

3.4.2. Log-Likelihood Statistics

Log-likelihood is a test for statistical significance, similar to the chi-squared measure that is often used in corpus analysis such as for collocation or keyword analysis and it is sometimes called G-square or G score (Baker et al, 2006). In statistical analysis of texts, to test the frequency distributions, LL test is a reliable alternative to Pearsons'Chi-square (Dunning, 1993). It is utilized to compare the observed and expected values for two data sets. Figure 5 shows a screenshot of the log-likelihood calculator (Ryson and Garside, 2000).

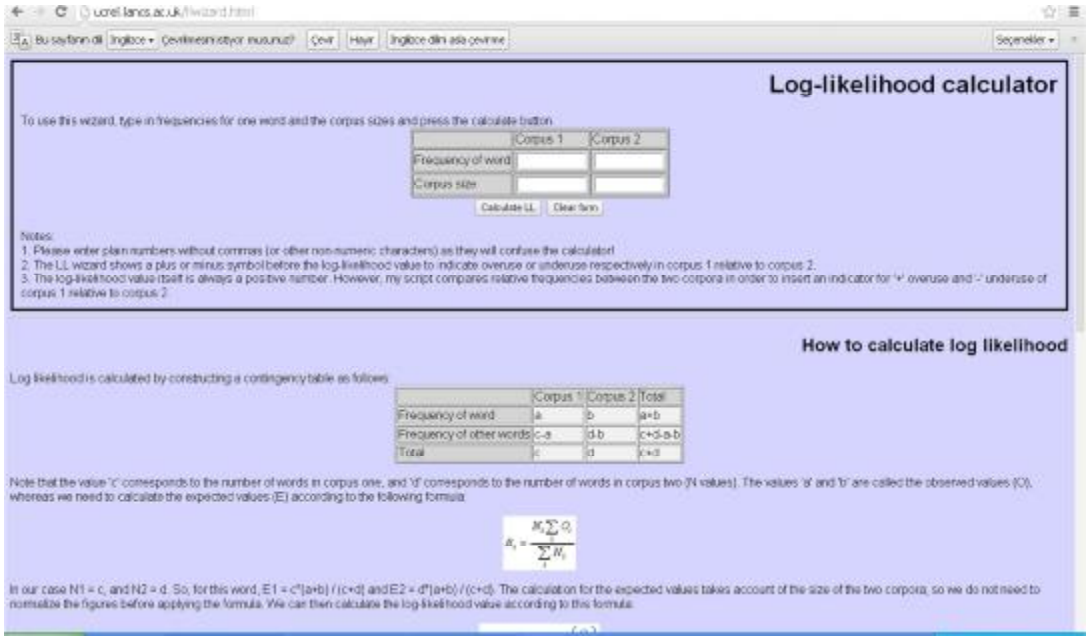


Figure 5. A screenshot of Log-likelihood Calculator

Log likelihood is calculated by constructing a contingency table (Rayson, 2000), as presented in Table 10.

Table 10

Contingency Table of LL Calculation

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

Source: Rayson & Garside, 2000, p. 3

Note that the value 'c' corresponds to the number of words in corpus one, and 'd' corresponds to the number of words in corpus two (N values). The values 'a' and 'b' are called the observed values (O), whereas we need to calculate the expected values (E) according to the following formula (Rayson and Garside, 2000):

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

The following is how Rayson and Garside (2000) clarify the formula in concern.

"In our case N1 = c, and N2 = d. So, for this word, E1 = c*(a+b) / (c+d) and E2 =

d*(a+b) / (c+d). The calculation for the expected values takes account of the size of the two corpora, so we do not need to normalize the figures before applying the formula. We can then calculate the log-likelihood value according to this formula:"

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

"This equates to calculating log-likelihood G2 as follows: G2 = 2*((a*ln (a/E1)) + (b*ln (b/E2)))….. 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"

Once frequency of a particular item and corpus size are entered, log-likelihood calculator results appear in the screen, as shown in Figure 6.

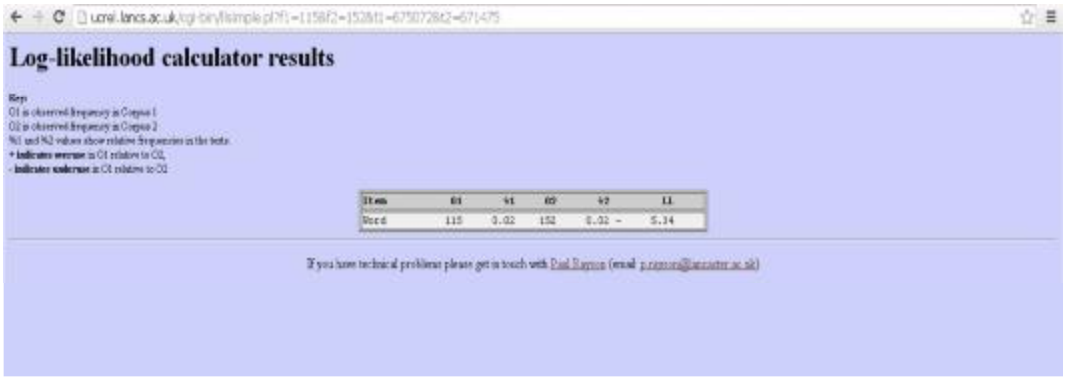


Figure 6. A screenshot of log likelihood calculator results

Rayson and Garside (2000) note that O1 and O2 refer to the observed frequencies of a particular item in Corpus 1 and O2, respectively and that %1 and %2 values show relative frequencies in the texts. The symbol (+) indicates overuse and (-) indicates underuse in O1 relative to O2. Log-likelihood statistics have been extensively utilized in data analysis process of significant studies conducted in the field of corpus linguistics

(e.g. Granger and Rayson, 1998; Scott, 2011).

LL calculation was one of the statistical measurements selected in data analysis of the present study due to the fact that it primarily involves Contrastive Interlanguage Analysis (Ibid). That is, two sets of non-native and a set of native corpora including doctoral dissertations of Turkish-speaking, Spanish-speaking and native authors of English were analysed through LL calculation to find out whether there is overuse/ underuse across them in the use of epistemic stance devices as well as lexical categories.

3.4.3. Wmatrix

Wmatrix is an analysis system for English texts and corpora developed by Rayson (2004) at Lancaster University. Accessed using a web-based interface, the programme functions as a concordance in addition to tagging text using CLAWS part-of-speech tagger and UCREL Semantic Analysis System (USAS) semantic tagger. It can calculate keywords, key part of speech tags (Baker et al., 2006).

Wmatrix users can upload their own corpus data to the system, so that it can be automatically annotated and viewed within the web browser. Each file is stored in a folder (equivalent to a folder in Windows or directory on Unix) (Rayson, 2004). The main screen of Wmatrix is provided in Figure 7.

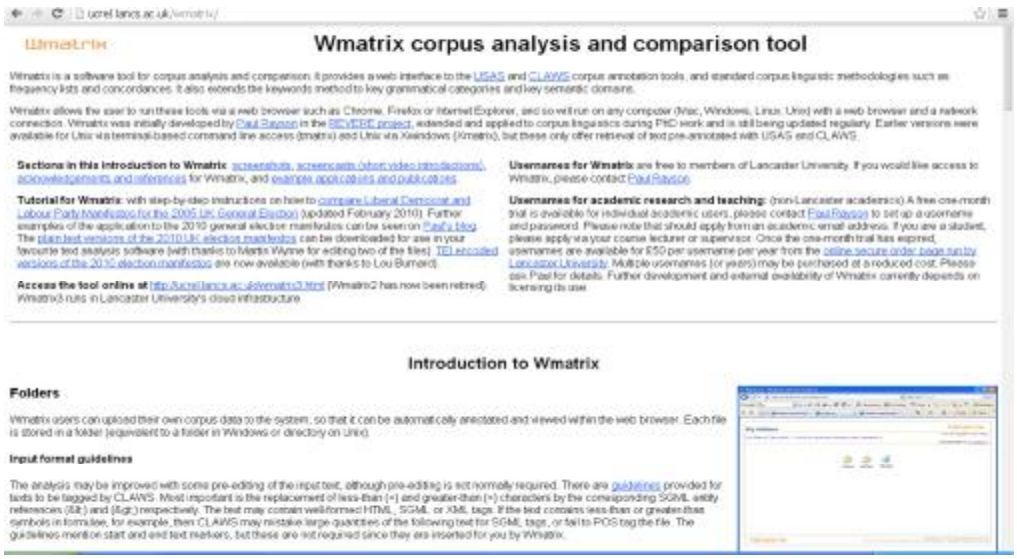


Figure 7. A Screenshot of the main page of Wmatrix

The files in each set of corpora were uploaded to the programme. However, it is noteworthy that the software does not allow the user to analyse files including more than 100.000 words. Accordingly, the corpora were restructured with creating new files containing a maximum of 100.000 words. As a consequence, each set consisted of 7 files and a total number of 21 files were individually uploaded to the programme, one of which is illustrated in the screenshot provided in Figure 8.

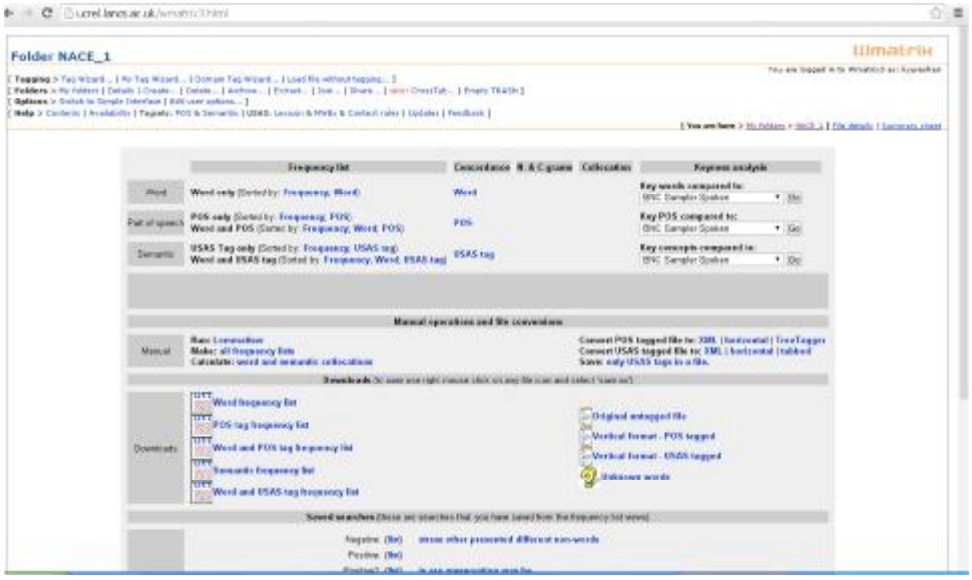


Figure 8. A Screenshot of a sample folder uploaded to Wmatrix

Rayson (2004) suggests that once the file has been uploaded to the web server, it is POS tagged by CLAWS and semantically tagged by USAS and that the format for a frequency list is a very simple two column format with a total line at the head of the file. The programme offers a variant of tag wizard "My Tag Wizard", which allows its users to override or extend the system dictionaries for their own data. The Figure 9 presents a screenshot of the tag wizard.



Figure 9. A Screenshot of the tag wizard in Wmatrix

Frequency lists are available for words in the simple interface, and in the advanced interface for POS tags and semantic tags and they can be sorted alphabetically or by frequency (Rayson, 2004). He offers that the frequency lists make it possible for the users to see standard concordances, which can show the usual word based concordance as well as all occurrences for words in one POS or semantic category. Figure 10 shows the POS frequencies of the words tagged in a folder in NACE.

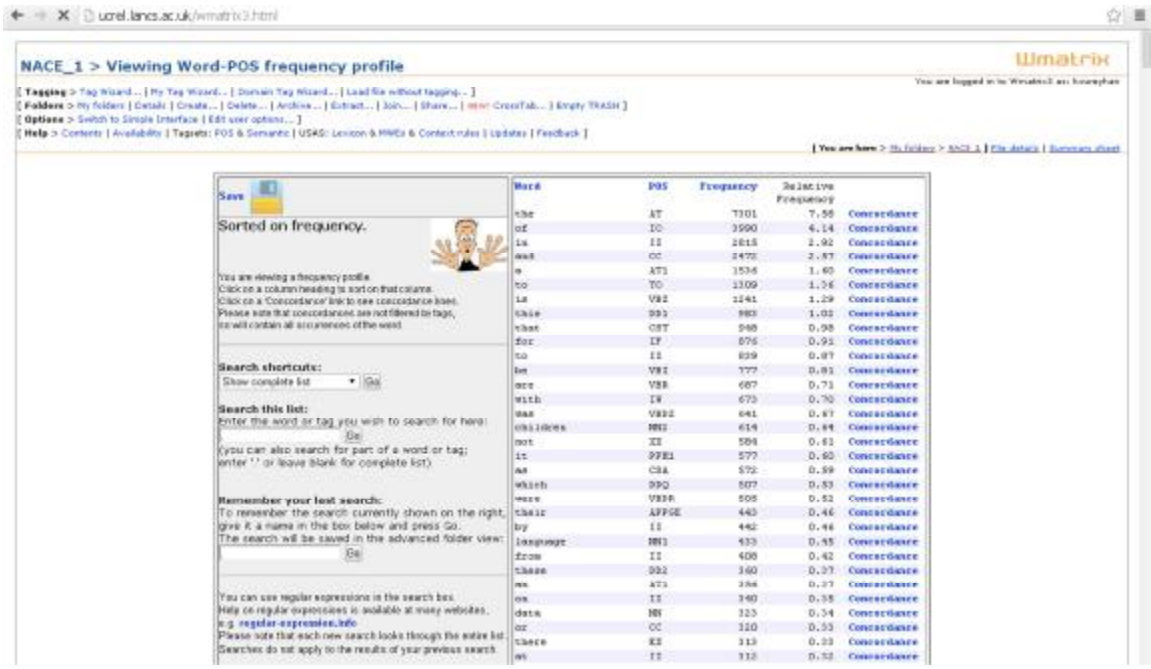


Figure 10. A screenshot of word POS frequencies

Wmatrix has been used in the present study particularly in analysing three corpora in terms of lexical categories. Since the programme automatically tagged the words, the researcher did not need to employ manual check over the results, which is an advantage in the data analysis process. Nonetheless, it was somewhat challenging and time-taking to compile the results obtained from the individual analysis of 21 folders in three corpora.

3.5. Data Analysis

Doctoral dissertations written by native and non-native academic authors of English (TAEs and SAEs) were analysed to see whether they significantly differ regarding the ESDs. The devices in concern were classified into four major groups: epistemic adjectives, epistemic adverbs, epistemic nouns and epistemic verbs. Subsequently, they were identified through Wordsmith Tools (6.0) and their frequencies were separately measured for each corpus. In order to reveal whether there is a statistically significant difference between the dissertations written by TAEs and those produced by NAEs with regard to ESDs, the Log Likelihood Test was administered to the findings. Then, the findings obtained from the dissertations produced by SAEs and those elicited from the dissertations written by NAEs were compared using the same statistical procedure. Finally, findings drawn from the dissertations authored by TAEs and those obtained from the works by SAEs were compared for Contrastive Interlanguage Analysis.

CHAPTER IV

RESULTS AND DISCUSSION

4.0. Introduction

The present chapter provides the findings drawn from the analyses of three corpora (sets of works produced by TAEs, SAEs and NAEs). As stated in the previous chapter, a total number of 117 types of English ESDs (under four main categories) were examined across three corpora via WordSmith Tools, a computer software package developed by Scott "for analysing the lexis of texts and corpora in order to produce frequency lists, to run concordance searches and calculate collocations for particular words, and to find keywords in a text and examine their distribution" (Baker et al. 2006, p. 169).

Statistical analyses for categorical and individual ESDs were carried out in order to calculate the frequencies of the devices in concern and to reveal whether native and non-native academic authors of English significantly differ from each other with respect to the use of ESDs.

4.1. Results

4.1.1. Epistemic Stance Devices in Three Corpora

Initially, ESDs occurring across three corpora were obtained and measured regardless of their category (*epistemic adjective, epistemic adverb, epistemic noun* and *epistemic verb*). Following Biber (2006), a total number of 117 types were identified over Turkish, Spanish and Native Corpora. Each type was identified and its frequency was calculated using the WordSmith Tools (Version 6.0) to reveal the proportion of ESDs used in three corpora. Those not found across three corpora are displayed in Table 11.

Table 11

ESD Types not Found in Three Corpora

ESD Category	Epistemic Adjective	Epistemic Noun	Epistemic Verb
ESD Type	Certain (-to)	Presumption	Gather
	Correct	Principle	Reckon
	Doubtful	Realization	
	Due		
	False		
	Inevitable		
	Possible		
	Right		

As can be seen in Table 11, most of the types not seen in three corpora are epistemic adjectives, six of which are the types expected to frame *–that* clause and two to frame *–to* clause, as in the following examples.

e.g. She looks ***certain to*** win an Oscar.

Is it ***correct that*** personnel cannot be charged to an award until their COI Training is completed?

With her injuries it's ***doubtful that*** she'll ever walk again.

Rose is ***due to*** start school in January.

Is it true or ***false that*** the Robinson projection shows the size of the continents accurately and the water areas are expanded to fill the extra space?

It was ***inevitable that*** there would be job losses.

It's just ***possible that*** I gave them the wrong directions.

Is it ***right that*** mostly college drop-out students run the biggest businesses in the world?

There is a general ***presumption that*** the doctor knows best.

Reflexology is based on the ***principle that*** specific areas on the feet correspond to different parts of the body.

There's a growing ***realization that*** changes must be made.

I ***gather*** from your letter ***that*** you're not enjoying your job.

I ***reckon (that)*** I am going to get that job.

Oxford Advanced Learner's Dictionary, 2010

Subsequently, the corpora were compared to each other regarding the use of the devices in question. Figure 11 shows the distribution of ESD types found across three corpora.

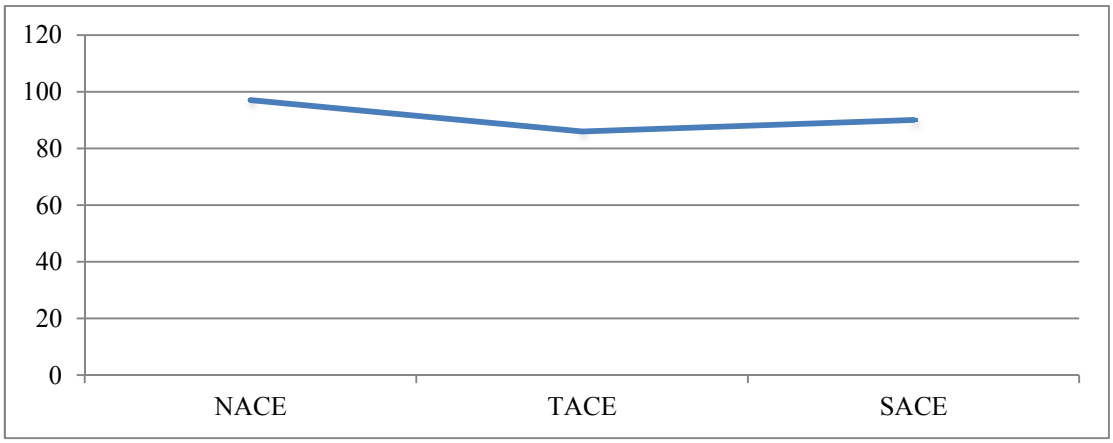


Figure 11. Distribution of ESD types in three corpora

As displayed in Figure 11, in NACE, 97 out of 117 types were used, which corresponds to the highest frequency among all corpora. SACE and TACE included 90 and 86 ESD types, respectively.

In accordance with the findings, ESDs were mostly frequented in NACE (4719 times) whereas they appeared 4455 and 4333 times in TACE and SACE, respectively. Thus, the frequency analysis has indicated that English ESDs were underused by TAEs and SAEs when compared to NAEs, which is illustrated in Table 12.

Table 12

Overall Comparative Frequency Distribution of ESDs in Three Corpora

	NACE (L1)	TACE (L2)	SACE (L2)
Corpus Size in words	671.475	675.072	668.256
ESDs (n)	4719	4455	4333
n per 10.000	70.28	65.99	64.84
T/t ratio (%)	0.7	0.7	0.6
Number of ESD types used	97	86	90

n= raw frequency of ESDs

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

As presented in Table 12, the highest and the lowest frequencies have been observed in NACE and SACE, respectively, which means there is an underuse in the dissertations produced by TAEs (4455 times) and those written by SAEs (4333 times). In order to investigate to what extent native and non-native authors of English use ESDs in their works, ESDs were analysed within their categorizations across three corpora. The main categories of ESDs previously proposed by Biber (2006) were measured regarding the total frequency of use. Figure 12 displays the frequency of categories in concern over three corpora.

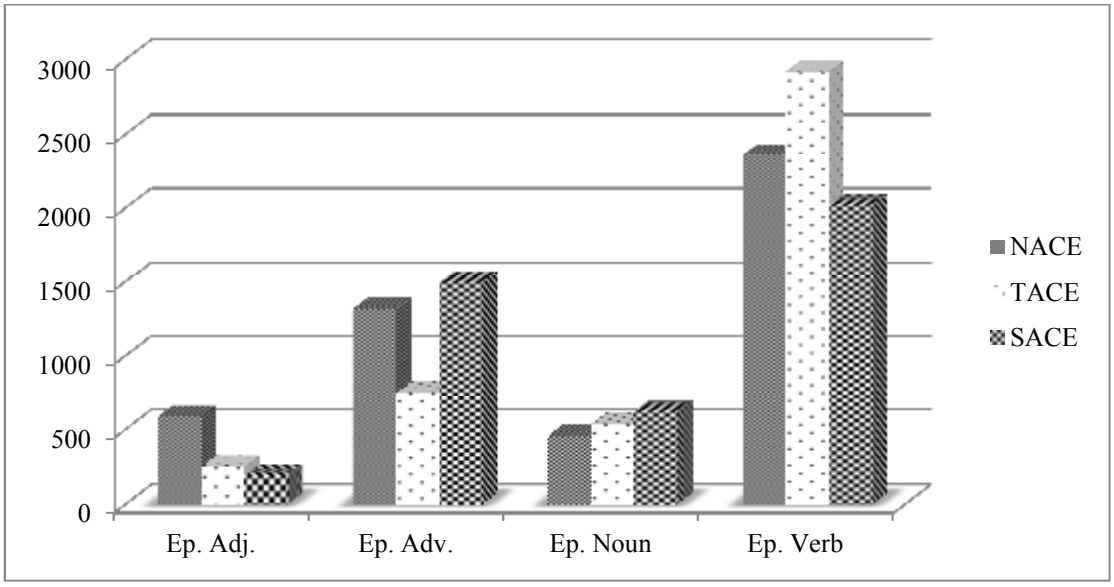


Figure 12. Epistemic categories in three corpora

Figure 12 suggests that three corpora significantly differ in the frequency distribution of the categories. That is, there is not a general tendency over the corpora with respect to categorical frequencies. For instance, epistemic adverbs and epistemic nouns were mostly preferred by SAEs while epistemic verbs were employed more frequently by TAEs. Namely, epistemic adverbs appeared 1496 times in SACE whereas they occurred 1315 and 749 times in NACE and TACE, respectively. Likewise, epistemic nouns were found more frequently in SACE than both TACE and NACE (SACE: 623; TACE: 535 and NACE: 455). As for epistemic verbs, approximately 40% of them were observed in TACE (2920 times) whereas 32.3% and approximately 28% appeared in NACE and SACE, respectively (NACE: 2360 and SACE: 2022). Epistemic adjectives, on the other hand, mostly occurred in NACE (586 times) while they were seen 251 and 192 times in TACE and SACE, respectively. These results indicate an

overuse and an underuse for certain categories in non-native corpora against the native corpus. For instance, epistemic verbs revealed to be overused by TAEs and epistemic adverbs and epistemic nouns were overused by SAEs against NAEs. In response, a significant underuse was observed in the use of epistemic adjectives by non-native academic authors against native authors.

It is considered essential to employ statistics to investigate whether the values of overuse and underuse are meaningful. LL ratio was calculated to make comparisons across corpora in terms of the use of ESDs. Initially, overall frequency of ESDs used in the corpora was calculated. The first LL analysis was made to reveal whether there is a significant difference between the TACE and NACE regarding ESDs. Table 13 provides the related results.

Table 13

LL Ratio of Overall ESDs in TACE and NACE

	TACE (O1)	%1	NACE (O2)	%2	LL Ratio (*p< 0.05)
ESDs	4455	0.66	4719	0.70	-9.07

O1 is observed frequency in Corpus 1

O2 is observed frequency in Corpus 2

%1 and %2 values show relative frequencies in the texts.

+ indicates overuse in O1 relative to O2

- indicates underuse in O1 relative to O2

In Table 13, O1 and O2 refer to overall frequency of ESDs found in TACE and NACE respectively. %1 value represents the relative frequency of the devices in the dissertations (i.e., 0.66 relative frequencies in TACE means that an approximate amount of 0.66 ESDs per 100 words were employed in TACE). In a similar vein, relative frequency of NACE was measured 0.70 ESDs per 100 words in NACE. In accordance with the results, ESDs were underused in TACE with -9.07 LL value ($p<0.05$), meaning that there is a statistically significant difference between two corpora with respect to ESD frequency; thereby, the underuse in TACE in relative to NACE was confirmed by LL calculation. As a next step, SACE and NACE were compared regarding the overall frequency of ESD usage. Table 14 outlines LL ratio of the frequency in concern.

Table 14

LL Ratio of Overall ESDs in SACE and NACE

	SACE (O1)	%1	NACE (O2)	%2	LL Ratio (*p< 0.05)
ESDs	4333	0.65	4719	0.70	-14.66

O1 is observed frequency in Corpus 1

O2 is observed frequency in Corpus 2

%1 and %2 values show relative frequencies in the texts.

+ indicates overuse in O1 relative to O2

- indicates underuse in O1 relative to O2

The underuse in SACE against NACE was approved by LL results with a value of -13.77. Additionally, relative frequency per 100 words in each corpus indicates a difference between the corpora in concern (SACE: 0.65 and NACE: 0.70 ESDs per 100 words). Lastly, TACE and SACE were compared via LL test. The results show that ESDs were employed by TAEs more frequently than SAEs, as depicted in Table 15.

Table 15

LL Ratio of Overall ESDs in TACE and SACE

	TACE (O1)	%1	SACE (O2)	%2	LL Ratio (*p< 0.05)
ESDs	4455	0.66	4333	0.65	+0.68

O1 is observed frequency in Corpus 1

O2 is observed frequency in Corpus 2

%1 and %2 values show relative frequencies in the texts.

+ indicates overuse in O1 relative to O2

- indicates underuse in O1 relative to O2

As viewed from Table 15, LL value between TACE and SACE revealed an overuse 0.68 as expected from the frequency difference observed between the two corpora. To sum up, this difference was measured lower than the difference between SACE and NACE (-14.66 LL value) and the one between TACE and NACE (-9.07 LL value).

Frequency ratio of each epistemic category needs to be compared with that of

each category for 10.000 words in NACE to see whether there is an overuse or underuse in epistemic category frequency per 10.000 in TACE, the results of which are presented in Table 16.

Table 16

Overused/underused ESD Categories in TACE in comparison with NACE

ESD Category	N	TACE n/ 10.000	N	NACE n/ 10.000	overuse/ underuse + / -
Epistemic		3.7		8.7	-6.0
Adjectives	251		586		
Epistemic		11.1		19.6	-8.5
Adverbs	749		1315		
Epistemic Nouns	535	7.9	455	6.8	+1.1
Epistemic Verbs	2920	43.3	2360	35.1	+8.2

n= raw frequency of epistemic categories

n/ 10.000= frequency of epistemic categories per 10.000 words

(-/+) = difference between relevant value and value in NACE per 10.000 words; (+) denotes overuse; (-) denotes underuse

As can be seen in Table 16, the frequency ratio of epistemic verbs in NACE was measured 35.1 per 10.000 words and the difference between two corpora revealed 8.2, which means these items were used in NACE 8 times less than TACE in every 10.000 words.

LL ratio of categorical frequencies was applied among three corpora. Thus, the categorical frequencies in NACE were compared with those in non-native corpora. LL ratio in ESD categories was calculated between NACE and TACE to reveal the overused or underused categories. As presented in Table 16, the frequency analysis of ESD categories in TACE in comparison to NACE approved the overuse and underuse in certain categories. Namely, frequency analysis per 10.000 words between TACE and NACE showed an underuse in the categories of epistemic adjectives and epistemic adverbs, and an overuse in the categories of epistemic nouns and epistemic verbs. LL ratio of categorical comparison confirmed the categorical frequency differences as well as overuse and underuse in TACE, as illustrated in Table 17.

Table 17

LL Ratio of ESD Categories in TACE and NACE

ESD Category	TACE n	NACE n	LL Ratio (*p< 0.05)
Epistemic Adjectives	251	586	-139.71
Epistemic Adverbs	749	1315	-160.26
Epistemic Nouns	535	455	+6.05
Epistemic Verbs	2920	2360	+56.55

n=raw frequency of ESDs in corpus
+ indicates overuse in TACE relative to NACE
- indicates underuse in TACE relative to NACE

Categorical frequencies of ESDs in SACE were compared with those in NACE. Table 18 exhibits the frequencies of ESDs per 10.000 words in SACE and NACE.

Table 18

Overused/underused ESD Categories in SACE in comparison with NACE

		SACE		NACE	overuse/ underuse
ESD Category	N	n/ 10.000	N	n/ 10.000	+ / -
Epistemic Adjectives	192	2.9	586	8.7	-5.8
Epistemic Adverbs	1496	22.4	1315	19.6	+2.8
Epistemic Nouns	623	9.3	455	6.8	+2.5
Epistemic Verbs	2022	30.3	2360	35.1	-4.8

n= raw frequency of epistemic categories
n/ 10.000= frequency of epistemic categories per 10.000 words
(-/+) = difference between relevant value and value in NACE per 10.000 words; (+) denotes overuse; (-) denotes underuse

The categorical frequency of ESDs in SACE indicated different results from those in TACE; namely, epistemic adverbs and epistemic nouns were used 2.8 and 2.5 more times per 10.000 words in SACE than NACE, respectively. As for epistemic adjectives and epistemic verbs, both categories revealed to be underused by SAEs. For instance, epistemic adjectives were preferred 5.8 and epistemic verbs 4.8 fewer times per 10.000 words in SACE than NACE. Categorical frequency differences between SACE and NACE were also approved by LL results. Table 19 contains the related results.

Table 19

LL Ratio of ESD Categories in SACE and NACE

ESD Category	**SACE n**	**NACE n**	**LL Ratio (*p< 0.05)**
Epistemic Adjectives	192	586	-207.19
Epistemic Adverbs	1496	1315	+12.55
Epistemic Nouns	623	455	+27.10
Epistemic Verbs	2022	2360	-24.50

n=raw frequency of ESDs in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

In contrast to the case in TACE, the highest difference in LL ratio between SACE and NACE was observed as underuse in epistemic adjective category with -207.19 LL value and the least difference was found as overuse in epistemic adverb category with +12.55 LL value. All the differences between SACE and NACE with respect to LL values of ESD categories were found statistically significant ($p < 0.05$).

So far, epistemic stance devices have been covered within four main categories as epistemic adjectives, epistemic adverbs, epistemic nouns and epistemic verbs. That is, frequency categories of these items across three corpora as well as the LL results obtained from the analyses made between the native and non-native corpora have been presented. However, this is not sufficient to account for whether these devices tend to convey the meaning of *certainty* or *likelihood*. Consequently, a further analysis has been made in order to examine three corpora with respect to the use of certainty and likelihood epistemic stance devices, which will be discussed in the following section.

4.1.2. Certainty and Likelihood Devices in Three Corpora

As mentioned in the previous section, the non-native corpora were individually compared with the native corpus with regard to epistemic stance devices. In this section, we will cover these devices in two broader categories as *certainty* and *likelihood* and report the results of the analysis administered to three corpora to investigate whether they significantly differ in the use of these items. Figure 13 provides the overall frequencies of the items in concern over three corpora.

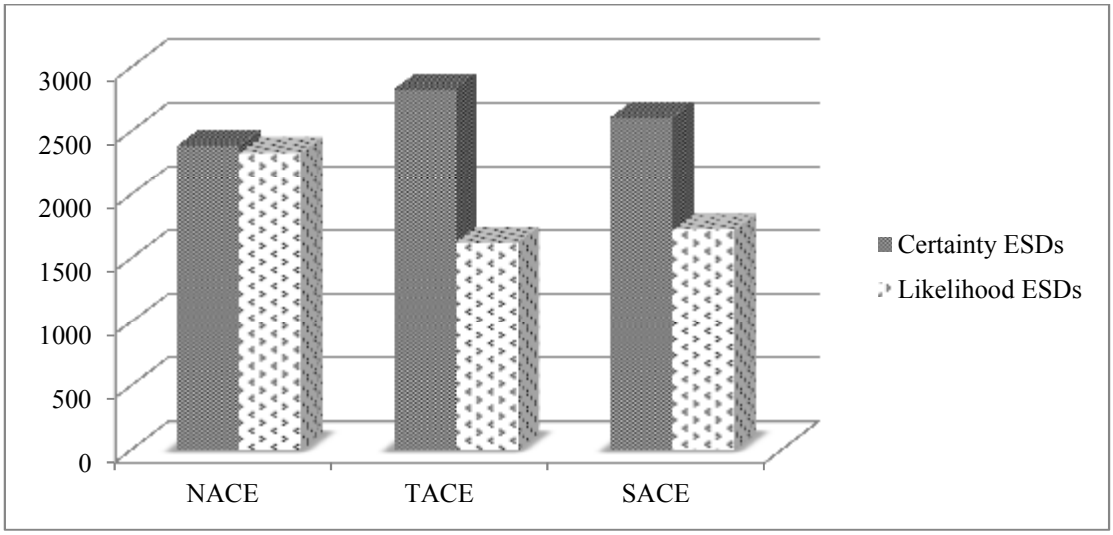


Figure 13. Certainty and likelihood devices in three corpora

As suggested in Figure 13, certainty devices were found more frequently than likelihood devices in all corpora. However, the corpora in concern considerably differ in the use of these items. Certainty devices occur in TACE with the highest rate when compared to SACE and NACE. Namely, they were used 2829 times in TACE, 2611 times in SACE and 2389 times in NACE. Likelihood devices, on the other hand, were mostly found in NACE with a frequency of 2330 while they appeared 1722 times in SACE and 1626 times in TACE.

Non-native corpora were individually compared to the native corpus for 10.000 words to see whether there is an overuse or underuse in the use of certainty and likelihood devices. The results are presented in Table 20.

Table 20

Overused/underused Certainty and Likelihood Devices in TACE in Comparison With NACE

Category	N	TACE n/ 10.000	N	NACE n/ 10.000	overuse/ underuse + / -
Certainty	2829	41.91	2389	35.58	+6.33
Likelihood	1626	26.09	2330	34.70	-8.61

n= raw frequency of epistemic categories

n/ 10.000= frequency of epistemic categories per 10.000 words

(-/+) = difference between relevant value and value in NACE per 10.000 words; (+) denotes overuse; (-) denotes underuse

As can be seen in Table 20, the frequency ratio of certainty devices in TACE was measured 41.91 per 10.000 words and the difference between two corpora is +6.33, which indicates an overuse in TACE against NACE. The analysis in concern also approved the underuse in TACE in the use of likelihood devices. Thus, these items were found 26.09 times in every 10.000 words in TACE whereas they were seen 34.70 times in NACE.

Frequencies in ESD categories in SACE were also compared with those in NACE. Table 21 displays the frequencies of ESDs per 10.000 words in both corpora.

Table 21

Overused/underused Certainty and Likelihood Devices in SACE in comparison with NACE

Category	N	SACE n/ 10.000	N	NACE n/ 10.000	overuse/ underuse + / -
Certainty	2611	39.07	2389	35.58	+3.49
Likelihood	1722	25.77	2330	34.70	-8.93

n= raw frequency of epistemic categories

n/ 10.000= frequency of epistemic categories per 10.000 words

(-/+) = difference between relevant value and value in NACE per 10.000 words; (+) denotes overuse; (-) denotes underuse

Table 21 indicates that certainty devices were overused in SACE while likelihood devices were underused in it. Namely, certainty devices were seen 39.07 times per

10.000 words in SACE whereas they appeared 35.58 times in NACE. As for likelihood devices, they were employed 25.77 times in every 10.000 in SACE and 34.70 times in NACE. Lastly, TACE and SACE were compared in terms of ESD categories per 10.000 words. The results are provided in Table 22.

Table 22

Overused/underused Certainty and Likelihood Devices in TACE in comparison with SACE

Category	TACE N	TACE n/ 10.000	SACE N	SACE n/ 10.000	overuse/ underuse + / -
Certainty	2829	41.91	2611	39.07	+2.84
Likelihood	1626	26.09	1722	25.77	+0.32

n= raw frequency of epistemic categories

n/ 10.000= frequency of epistemic categories per 10.000 words

(-/+) = difference between relevant value and value in SACE per 10.000 words; (+) denotes overuse; (-) denotes underuse

It is revealed that both certainty and likelihood devices were slightly overused by TAEs against SAEs. For instance, certainty devices were found approximately 42 times in every 10.000 words in TACE while they occurred less 39.07 times in SACE. Similarly, likelihood devices were seen over 26 times per 10.000 words in TACE whereas they appeared approximately 26 times in SACE, which shows likelihood devices were less overused than certainty devices by TAEs in comparison to SAEs.

A log likelihood test was administered to three corpora to find out whether these differences are statistically significant or not. Table 23 includes the related results.

Table 23

LL Ratio of Certainty and Likelihood Devices in TACE and NACE

Category	TACE n	NACE n	LL Ratio (*p< 0.05)
Certainty	2829	2389	+34.83
Likelihood	1626	2330	-129.74

n=raw frequency of ESDs in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As seen in Table 23, certainty devices were significantly overused in TACE against NACE, which was confirmed by a +34.83 LL value. Likelihood devices, on the contrary, reveal to be underused in TACE; thus, it appeared 1626 times in TACE and 2330 times in NACE. The difference was again approved by LL results ($p < 0.05$). The results of the analysis made between SACE and NACE are given in Table 24.

Table 24

LL Ratio of Certainty and Likelihood Devices in SACE and NACE

Category	SACE n	NACE n	LL Ratio (*$p < 0.05$)
Certainty	2611	2389	+10.96
Likelihood	1722	2330	-88.68

n=raw frequency of ESDs in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

As in the case between TACE and NACE, the overuse concerning the use of certainty devices and the underuse concerning that of likelihood devices in SACE against NACE were confirmed by +10.96 and -88.68 LL values.

LL ratio of categorical frequencies was measured for the non-native corpora and the related results can be seen in Table 25.

Table 25

LL Ratio of Certainty and Likelihood Devices in TACE and SACE

Epistemic Categories	TACE n	SACE n	LL Ratio (*$p < 0.05$)
Certainty	2829	2611	+6.67
Likelihood	1626	1722	-3.81

n=raw frequency of ESDs in corpus

+ indicates overuse in TACE relative to SACE

- indicates underuse in TACE relative to SACE

As seen in Table 25, the overuse of certainty devices and underuse of likelihood devices in TACE against SACE were approved with +6.67 and -3.81 LL values, respectively.

The following section is intended to cover certainty and likelihood devices across three corpora in more detail. That is to say that the distribution of epistemic categories (i.e. epistemic adjectives, epistemic adverbs, epistemic nouns and epistemic verbs) over three corpora will be discussed throughout it.

4.1.2.1. Certainty Devices in Three Corpora

Certainty devices include the following epistemic categories: epistemic adjectives (epistemic adjectives controlling *that*- clauses and epistemic adjectives controlling *to*-clauses), epistemic adverbs, epistemic nouns controlling *that*- clauses and epistemic verbs controlling *that*- clauses. Following Biber (2006), types belonging to these categories are given in Table 26.

Table 26

ESDs in Certainty Categories

Certainty Category	ESD Types
Adj.+ *that*-clause	*apparent, certain, clear, confident, convinced, correct, evident, false, impossible, inevitable, obvious, positive, right, sure, true, well-known*
Adj.+ *to*-clause	*apt, certain, due, guaranteed, prone, sure*
Adverbs	*actually, always, certainly, definitely, indeed, inevitably, in fact, never, of course, obviously, really, undoubtedly, without doubt, no doubt*
Noun+ *that*-clause	*assertion, conclusion, conviction, discovery, doubt, fact, knowledge, observation, principle, realization, result, statement*
Verb+ *that*-clause	*conclude, demonstrate, determine, discover, find, know, learn, mean, notice, observe, prove, realize, recognize, remember, see, show, understand*

Source: Biber, 2006, p. 92-93

Non-native corpora were compared to native corpus regarding a total number of above-mentioned 65 types and the results are illustrated in Figure 14.

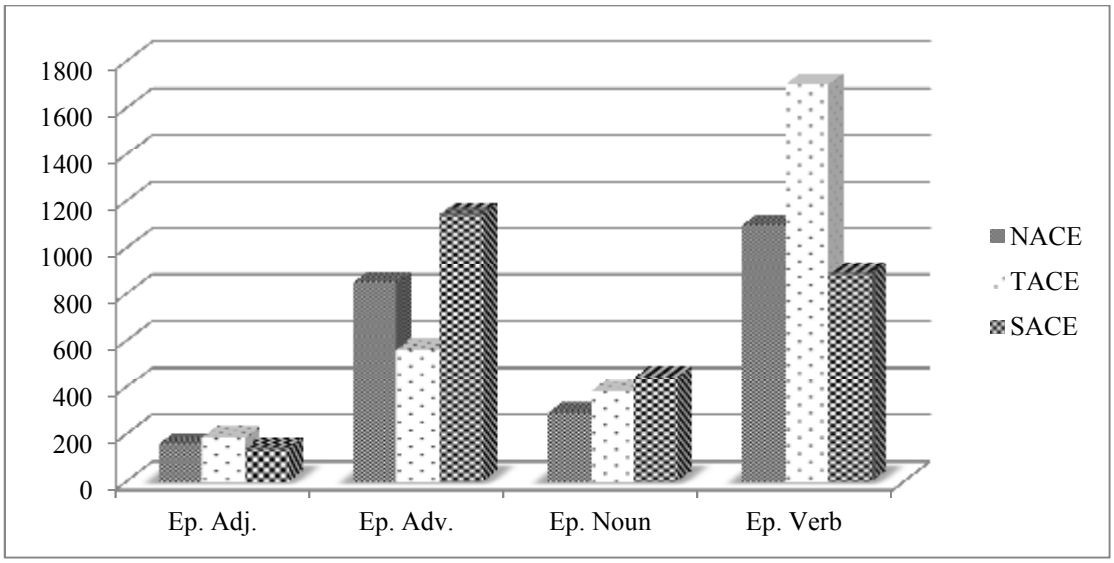


Figure 14. Certainty categories in three corpora

As indicated in Figure 14, certainty verbs, which are epistemic verbs controlling *that*-clauses, reveal to be most frequently used category in NACE and TACE appearing 1095 and 1700 times, respectively. Certainty adverbs are the mostly used epistemic items in SACE and the second most frequented ones in NACE and TACE. Three corpora are alike in the choice of the other devices; that is, epistemic nouns constitute the third and epistemic adjectives the least frequented categories in three corpora. Nonetheless, three corpora differ from each other with respect to the frequency of these items, which is approved via LL calculation. Table 27 presents the results of log-likelihood test administered to TACE and NACE.

Table 27

LL Ratio of Certainty Categories in TACE and NACE

Certainty Category	TACE n	NACE n	LL Ratio (*p< 0.05)
Ep. Adjective	185	158	+1.99
Ep. Adverb	563	848	-59.50
Ep. Noun	381	288	+12.48
Ep. Verb	1700	1095	+128.79

n=raw frequency of certainty categories in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As suggested in Table 27, both overused and underused categories in TACE in compared to NACE were confirmed by LL results. Epistemic verbs for instance, were preferred much more frequently by TAEs than NAEs, which was approved with a +128.79 LL value. The following are the expressions obtained from two corpora.

Example 1.

[However, the literature incorporates studies **showing** that the level of autonomy changes in accordance with the economic status of learners.]

Extracted from <TACE-CU-2011-MB>

[The analysis of the first reflection sheets **showed** that students were aware of the fact that they had to take part in such an implementation so as to gain some specific skills.

Extracted from <TACE-CU-2006-NK>

Likewise, epistemic adverbs proved to be significantly underused in TACE, which was again confirmed by LL test result ($p < 0.05$). The following are the expressions taken from NACE.

Example 2.

[The verbal nature of the questionnaire delivery **certainly** discouraged opportunities to return to an earlier question and reconsider the first response given.]

[The dominance of Urdu over other Pakistani heritage languages was **perhaps** best highlighted by a statement made to the researcher by a male Pakistani heritage teacher.]

[The comments appended to this question suggested that some staff might **indeed** be adopting an unofficial 'wait and see policy'.]

Extracted from <NACE-NU-2006-CS>

Table 28 displays the results of LL test applied to SACE and NACE in order to see whether the categorical differences found between two corpora are statistically significant or not.

Table 28

LL Ratio of Certainty Categories in SACE and NACE

Certainty Category	SACE n	NACE n	LL Ratio (*p< 0.05)
Ep. Adjective	139	158	-1.13
Ep. Adverb	1143	848	+45.30
Ep. Noun	441	288	+33.09
Ep. Verb	888	1095	-20.66

n=raw frequency of certainty categories in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

As illustrated in Table 28, both overused and underused categories in SACE in compared to NACE were confirmed by LL results. For instance, epistemic adverbs were employed more frequently by SAEs than NAEs, which was approved with a +45.30 LL value. The following are the extracts taken from two corpora.

Example 3.

[The data obtained from teachers and students through questionnaires revealed that, **in fact**, both groups tend to believe that more progress is made in speaking skills in intensive courses than in extensive courses.]

Extracted from <SACE-UB-2007-RSS>

[In addition to this, my expectations were fulfilled with UNI students, the highest achievers, **in fact**, within the scope of upper-intermediate learners (2000 words).]

Extracted from <SACE-USC-2009-RFC>

Similarly, epistemic verbs revealed to be significantly underused in SACE, which was confirmed by LL test result (p< 0.05). The following are the statements obtained from NACE.

Example 4.

[Research has **shown** that genres such as letters, news and fiction have become more informal and involved since the eighteenth century, while genres such as science and medicine have become more “literate“ and formal.]

Extracted from <NACE-UG-20010-CW>

[Andrew reflected on the idea of teacher accountability, and how illuminating it was for him to see teachers in action working on **showing** that they are indeed reaching benchmarks and teaching according to the standards in place.]

Extracted from <NACE-UI-2012-EJE>

Table 29 shows the results of LL test applied to TACE and SACE to find out whether the categorical differences between two corpora are statistically significant or not.

Table 29

LL Ratio of Certainty Categories in TACE and SACE

Certainty Category	TACE n	SACE n	LL Ratio (*p< 0.05)
Ep. Adjective	185	139	+6.09
Ep. Adverb	563	1143	-207.10
Ep. Noun	381	441	-5.01
Ep. Verb	1700	888	+250.95

n=raw frequency of certainty categories in corpus

+ indicates overuse in TACE relative to SACE

- indicates underuse in TACE relative to SACE

Table 29 exhibits that overused and underused categories in TACE against SACE were confirmed by LL results. For example, epistemic verbs were employed much more frequently by TAEs than SAEs, which was approved with a +250.95 LL value. The following expressions are obtained from TACE.

Example 5.

[In the formal surface changes category, it can be **seen** that the control group students have made slightly more changes compared to the experimental group students.]

Extracted from <TACE-METU-2009-ETD>

[Nevertheless, when the pre-test/post-test results for vocabulary recognition are taken into consideration, we can **see** that the interview results support the test results.]

Extracted from <TACE-CU-2011-IU>

Likewise, that certainty adverbs were significantly underused in TACE against SACE was approved with -207.10 LL value. The following are the statements taken from SACE.

Example 6.
[Jazz and the blues idiom have **always** been highly protean aesthetics when it comes to the field of exegesis, both at a musical and at a cultural level.]

Extracted from <SACE-UCM-2012-CAR>

[Additionally, it is likely the case that there is a collocation effect, after all, fries in the context of food is almost **always** pluralized.]

Extracted from <SACE-UI-2011-GSCD>

The following section will describe and outline the likelihood devices, which include four major categories, across three corpora.

4.1.2.1. Likelihood Devices in Three Corpora

Likelihood devices contain four main epistemic categories, which are epistemic adjectives (epistemic adjectives controlling *that*- clauses and epistemic adjectives controlling *to*- clauses), epistemic adverbs, epistemic nouns controlling *that*- clauses, and epistemic verbs (epistemic verbs controlling *that*- clauses and epistemic verbs controlling *to*- clauses). Table 30 provides the epistemic types falling into these categories.

Table 30

ESDs in Likelihood Categories

Likelihood Category	ESD Types
Adj.+ *that*-clause	*Doubtful, likely, possible, probable*
Adj.+ *to*-clause	*Liable, likely, unlikely*
Adverb	*Apparently, evidently, kind of, in most cases/instances, perhaps, possibly, predictably, probably, roughly, sort of, maybe*
Noun+ *that*-clause	*Assumption, belief, claim, contention, feeling, hypothesis, idea, implication, impression, notion, opinion, possibility, presumption, suggestion*
Verb+ *that*-clause	*Assume, believe, doubt, gather, guess, hypothesize, imagine, predict, presuppose, presume, reckon, seem, speculate, suppose, suspect, think*
Verb+ *to*-clause	*Appear, happen, seem, tend*

NACE, TACE and SACE were compared to each other in order to see whether there is a significant difference across them in the choice of likelihood categories. The results are illustrated in Figure 15.

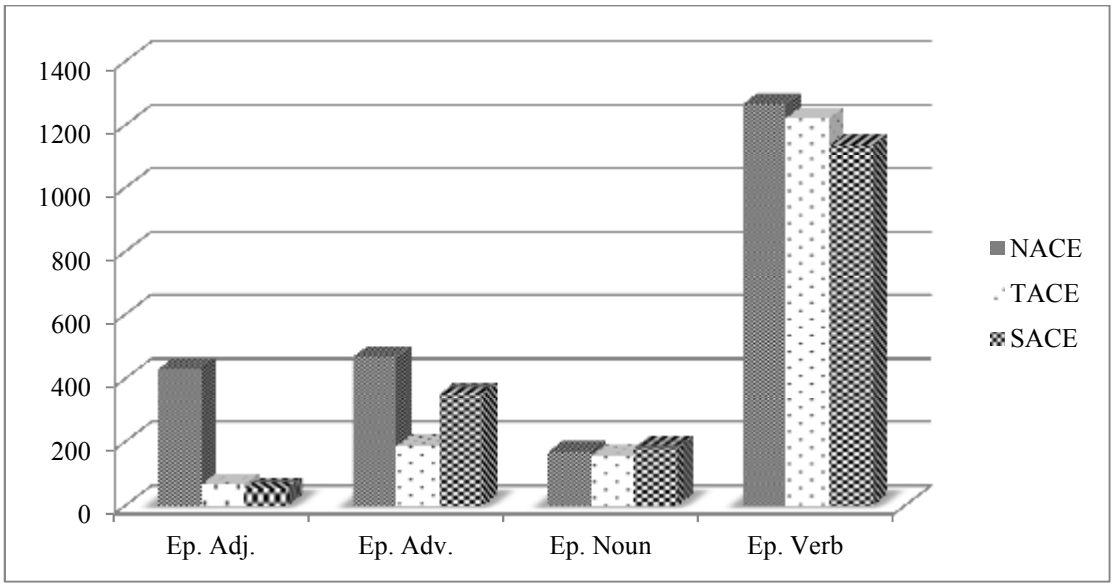


Figure 15. Likelihood categories in three corpora

As shown in Figure 15, three corpora are identical in likelihood categories of the highest and second highest frequencies. Thus, epistemic verbs are the most extensively used category (NACE: 1265; TACE: 1220; SACE: 1134) and epistemic adverbs the second most frequently used one in three corpora (NACE: 467; SACE: 353; TACE: 186). Non-native corpora differ from the native corpus in the least preferred category;

namely, TAEs and SAEs seemed to have employed epistemic adjectives less frequently than other categories in their writing while these items were found more frequently than epistemic nouns in NACE. Epistemic adjectives, for example, occurred 428 times in the native corpus whereas they were seen 66 and 53 times in TACE and SACE, respectively. Epistemic nouns, on the other hand, appeared 182 times in SACE, 167 in NACE and 154 in TACE. A log likelihood test was applied to find out whether the differences found across three corpora are statistically significant. Results of the first analysis, which was administered between TACE and NACE, are displayed in Table 31.

Table 31

LL Ratio of Likelihood Categories in TACE and NACE

Likelihood Category	TACE n	NACE n	LL Ratio (*p< 0.05)
Ep. Adjective	66	428	-298.31
Ep. Adverb	186	467	-126.47
Ep. Noun	154	167	-0.60
Ep. Verb	1220	1265	-1.07

n=raw frequency of likelihood categories in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As illustrated in Table 31, overuse concerning all likelihood categories in TACE against NACE was confirmed through LL calculation. Epistemic adjectives, for example, were employed considerably more often by NAEs against TAEs, which was approved with a -298.31 LL value. Two of them are illustrated in the following extracts drawn from NACE.

Example 7.

[It also construes a relatively common social activity that is **likely to** be approached by the reader in a perfunctory way rather than as something they devote a lot of time and attention to.]

Extracted from <NACE-UM-2009-CES>

[Such a recommendation however is not easily followed as such a study design would be costly, time consuming, **liable to** participant drop-out and a considerable commitment for the participants involved.]

Extracted from <NACE-NU-2009-CM>

Likewise, epistemic adverbs were significantly underused by TAEs against NAEs, which was confirmed with -126.47 LL value. Epistemic nouns and epistemic verbs were also underused by TAEs; however, the difference was not found so significant as the ones measured between two corpora with regard to epistemic adjectives and epistemic adverbs.

Finally, a log likelihood analysis was administered between SACE and NACE and the results are displayed in Table 32.

Table 32

LL Ratio of Likelihood Categories in SACE and NACE

Likelihood Category	SACE n	NACE n	LL Ratio (*p< 0.05)
Ep. Adjective	53	428	-331.28
Ep. Adverb	353	467	-15.36
Ep. Noun	182	167	+0.72
Ep. Verb	1134	1265	-6.54

n=raw frequency of likelihood categories in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

As seen in Table 32, both overused and underused categories in SACE against NACE were confirmed by LL results. As in the case of TACE, epistemic adjectives revealed the mostly underused items among likelihood devices in SACE. They occurred in NACE over than ten times as many in SACE. Similarly, The following are the statements extracted from NACE to illustrate one of the items in concern.

Example 8.

[It is **possible that** this is reflective of the chat rooms nature as English language practising rooms.]

Extracted from <NACE-NU-2011-AB>

[It is therefore **likely that** the repetition task which DEAP effectively became for them was in fact a NWR task as they had no lexical representations available to assist them.]

Extracted from <NACE-NU-2006-CS>

Epistemic adverbs and epistemic verbs are the other likelihood categories that SACE and NACE differ in use; however, it would not be sufficient to specify them as significant. Epistemic nouns constitute the only category that was overused in SACE against NACE. Yet, it is found much less significant when compared to the underuse measured between two corpora with respect to other likelihood categories. Lastly, log likelihood categories of likelihood categories in TACE against SACE are provided in Table 33.

Table 33

LL Ratio of Likelihood Categories in TACE and SACE

Likelihood Category	TACE n	SACE n	LL Ratio (*p< 0.05)
Ep. Adjective	66	53	+1.29
Ep. Adverb	186	353	-54.31
Ep. Noun	154	182	-2.63
Ep. Verb	1220	1134	+2.33

n=raw frequency of likelihood categories in corpus

+ indicates overuse in TACE relative to SACE

- indicates underuse in TACE relative to SACE

As shown in Table 33, likelihood adjectives and likelihood verbs were slightly overused by TAEs against SAEs, which was approved by LL calculation. In response, likelihood nouns were slightly underused in TACE against SACE, which was approved by -2.63 LL value. Likelihood adverbs, on the other hand, appeared in TACE considerably less than SACE. Namely, they were used by SAEs twice as frequently as TAEs. In conclusion, they have revealed mostly underused devices in TACE against SACE, which was confirmed with -54.31 LL value. Following extracts drawn from SACE are intended to illustrate one of the most underused likelihood adverbs in TACE against SACE.

Example 9.

[A native speaker of Danish reading the words Lamine Revère would **probably** not be able to relate them to his/her language.]

[For the Russian listener, **probably** the cultural assumption that best fits this joke is the well-known quality of Armenian brandy.]

[But, all things considered, the translatability of this joke is **probably** very limited because of all the cultural aspects that operate within it.]

Extracted from <SACE-UPF-2012-JFMB>

Section 4.1.3 is intended to examine epistemic stance devices across three corpora.

4.1.3. Individual Epistemic Stance Devices

ESDs were examined individually in the data they appear. Initially, top ten ESDs in NACE have been determined and are presented in Table 34.

Table 34

Most Frequently Used 10 ESDs in NACE

ESD	n	T/t %	n per 10.000
Show	502	0.1	7.5
Seem	488	0.1	7.3
Appear	320	0.0	4.8
Likely	302	0.0	4.5
Fact	235	0.0	3.5
Tend	225	0.0	3.4
Perhaps	220	0.0	3.3
Always	155	0.0	2.3
Indeed	126	0.0	1.9
Mean	119	0.0	1.8
Total	2692	0.2	40.3

n= raw frequency of ESD in corpus

% T/t= Type/token ratio, percentage of the device in overall ESD types in corpus

In accordance with the frequency results, over 57% of total ESDs in NACE consist of top ten epistemic stance devices found in the corpus at stake. Namely, ESDs such as *show, seem* and *appear* occurred 40.3 times in every 10.000 words in NACE. *Show* is the most extensively used in NACE occurring approximately 8 times in every 10.000 words in the corpus. The following are the expressions drawn from a single work in NACE, in which show was employed 18 times.

Example 10.

[However, in analysing various types of OIR sequences, the present study has **shown** that this doesn't quite appear to be the case in this corpus of online chat room interaction.]

[Analysis of such sequences again **showed** that repair dealt with the possibility of trouble in hearing or understanding.]

[In Section 6.2, it was first **shown** that one way to treat a nonresponse is to check for the presence and recipiency of the intended respondent.]

Extracted from <NACE-NU-2011-AB>

Secondly, the most frequently used ESDs in TACE have been obtained and are given in Table 35.

Table 35

Most Frequently Used 10 ESDs in TACE

ESD	n	T/t %	n per 10.000
Show	616	0.1	9.1
Seem	549	0.1	8.1
Fact	342	0.1	5.1
See	223	0.0	3.3
Believe	210	0.0	3.1
Mean	187	0.0	2.8
Tend	168	0.0	2.5
Observe	151	0.0	2.2
Think	129	0.0	1.9
Find	116	0.0	1.7
Total	2691	0.3	39.8

n= raw frequency of ESD in corpus

% T/t= Type/token ratio, percentage of the device in overall ESD types in corpus

The most frequented 10 ESDs in TACE constitute 60.4% of all ESDs occurring in the corpus at stake. As illustrated in Table 35, these devices were found approximately 40 times in every 10.000 words in TACE whereas they occurred 39.6 times per 10.000 words in NACE. Five ESDs (*fact, mean, seem, show* and *tend*) are identical in the two corpora. However, ESDs in TACE such as *believe, find, observe, see* and *think* do not appear in top ten ESD list of NACE. It was revealed that four out of five identical ESDs

were overused by TAEs. For example, *show* was employed 616 times in TACE, which corresponds to 9.1 times in every 10.000 words in it whereas it was found 7.5 times per 10.000 words in NACE. It is interesting that *appear,* the third most frequent ESD in NACE, does not exist in the top ten list of TACE. Similarly, *always, indeed, likely,* and *perhaps* which are included in the top ten ESD list in NACE are not found in the counterpart list of TACE.

In order to test the significance of mostly used ESDs in TACE, LL ratio was administered comparing it with NACE. Frequency differences of the ESDs between the corpora in question were confirmed by the statistical analysis. Table 36 displays the ratio of overused ESDs in TACE against NACE.

Table 36

LL ratio of Overused ESDs in TACE

ESD	TACE n	NACE n	LL Ratio (*p< 0.05)
Believe	210	31	+148.16
Observe	151	18	+118.84
Think	129	20	+88.46
See	229	70	+88.22
Understand	49	3	+48.90
Fact	342	235	+19.39
Obvious	26	4	+17.91
Realize	75	33	+16.55
Mean	186	116	+16.00
Idea	65	27	+15.97
Conclude	109	58	+15.56
(Im)probable	21	3	+15.09
Prove	22	4	+13.62
Apparent	18	3	+11.81
Opinion	15	2	+11.18
Show	616	502	+11.04
Clear	81	49	+7.79
Probably	91	61	+5.80
Evident	18	7	+4.95
Obviously	59	38	+4.47
Conclusion	12	4	+4.14
Seem	549	488	+3.27
Guess	2	0	+2.76
Well-known	2	0	+2.76
Recognize	17	9	+2.46
Find	116	96	+1.78
Presume	1	0	+1.38
Notice	19	13	+1.10
Apt	6	3	+1.00

(Table 36 Continued)

Statement	10	7	+0.52
Suppose	5	3	+0.49
In most cases/ instances	6	4	+0.39
Know	36	34	+0.05
Really	50	48	+0.03
Assumption	23	22	+0.02

n=raw frequency of ESDs in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As seen in Table 36, the highest LL value belongs to *believe* with +148.16 value, which indicates a significant difference between TACE and NACE in terms of frequency and statistics. *Observe*, another overused ESD, placed in the second range in TACE with 118.84 plus value. *Think, see* and *understand* are some of the other overused ESDs in TACE with high LL values and statistical differences (*think: +88.46; see:* +88.22; *understand*: +48.90). The following example is intended to illustrate the use of *believe*, which was found 36 times in a single work by a TAE:

Example 11.

[Almost half of all participants seemed to **believe** that the most important part of learning English is learning how to translate from Turkish to English.]

[These two items posit contradicting assertions: If students do not **believe** that some people have a special ability for learning foreign languages, then they could be expected to **believe** that everybody can learn a foreign language.]

[Also, it is possible to **believe** that some people have a special ability for learning foreign languages and at the same time that ability or talent is not so important because anyone can learn a foreign language if they work hard enough.]

Extracted from <TACE-CU-2010-SB>

Similarly, the following examples are given to display another overused type *observe*, which was preferred 10 times in a single work by TAE.

Example 12.

[It can be easily **observed** that she referred to going back and evaluating what she did during the teaching process.]

[It was also **observed** that the student teachers made good use of the opportunities the study gave them to engage in self-directed development.]

[In her second teaching video, it was frequently **observed** that she became mostly aware of her teaching practice as time went by.]

Extracted from <TACE-GU-2010-CB>

LL measurement was applied to test the underused ESDs in accordance with frequency analysis. Table 37 shows the underused ESDs in TACE in comparison with NACE.

Table 37

LL Ratio of Underused ESDs in TACE

ESD	TACE n	NACE n	LL Ratio (*p< 0.05)
Likely	36	302	-240.72
Perhaps	12	220	-228.28
Appear	69	320	-176.98
Possible	9	93	-80.97
Predict	0	42	-58.45
Certainly	10	67	-47.58
Unlikely	0	29	-40.36
Possibly	12	59	-34.17
Possibility	5	29	-18.87
Always	95	155	-14.86
Indeed	74	126	-13.96
Prone	7	28	-13.60
Observation	1	14	-13.52
In fact	67	115	-13.07
Suggestion	4	21	-12.76
Predictably	0	9	-12.52
Demonstrate	62	105	-11.43
True	1	12	-11.03
Apparently	19	45	-11.01
No doubt	3	17	-10.89
Of course	42	74	-9.11
Tend	168	225	-8.60
Suspect	0	6	-8.35
Sure	12	29	-7.36
Discover	0	5	-6.96
Speculate	8	21	-6.11
Notion	0	4	-5.57
Hypothesize	9	21	-5.00
Guaranteed	0	3	-4.17
Kind of	0	2	-2.78
Inevitably	6	13	-2.68
Knowledge	6	13	-2.68

(Table 37 Continued)

Hypothesis	12	21	-2.53
Determine	1	4	-1.94
Happen	3	7	-1.67
Liable	0	1	-1.39
Roughly	10	25	-1.43
Impossible	0	1	-1.39
Presuppose	0	1	-1.39
Discovery	0	1	-1.39
Doubt	0	1	-1.39
Remember	7	12	-1.36
Assume	66	78	-1.07
Never	65	73	-0.51
Undoubtedly	5	7	-0.35
Confident	1	2	-0.35
Certain	6	8	-0.30
Maybe	15	18	-0.29
Sort of	17	20	-0.26
Convinced	2	3	-0.21
Implication	3	4	-0.15
Result	4	5	-0.12
Impression	5	6	-0.10
Assertion	6	7	-0.08
Definitely	11	11	-0.00
Evidently	4	4	-0.00
Feeling	3	3	-0.00
Imagine	1	1	-0.00

n=raw frequency of ESDs in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

According to LL ratio of underused ESDs, a total number of 60 ESDs have revealed a difference in TACE against NACE. *Likely* and *perhaps* are the most and the second most underused ESDs with -240.72 LL and -231.61 LL values, respectively. It is noteworthy that most of the top ten underused ESDs in TACE are epistemic adverbs and epistemic adjectives. More specifically, there are just two epistemic verbs (*appear* and *predict*) and an epistemic noun (*possibility*) among the items in concern. Following this, top ten ESDs in SACE have been determined and given in Table 38.

Table 38

Most Frequently Used 10 ESDs in SACE

ESD	n	T/t %	n per 10.000
Seem	646	0.0	9.7
Fact	411	0.0	6.2
Show	314	0.0	4.7
In fact	224	0.0	3.4
Always	171	0.0	2.6
Indeed	154	0.0	2.3
Actually	154	0.0	2.3
Probably	143	0.0	2.1
Tend	141	0.0	2.1
Appear	133	0.0	2.0
Total	2491	0.1	37.4

n= raw frequency of ESD in corpus

% T/t= Type/token ratio, percentage of the device in overall ESD types in corpus

As can be observed in Table 38, top ten ESDs cover 57.5% of all ESDs in SACE and they occurred 37.4 times per 10.000 words in this corpus. Seven ESDs are identical in SACE with NACE (*always, appear, fact, indeed, seem, show* and *tend*). ESDs such as *actually, in fact* and *probably* do not appear in top ten list of NACE, which might mean they were not significantly overused by SAEs. All common ESDs differ in their frequencies across the corpora in concern. For instance, *seem* has revealed to be the most frequently used ESD by SAEs (646 times) whereas it is the second most preferred item by NAEs (488 times). Similarly, *fact*, the second most frequented ESD in SACE (411 times), appears the fifth item in the most frequently used ESD list of NACE (235 times). The corpora in concern were compared via LL Ratio to test the ESDs that seem to be overused and underused in accordance with the frequency results in SACE. Table 39 shows LL ratio of the overused items in this corpus.

Table 39

LL ratio of Overused ESDs in SACE

ESD	SACE n	NACE n	LL Ratio (*p< 0.05)
Fact	411	235	+49.41
Observe	79	18	+41.70
Idea	92	27	+37.83
In fact	224	115	+36.20
Probably	143	61	+34.31
Seem	646	488	+22.85
Really	102	48	+20.14
Think	55	20	+17.15
Contention	12	0	+16.69
Prove	23	4	+14.87
Actually	154	99	+12.32
Notice	35	13	+10.58
Believe	61	31	+10.11
Obviously	61	38	+5.50
Definitely	23	11	+4.39
In most cases/ instances	12	4	+4.22
Never	104	78	+3.85
Indeed	154	126	+2.94
Presume	2	0	+2.78
Belief	17	9	+2.54
Undoubtedly	14	7	+2.41
Happen	14	7	+2.41
Convinced	8	3	+2.38
True	20	12	+2.06
Evidently	9	4	+2.00
Conclusion	9	4	+2.00
Remember	19	12	+1.63
Without doubt	1	0	+1.39
Conviction	1	0	+1.39
Apt	6	3	+1.03
Always	171	155	+0.86
Sort of	25	20	+0.58
Implication	6	4	+0.41
Liable	2	1	+0.34
Understand	5	3	+0.52
Imagine	2	1	+0.34
Certainly	72	67	+0.18
Apparent	4	3	+0.15
Maybe	20	18	+0.12
Recognize	10	9	+0.06
Evident	7	7	+0.00
Doubt	1	1	+0.00
Presuppose	1	1	+0.00
Due	0	0	+0.00

n=raw frequency of ESDs in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

It is interesting to see that most of the overused ESDs in frequency table for SACE do not exist in the top ten ESD list of the corpus in concern. As presented in Table 39, *fact* is the most overused type in SACE against NACE with +49.41 LL value even though *seem* reveals to be the most frequently used ESD across SACE. However, ESDs such as *observe, idea, probably, really, think, prove* and *contention* which are not seen in the top ten ESDs in SACE are among the most overused items in SACE in contrast to NACE by means of their LL value. The related value of *always* has not indicated a significant overuse in SACE when compared to NACE although it is in the top ten ESD list in SACE. Like in TACE, *fact*, o*bserve* and *think* are some of the most overused ESDs in SACE in accordance with LL ratio. The example presented below was taken from a single work in SACE where the most overused item in SACE occurred 13 times:

Example 13.

[The tangential essence lies in the **fact** that most of these schools accept myth to be intrinsically bound to some form of plot, in the Aristotelian sense.]

[On the contrary, the concealment of the proof that those binarisms were once vibrant in jazz discourse attests to the **fact** that they still remain so.]

[The **fact** that the jazz literary canon exalts black writers over white ones is, for the most part, an accepted and welcomed premise.]

Extracted from <SACE-UCM-2012-CAR>

Another overused ESD in SACE is *probably,* which appeared 11 times in a single work from the corpus in concern.

Example 14.

[The punch line in this joke inheres in combining fundamentalism. Saudi Arabia—the spiritual place of the Islamic world and **probably** one of the most conservative Muslim countries—the Arabic language and off-colour humour.]

[Because of the topic chosen, it is a joke that is **probably** typical of children, and since it is inoffensive in its configuration and its main intention is to surprise and entertain the listener, its sounds can remind a Catalan speaker of an African language.]

[As "doping" is an English word borrowed by many languages, it can also **probably** be adapted to different purposes even though the degree of the joke's humorousness might be different from culture to culture.]

Extracted from <SACE-UPF-2012-JFMB>

In this single work, *probably* was used 11 times to convey the meaning of likelihood. ESDs significantly underused in SACE against NACE are outlined in Table 40.

Table 40

LL ratio of Underused ESDs in SACE

ESD	SACE n	NACE n	LL Ratio (*p< 0.05)
Likely	37	302	-234.96
Perhaps	46	220	-122.93
Appear	133	320	-78.66
Possible	11	93	-73.57
Show	314	502	-42.81
Demonstrate	42	105	-27.59
Unlikely	2	29	-28.01
Of course	28	74	-21.29
Tend	141	225	-19.05
Predict	12	42	-17.51
Speculate	3	21	-15.10
Suggestion	6	21	-8.75
Assume	46	78	-8.20
Predictably	1	9	-7.32
Possibility	12	29	-7.18
Result	0	5	-6.91
Certain	1	8	-6.16
Notion	0	4	-5.53
Possibly	36	59	-5.51
Assertion	1	7	-5.03
Realize	18	33	-4.41
Know	19	34	-4.23
Guaranteed	0	3	-4.14
Hypothesis	10	21	-3.94
No doubt	8	17	-3.27

(Table 40 Continued)

See	50	70	-3.25
Sure	17	29	-3.11
Hypothesize	11	21	-3.13
Opinion	0	2	-2.76
Learn	4	10	-2.63
Mean	95	119	-2.58
Observation	7	14	-2.35
Knowledge	7	13	-1.80
Impossible	0	1	-1.38
Discovery	0	1	-1.38
Realization	0	1	-1.38
Claim	15	22	-1.30
Find	81	96	-1.20
Roughly	18	25	-1.11
Feeling	1	3	-1.04
(Im)probable	1	3	-1.04
Assumption	16	22	0.92
Impression	4	6	-0.39
Suspect	4	6	-0.39
Confident	1	2	-0.34
Statement	5	7	-0.33
Suppose	2	3	-0.20
Prone	25	28	-0.16
Obvious	3	4	-0.14
Apparently	43	45	-0.04
Clear	47	49	0.03

n=raw frequency of ESDs in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

The highest ratio was calculated for *likely* as -234.96, which is followed by *perhaps* with -122.93 and *appear* with -78.66 LL values. Interestingly, *show* and *tend* which occur in mostly frequented ESDs in SACE have shown underuse in SACE against NACE in terms of LL ratio (*show*: -42.81; *tend*: -19.05).

The following section will offer the results of the analyses administered to three corpora in order to reveal whether they significantly differ in the use of four lexical categories, which are *adjectives, adverbs, nouns* and *verbs*.

4.1.4. Lexical Categories in Three Corpora

Each corpus was analysed to investigate four lexical categories as *adjectives, adverbs, nouns* and *verbs* using software package programmes (i.e. VMatrix and Log

Likelihood Calculator). First of all, it is seen that approximately half of each corpus is composed of the items belonging to these categories (NACE: 50%; TACE: SACE: 41%). Figure 16 displays frequency distribution of lexical categories across three corpora.

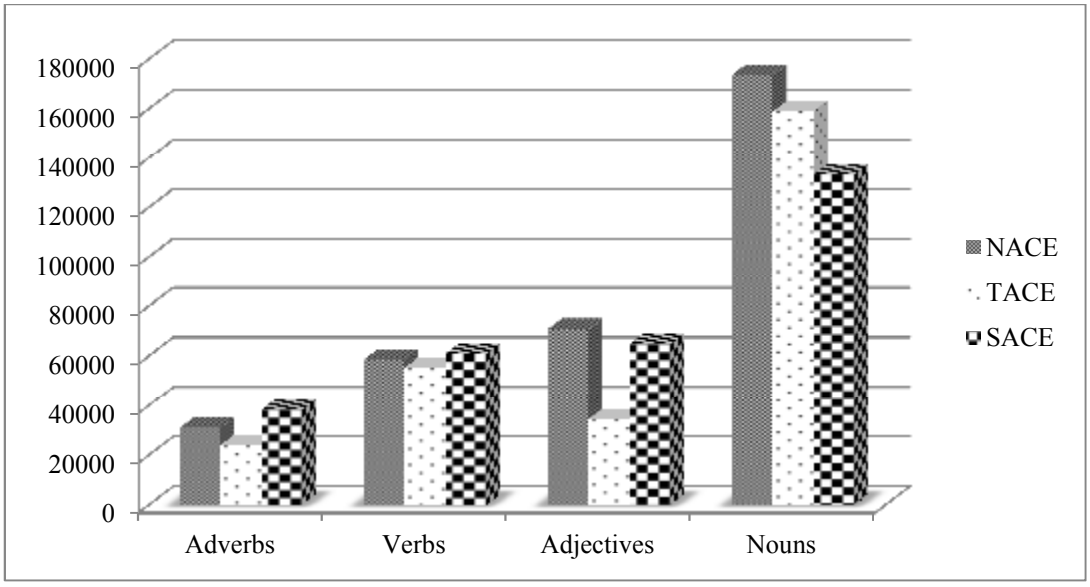


Figure 16. Lexical categories across three corpora

It is not much surprising that nouns are the most frequently used tokens in all corpora (NACE: 25.86%; TACE: 23.56%; SACE: 19.97%) and that adverbs are the least frequently used ones over the corpora in concern (NACE: 4.61%; TACE: 3.54%; SACE: 5.73%). However, there has revealed several interesting differences across three corpora in terms of the distribution of adjectives and verbs. In NACE, for instance, verbs were relatively less used than adjectives (Verbs: 8.71%; Adjectives: 10.54%). On the contrary, in TACE, verbs were employed more frequently than adjectives (Verbs: 8.27%; Adjectives: 5.10%). Lastly, in SACE, adjectives were slightly more frequented than verbs (Adjectives: 9.70%; Verbs: 9.12%). At this point, percentages of each category were calculated in accordance with the amount of lexical categories seen in each corpus since it was found more functional to examine lexical categories in their own right rather than across the entire corpora. The related results are provided in Table 41.

Table 41

Frequency Distribution of Lexical Categories in Three Corpora

	NACE		TACE		SACE	
Category	n	%	n	%	n	%
Adverbs	30955	9.27	23914	8.77	38309	12.88
Verbs	58491	17.52	55288	20.28	60971	20.49
Adjectives	70786	21.20	34405	12.62	64827	21.79
Nouns	173647	52.01	159059	58.33	133421	44.84

n=raw frequency of each category

%=percentage of each category to overall frequency of lexical categories (in-group T/t ratio)

In accordance with the results, not much has changed with the frequency distribution of lexical tokens across lexical categories in three corpora. For instance, in parallel with the previous results, adverbs were the least and nouns are the most found tokens in lexical categories in three corpora. However, it is noteworthy that adjectives constitute nearly 22% of the lexical tokens in SACE while they have a slightly lower frequency in NACE, in which nouns appear as the largest lexical category when compared to other two corpora. It is also interesting that nouns compose approximately 60% of the lexical tokens in TACE.

The picture has not significantly changed for adverbs, as well. Namely, they consist nearly 13% of the lexical tokens in SACE, 9.27% of those in NACE and 8.77% of those in TACE.

Adjectives seem to be used less frequently than verbs in NACE (Adjectives: 10.54%; Verbs: 8.71%). It is also observed that verbs have relatively a higher percentage than adjectives within lexical tokens in TACE (Verbs: 20.28%; Adjectives: 12.62%). Finally, they seem to have similar proportions over the lexical categories in SACE (Adjectives: 21.79%; Verbs: 20.49%).

Native and non-native corpora required to be compared with respect to the frequency ratio of each lexical category for 10.000 to find out whether there is an overuse or underuse in lexical category frequency per 10.000 words in TACE and SACE against NACE. The results can be seen in Table 42.

Table 42

Overused/underused Lexical Categories in TACE in comparison with NACE

Lexical Category	TACE		NACE		overuse/ underuse
	N	n/ 10.000	n	n/ 10.000	+ / -
Adjectives	34405	509.65	70786	1054.19	-544.54
Adverbs	23914	354.24	30955	461.00	-106.76
Nouns	159059	2356.18	173647	2586.05	-229.87
Verbs	55288	818.99	58491	871.08	-52.09

n= raw frequency of epistemic categories

n/ 10.000= frequency of epistemic categories per 10.000 words

(-/+) = difference between relevant value and value in NACE per 10.000 words; (+) denotes overuse; (-) denotes underuse

As suggested in Table 42, the frequency ratio of verbs in NACE was calculated over 871 per 10.000 words and the difference between two corpora was found -52.09, which means they were used in NACE much more frequently than TACE. It is also seen that the other lexical categories were underused by TAEs in comparison to NAEs. For instance, adverbs appeared 7041 fewer times in TACE than NACE with a -106.76 value. Likewise, adjectives and nouns seem to be remarkably underused in TACE against NACE (Adjectives: -544.54; Nouns: -229.87), which means that they were preferred approximately 545 and 230 fewer times per 10.000 words by TAEs than NAEs. As a following step, LL ratio was calculated to make comparisons across corpora with regard to lexical tokens in question. The first LL analysis was administered to TACE and NACE to reveal the overused or underused categories. Table 43 provides the related results.

Table 43

LL Ratio of Lexical Categories in TACE and NACE

Lexical Category	TACE n	NACE n	LL Ratio (*p< 0.05)
Adjectives	34405	70786	-13040.32
Adverbs	23914	30955	-943.81
Nouns	159059	173647	-719.73
Verbs	55288	58491	-108.01

TACE is observed frequency in Corpus 1

NACE is observed frequency in Corpus 2

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As indicated in Table 43, LL ratio of categorical comparison has approved some of the categorical frequency differences in TACE against NACE. It is really interesting to see that all differences revealed to be underuse in the former corpus. Namely, a significant difference was observed between TACE and NACE in the category of adjectives with -13040.32 LL value and the least difference was found between the two corpora in the category of verbs with -108.01 LL value. The second and the third highest differences were seen in the category of adverbs and nouns, respectively (Adverbs: -943.81; Nouns: -719.73). To sum up, all the differences between TACE and NACE with respect to LL values of lexical categories were found statistically significant ($p < 0.05$). Another interesting finding might be that LL values counted for the lexical categories are inconsistent with those previously calculated for epistemic categories (See Table 17). For instance, LL values obtained from the analyses for epistemic verbs and epistemic nouns revealed to be overuse in TACE against NACE (Epistemic Verbs: +56.55; Epistemic Nouns: +6.05); however, those drawn from the analyses for verbs and nouns were observed as underuse in TACE against NACE (Verbs: -108.01; Nouns: -719.73). Similarly, the differences found between two corpora in the categories of epistemic adverbs and epistemic adjectives were not so high as those found in the categories of adverbs and adjectives, respectively (Epistemic Adverbs: -160.26; Epistemic Adjectives: -139.71). All in all, it can be concluded that ESD categories and lexical categories were preferred relatively more often by NAEs than TAEs. The same analyses were repeated to reveal whether there is an overuse or underuse in SACE against NACE regarding lexical categories and to determine whether the overuse or underuse values are statistically significant -if any. The results are exhibited in Table 44 and Table 45.

Table 44

Overused/underused Lexical Categories in SACE in comparison with NACE

		SACE		NACE	overuse/ underuse
Category	N	n/ 10.000	n	n/ 10.000	+ / -
Adjectives	64827	970.09	70786	1054.19	-84.09
Adverbs	38309	573.27	30955	461.00	+112.27
Nouns	133421	1996.56	173647	2586.05	-589.50
Verbs	60971	912.39	58491	871.08	+41.31

n= raw frequency of lexical categories

n/ 10.000= frequency of lexical categories per 10.000 words

(-/+) = difference between relevant value and value in NACE per 10.000 words; (+) denotes overuse; (-) denotes underuse

Table 44 indicates an overuse in SACE in comparison to NACE for the categories of adverbs and verbs. In other words, these lexical tokens were employed by SAEs more frequently than NAEs (Adverbs: +112.27; Verbs: +41.31); namely, adverbs were used by SAEs over 112 more times and verbs 41 more times per 10.000 words than NAEs. On the contrary, adjectives and nouns revealed to be underused by SAEs. For instance, adjectives and nouns were employed in SACE 84 and 589 fewer times per 10.000 words, respectively than in NACE. LL test was applied between the two corpora to see whether the overuse and underuse in SACE in comparison to NACE would be confirmed. Table 45 offers the test results.

Table 45

LL Ratio of Lexical Categories in SACE and NACE

Lexical Categories	SACE n	NACE n	LL Ratio (*p< 0.05)
Adjectives	64827	70786	-234.07
Adverbs	38309	30955	+818.01
Nouns	133421	173647	-5093.28
Verbs	60971	58491	+64.10

SACE is observed frequency in Corpus 1

NACE is observed frequency in Corpus 2

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

Table 45 suggests that the highest LL value was reported as underuse in SACE against NACE. It was found between the corpora in concern in the category of nouns with a LL value of -5093.28. The second highest value revealed to be overuse in SACE within the category of adverbs (p= +818.01). The third highest difference realized as underuse in the category of adjectives with a LL value of -234.07. Finally, the lowest difference was found between the two corpora was overuse in SACE against NACE in the category of verbs (p= +64.10). Despite not being so striking as those found between TACE and NACE in terms of the LL values calculated for epistemic categories and lexical categories, the differences calculated between SACE and NACE with respect to the values in concern are worth mentioning. There seem to be significant differences between the two sets of values counted for epistemic categories and lexical categories (See Table 19). The lowest value in the epistemic categories was found as underuse in the category of epistemic adjectives (-207.19), which was approved by the results of the analysis conducted for the category of adjectives with a similar LL value (-234.37). In a

similar vein, LL value calculated for the category of epistemic adverbs revealed to be overuse in SACE (+12.55) and this was confirmed with a much higher LL value (+818.01) obtained from the later analysis made for the category of adverbs. Epistemic verbs, on the other hand, were reported as underuse in SACE against NACE with a LL value of -24.50; however, the results of the analysis administered for the category of verbs indicated an overuse in SACE (+64.10 LL) against the native corpus. The most meaningful difference was found as underuse in the category of nouns with a LL value of -5093.28, which seems to contradict with the value previously measured between two corpora for the category of epistemic nouns (+27.10).

4.1.5. Individual Lexical Categories in Three Corpora

Tokens in the identified lexical categories (*adverbs, adjectives, nouns* and *verbs*) have been examined individually in the corpora they appeared. The following sections are intended to describe and outline the top 50 items in each category across three corpora.

4.1.5.1. Adjectives in Three Corpora

All three corpora were tagged via the CLAWS tagger in VMatrix (Rayson, 2013) to determine top fifty adjectives in each. Before discussing the items in concern, it might be useful to mention the proportion of epistemic adjectives in all adjectives found in each corpus, as given in Figure 17.

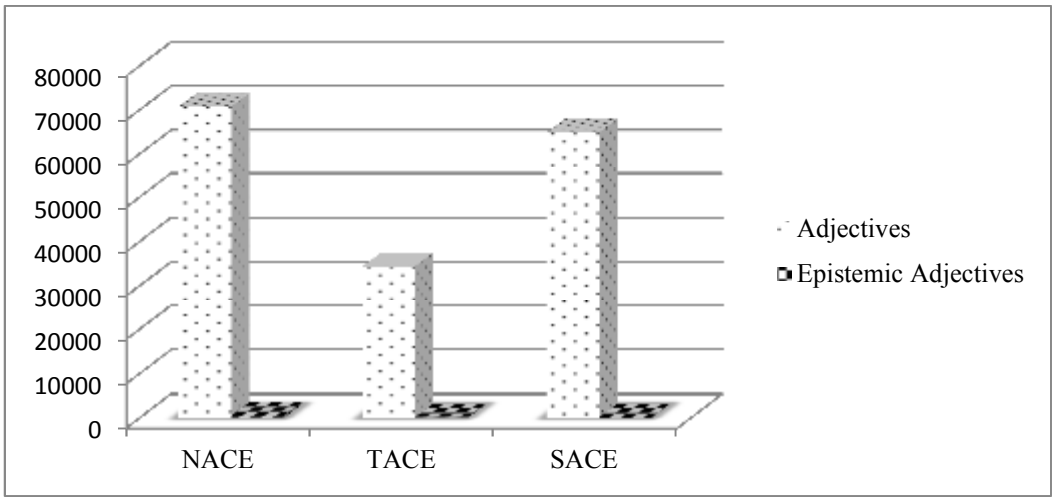


Figure 17. Epistemic adjectives and all adjectives in three corpora

As outlined in Figure 17, epistemic adjectives constitute a very small portion of all adjectives across three corpora (NACE: 0.8%; TACE: 0.7%; SACE: 0.3%). NACE and TACE seem to slightly differ in this sense whereas a significant deviance is observed in SACE when compared to other corpora. Top fifty adjectives found in NACE are displayed in Table 46.

Table 46

Most Frequently Used 50 Adjectives in NACE

Adjectives	n	T/t %	n per 10.000
Other	1261	0.19	18.78
Significant	1175	0.17	17.50
Low	1131	0.17	16.84
High	1070	0.16	15.94
Different	1031	0.15	15.35
Native	995	0.15	14.82
English	721	0.11	10.74
Passive	717	0.11	10.68
Main	641	0.10	9.55
Lexical	579	0.09	8.62
Spanish	537	0.08	8.00
Possible	511	0.08	7.61
Social	455	0.07	6.78
Verbal	423	0.06	6.30
Previous	418	0.06	6.23
Great	416	0.06	6.20
New	406	0.06	6.05
Likely	390	0.06	5.81
Individual	346	0.05	5.15
Present	343	0.05	5.11
Similar	341	0.05	5.08
Further	328	0.05	4.88
Specific	324	0.05	4.83
Large	321	0.05	4.78
Linguistic	319	0.05	4.75
Old	315	0.05	4.69
Visual	312	0.05	4.65
Small	311	0.05	4.63
Particular	302	0.04	4.50
Overall	300	0.04	4.47
Important	299	0.04	4.45
Able	285	0.04	4.24
Clear	275	0.04	4.10
Active	265	0.04	3.95
Final	261	0.04	3.89
Phonological	258	0.04	3.84
Full	253	0.04	3.77
Semantic	242	0.04	3.60
German	241	0.04	3.59
Moral	240	0.04	3.57
Following	233	0.03	3.47

(Table 46 Continued)

Frequent	230	0.03	3.43
Common	229	0.03	3.41
Only	222	0.03	3.31
Strong	220	0.03	3.28
Current	219	0.03	3.26
Consistent	207	0.03	3.08
Advanced	205	0.03	3.05
Relative	191	0.03	2.84
Perfective	188	0.03	2.80
Total	21502	3.22	320.22

n= raw frequency of adjectives in corpus

% T/t= Type/token ratio, percentage of the adjectives in overall tokens in corpus

As can be concluded from Table 46, the most frequented 50 adjectives in NACE occurred more than 320 times per 10.000 words in it and they compose over 30% of all adjectives found across the entire corpus. *Other* seems to be the mostly used adjective occurring 18.87 times in every 10.000 words in NACE. *Significant* and *low* that appeared 17.50 and 16.84 times per 10.000 words in the corpus, respectively are the second and third most frequented ones. The counterpart list of TACE can be found in Table 47.

Table 47

Most Frequently Used 50 Adjectives in TACE

Adjectives	n	T/t %	n per 10.000
High	938	0.14	13.89
Other	666	0.10	9.87
Significant	616	0.09	9.12
English	583	0.09	8.64
Turkish	567	0.08	8.40
Different	520	0.08	7.70
Following	385	0.06	5.70
Low	344	0.05	5.10
Effective	342	0.05	5.07
Experimental	339	0.05	5.02
Correct	299	0.04	4.43
Important	293	0.04	4.34
Aware	271	0.04	4.01
Similar	264	0.04	3.91
Positive	249	0.04	3.69
Good	241	0.04	3.57
Able	217	0.03	3.21
New	213	0.03	3.16
Present	210	0.03	3.11
Foreign	202	0.03	2.99
Main	195	0.03	2.89
Native	194	0.03	2.87
Further	185	0.03	2.74
Personal	184	0.03	2.73
Direct	183	0.03	2.71

(Table 47 Continued)

Possible	182	0.03	2.70
Incorrect	177	0.03	2.62
Verbal	174	0.03	2.58
Previous	170	0.03	2.52
Clear	169	0.03	2.50
Total	161	0.02	2.38
Difficult	159	0.02	2.36
Great	152	0.02	2.25
Lexical	149	0.02	2.21
Social	147	0.02	2.18
Specific	144	0.02	2.13
Young	143	0.02	2.12
Mean	140	0.02	2.07
Written	137	0.02	2.03
Only	131	0.02	1.94
Patriarchal – Post-Test	128	0.02	1.90
Current	123	0.02	1.82
Certain	121	0.02	1.79
Ideal - Regular	120	0.02	1.78
Various - Free	119	0.02	1.76
Critical – Tense - Intrinsic	117	0.02	1.73
Pragmatic	116	0.02	1.72
General	114	0.02	1.69
Particular	112	0.02	1.66
Desired – Developmental - Negative	110	0.02	1.63
Total	12215	1.85	180.94

n= raw frequency of adjectives in corpus

% T/t= Type/token ratio, percentage of the adjectives in overall tokens in corpus

As Table 47 suggests, the most widely used adjectives in TACE constitute 35.50% of all adjectives found in it. They occurred approximately 181 times in every 10.000 words in TACE while they were seen over 323 times per 10.000 words in NACE. Nonetheless, the corpora in concern do not significantly differ in this sense moving from the fact that 26 items seem to be identical (*high, other, significant, English, different, following, low, important, similar, able, new, main, native, further, possible, verbal, previous, clear, great, lexical, social, specific, only, current* and *particular*). The followings are the adjectives that appear among the top 50 adjectives in TACE and that do not exist in the counterpart list of NACE: *Turkish, effective, experimental, correct, aware, positive, good, foreign, personal, direct, incorrect, total, difficult, young, mean, written, patriarchal, post-test, certain, ideal, regular, various, free, critical, tense, intrinsic, pragmatic, general, desired, developmental* and *negative.* Table 48 contains the most frequented 50 adjectives in SACE.

Table 48

Most Frequently Used 50 Adjectives in SACE

Adjectives	n	T/t %	n per 10.000
Spanish	2727	0.41	40.81
English	1402	0.21	20.98
Other	1285	0.19	19.23
Different	1222	0.18	18.29
High	1005	0.15	15.04
Low	579	0.09	8.66
Overt	517	0.08	7.74
New	513	0.08	7.68
Lexical	509	0.08	7.62
Present	501	0.07	7.50
Similar	495	0.07	7.41
Significant	486	0.07	7.27
Positive	477	0.07	7.14
Irish	458	0.07	6.85
Social	449	0.07	6.72
Written	423	0.06	6.33
Common	420	0.06	6.29
Negative - Main	410	0.06	6.14
Following	384	0.06	5.75
Female	362	0.05	5.42
Previous	358	0.05	5.36
Important	355	0.05	5.31
Linguistic	348	0.05	5.21
False	341	0.05	5.10
Specific	327	0.05	4.89
Single	322	0.05	4.82
Clear	294	0.04	4.40
Basque	286	0.04	4.28
Possible	285	0.04	4.26
Simple	282	0.04	4.22
Academic	281	0.04	4.20
Cultural	277	0.04	4.15
Frequent	276	0.04	4.13
Semantic	263	0.04	3.94
Pronominal	260	0.04	3.89
Direct	251	0.04	3.76
General	250	0.04	3.74
Relevant - Iconical	241	0.04	3.61
Personal	239	0.04	3.58
Certain	229	0.03	3.43
Japanese	226	0.03	3.38
Past	220	0.03	3.29
Oral	214	0.03	3.20
Phonological	212	0.03	3.17
Only	211	0.03	3.16
Political	210	0.03	3.14
Recurrent	202	0.03	3.02
External	201	0.03	3.01
Further	197	0.03	2.95
Null-Visual	192	0.03	2.87
Total	22474	3.33	336.31

n= raw frequency of adjectives in corpus

% T/t= Type/token ratio, percentage of the adjectives in overall tokens in corpus

It is seen that the most frequently used adjectives in SACE accounts for 34.67% of all adjectives found in the corpus in concern. Thus, they appeared over 336 times in every 10.000 words in SACE. Half the adjectives appearing in the top 50 lists in SACE and NACE are identical. 26 adjectives in the list of SACE do not exist in that of NACE. In return for this, the top fifty list of NACE includes 24 adjectives that do not appear in the counterpart list of SACE. The following are the items in question: *able, active, advanced, common, consistent, current, final, German, great, individual, large, likely, moral, native, old, overall, particular, passive, perfective, relative, small, strong, verbal* and *visual.*

It is not surprising that *Spanish* and *English* revealed to be the mostly used adjectives in SACE when it is considered *Turkish* and *English* appeared among the most frequented adjectives in the counterpart list in TACE. However, the former couple were seen relatively more often in SACE than the latter in TACE. Namely, *Spanish* were used 40.81 times per 10.000 words in SACE while *Turkish* was employed 8.40 times in every 10.000 words in TACE. Likewise, it was observed that *English* occurred 20.98 times in SACE whereas it was preferred 8.64 times per 10.000 words in TACE. In NACE, *English* and *Spanish* appeared 10.74 and 8.00 times in every 10.000 words, respectively. *Other, different* and *high,* which were extensively used in TACE and NACE, seemed to be frequently preferred in SACE, as well. For example, *other* was found 19.23, 18.78 and 9.87 times in every 10.000 words in SACE, NACE and TACE, respectively, which means it was preferred by SAEs and NAEs much more frequently than TAEs. Similarly, *high* was extensively found across all three corpora (NACE: 15.94 times per 10.000 words; SACE: 15.04 times per 10.000 words; TACE: 13.89 times per 10.000 words). Finally, *different* was employed by SAEs relatively more often than NAEs and TAEs; namely, it occurred 18.29 times per 10.000 words in SACE whereas it was found 15.35 and 7.70 times in NACE and TACE, respectively.

The three corpora were analysed through a log likelihood test in order to reveal whether TACE and SACE significantly differ from NACE in the use of identical adjectives seen in the top fifty list of each. Initially, TACE and NACE were compared via the analysis in concern. The related results are shown in Table 49.

Table 49

LL Ratio of Adjectives in TACE and NACE (Top 50)

Adjectives	TACE n	NACE T/t %	LL Ratio (*p< 0.05)
Following	385	233	+36.97
Important	293	299	-0.10
High	938	1070	-9.40
Able	217	285	-9.60
Similar	264	341	-10.24
English	583	721	-15.37
Only	131	222	-24.21
Clear	169	275	-26.12
Current	123	219	-27.83
Present	210	343	-33.01
Further	185	328	-41.16
New	213	406	-62.22
Specific	144	324	-72.01
Particular	112	302	-79.85
Verbal	174	423	-108.42
Previous	170	418	-109.27
Great	152	416	-128.96
Possible	182	511	-164.42
Social	147	455	-166.94
Different	520	1031	-174.27
Significant	616	1175	-180.40
Other	666	1261	-189.93
Main	195	641	-253.12
Lexical	149	579	-273.60
Low	344	1131	-446.72

n=raw frequency of adjectives in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As indicated in Table 49, most of the identical adjectives in TACE and NACE were underused by TAEs against NAEs. *Following* seems unique to be overused by TAEs with a +36.97 LL value. That is, it was employed 368 times by TAEs whereas it was preferred 134 times by NAEs. The following are the examples extracted from a single work by a TAE.

Example 15.

[The first three parts deal with student teacher anxiety while the **following** four parts deal with foreign language student teacher anxiety.]

[The overall findings and findings according to the six factors of the scale are presented in the **following** subsections.]

[The **following** extracts from student teachers' experience about their stress when they were about to be observed would be enough to clarify the issue.]

Extracted from <TACE-AU-2010-MA>

All the other adjectives in this category were underused by TAEs. *Important* revealed to be the least underused adjective by TAEs with a -0.1 LL value, as illustrated in Example 16.

Example 16.
[That is why aphasiology has an important role in explaining the relationship between language and the brain.]

Extracted from <TACE-HU-2008-AAB>

[Insight has been provided into **important** aspects of TMI settings, such as how the affordances and constraints of technology can impact upon interaction.]

Extracted from <NACE-NU-2011-AB>

High is the other adjective TAEs and NAEs do not significantly differ in using with a -9.4 LL value. The following are the sentences obtained from TACE and NACE, respectively.

Example 17.
[The results of the (experimental) study indicated that the (experimental) group displayed a **high** degree of teaching metacognitive awareness according to the findings of the Wilcoxon Matched-Pairs Signed-Rank Test.

Extracted from <TACE-GU-2010-CB>

[As the focus of this study is the effect of PP on non-word repetition, the scores across lengths were collapsed yielding one overall score for **high** PP and one for low PP non-word repetition ability.]

Extracted from <NACE-NU-2009-CM>

It is seen that *able* was the third least underused adjective in its category with a -9.6 LL value. Namely, it occurred 217 and 285 times in TACE and NACE, respectively. Extracts from both corpora are given in the following.

Example 18.

[In the suggested syllabus, the researcher has tried to cover as much topic as possible but because of the time limitation she is not **able** to handle all of them in as much detail as she wishes.]

Extracted from <TACE-GU-2012-DBA>

[It was also suggested, with reference to previous research, that the three parties of the Scottish centre, over the last 40 years, have been **able** to successfully link left-wing ideology with Scottish national identity.]

Extracted from <NACE-GU-2006-DJS>

The two corpora significantly differ in the use of 25 identical adjectives. *Low* is the most underused adjective by TAEs with a -446.72 LL value. It appeared 1131 times in NACE while it was found 344 times in TACE. Sentences in Example 19 were extracted from NACE.

Example 19.

[High lexical density was associated with high measures of technicality and grammatical metaphor, for example in the biology textbook, and **low** lexical density with low measures of grammatical metaphor and technicality, for example in the physics CD-ROM.]

Extracted from <NACE-US-2006-JJ>

[The EPE/ higher proficiency students have a slightly larger range in the number of re-mentions, spanning from a **low** of three (Students 7 and 8) to a high of 20 (Student 6).]

Extracted from <NACE-UI-2010-JEG>

Similarly, *lexical* was found 579 times in NACE while it was used 149 times in TACE (-273.6 LL value). It is illustrated in the sentences taken from NACE in Example 20.

Example 20.
[Following the eye-tracking session participants were tested on their knowledge of the **lexical** items used in the experiment.]

[Whereas English is typically limited to **lexical** NPs or personal pronouns, German also has anaphoric demonstrative pronouns which are used for less salient entities.]

[It is possible, for example that the extra **lexical** material in German passive sentences allows L2 learners more time to integrate different sources of information, resulting in a target like preference, or, indeed, lack or preference.]

Extracted from <NACE-UE-2009-FW>

On the contrary to the case in TACE, most of the identical adjectives in SACE and NACE seems to be overused by SAEs against NAEs. Namely, it is seen that the SAEs overused 15 out of 26 items in comparison to NAEs. Table 50 shows all identical adjectives in SACE and NACE.

Table 50

LL Ratio of Adjectives in SACE and NACE (Top 50)

Adjectives	SACE n	NACE n	LL Ratio (*p< 0.05)
Spanish	2727	537	+1616.78
English	1402	721	+225.64
Common	420	229	+57.97
Following	384	233	+38.06
Present	501	343	+30.52
Similar	495	341	+29.28
Different	1222	1031	+17.14
New	513	406	+13.01
Important	355	299	+5.07
Frequent	276	230	+4.41
Linguistic	348	319	+1.40
Semantic	263	242	+0.98
Clear	294	275	+0.73
Other	1285	1261	+0.36
Specific	327	324	+0.03
Social	449	455	-0.02
Only	211	222	-0.23
High	1005	1070	-1.74
Previous	358	418	-3.67
Lexical	509	579	-4.18
Phonological	212	258	-4.29
Further	197	328	-32.41
Main	410	641	-50.08
Possible	285	511	-63.98
Low	579	1131	-178.78
Significant	486	1175	-291.32

n=raw frequency of adjectives in corpus

\+ indicates overuse in SACE relative to NACE

\- indicates underuse in SACE relative to NACE

Not surprisingly, *Spanish* was extensively used by SAEs. It was found 2727 times in SACE whereas it occurred 537 times in NACE, which corresponds to an enormous LL value. Example 21 is intended to illustrate a couple of sample statements including it.

Example 21.

[From considering their experiences, it is clear that the **Spanish** government needs to radically rethink its plans for the English PAU if it does not want to repeat the mistakes and failures of the past.]

[Learning a foreign language, and particularly English, is paramount in modern **Spanish** society.]

[The second area of study of **Spanish** academics has been proposals for improvement of the test.]

Extracted from <SACE-UPV-2012-DLN>

English, on the other hand, holds the second rank in the list of identical adjectives where the two groups of academic writers significantly differ. It was found in SACE approximately twice as many in NACE with a +225 LL value. The following are the statements taken from SACE.

Example 22.
[Lastly, female participants spent more time on planning and learning **English** vocabulary.]

[What is more, 12-14 year old students admitted learning only vocabulary taught by their **English** teacher.]

[The last factor considered in this study was the amount of **English** instruction received by our participants.]

Extracted from <SACE-USC-2009-RFC>

The most underused adjective in SACE is *significant* with a -291.32 LL value. It seems to be employed 1175 times by NAEs and 486 times by SAEs. It might be beneficial to recall here that it was also underused by TAEs with a -180.4 LL value. The following statements that contain the item in question were taken from NACE.

Example 23.
[Of course, individual stories can be observed to increase or decrease by much more **significant** amounts.]

[Overall, the LD differences between the rejected and the accepted stories are not **significant** enough to warrant the claim of them being categorical.]

[As shown in figure 4.5, a **significant** number of stories exhibit a lower LD after revision.]

Extracted from <NACE-UE-2008-OS>

As in the case of TACE, *low* and *possible* are among the mostly underused adjectives in SACE. For instance, *low* is the second and *possible* is the third in this category with -178.78 and -63.98 LL values, respectively. A few statements extracted from both corpora are given in Examples 24 and 25.

Example 24.

[Across the different genres, however, high and **low** LD percentages are evenly distributed.]

Extracted from <NACE-UE-2008-OS>

[The structured content of the songs also provided benefits with regard to grammar, as in the exam, when the students had to fill out grammar exercises they often sang in a **low** voice, as they knew the grammar was contained in the lyrics.]

Extracted from <SACE-UZ-2012-CMTF>

Example 25.

[At the text level, it is not **possible** to spell out more specific characteristics, which would hold generally for all main narrative bodies.]

Extracted from <NACE-UE-2008-OS>

[In his seminal work, Alberti developed all possible mathematical formulas to create the illusion of depth, and therefore, three-dimensionality, in a flat, two-dimensional surface.]

Extracted from <SACE-UZ-2012-ARJ>

It is noteworthy that SACE do not significantly differ from NACE in the use of particular adjectives that are among the mostly underused ones in TACE such as *clear, high*, *only, other, social* and *specific*. Taking into consideration the facts that 96% of the identical adjectives in TACE and NACE were underused by TAEs and that 70% of

those in SACE and NACE were overused by SAEs, it can be concluded that TACE and SACE considerably differ regarding the use of adjectives identical in NACE. Finally, it might be practical to cover the adjectives identical in three corpora. Figure 18 shows the distribution of the items in concern across the corpora.

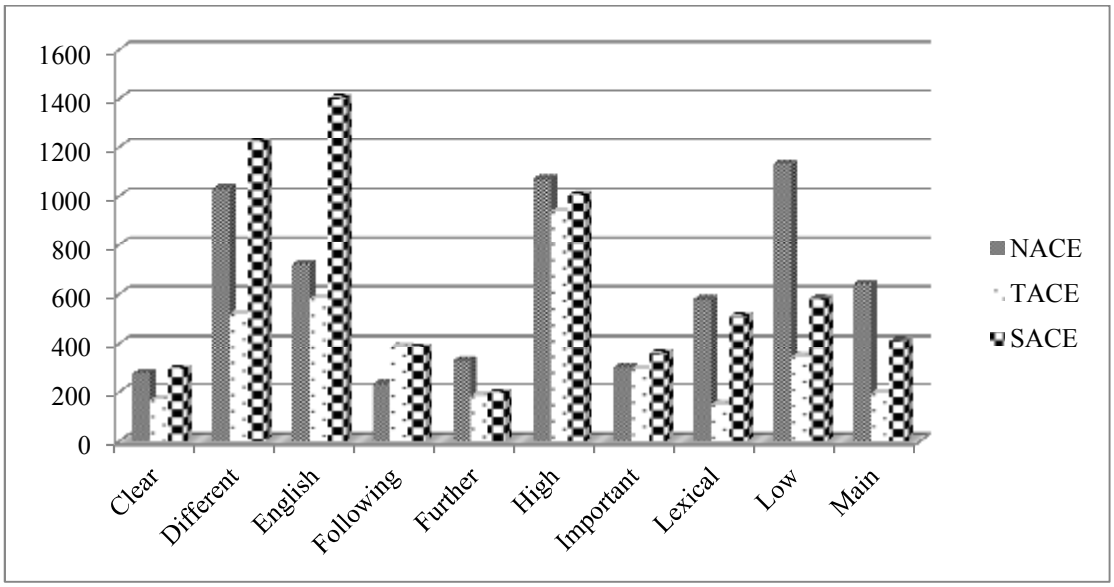


Figure 18. Identical adjectives in three corpora (1-10)

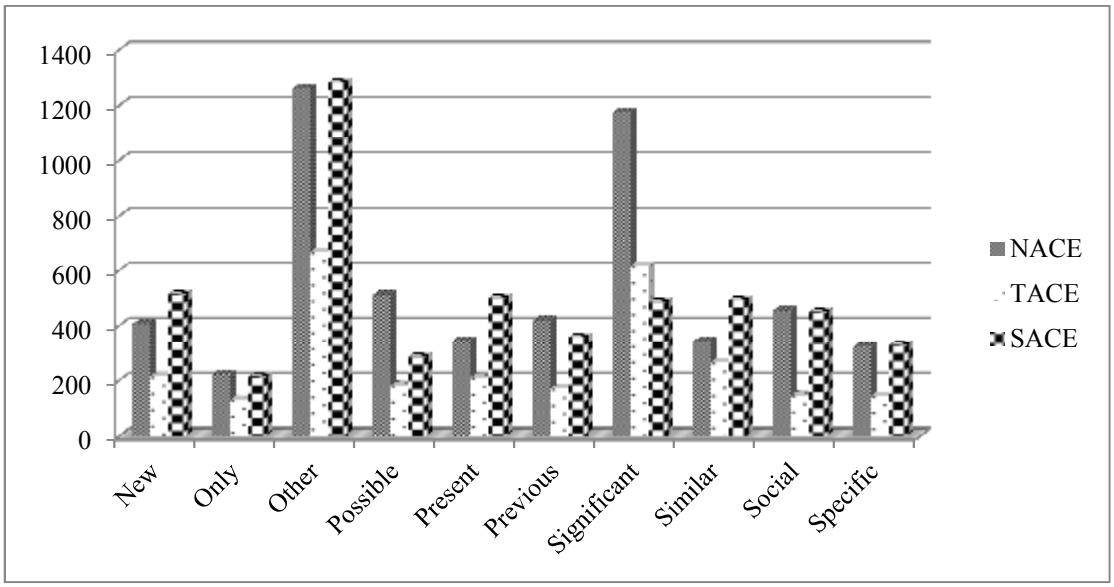


Figure 18. Identical adjectives in three corpora (11-20)

As seen in Figure 18, approximately half of the adjectives are common in the top 50 adjective lists of three corpora; however, this does not necessarily mean that they display a balanced distribution across the corpora in concern. *High* is the only item the

three corpora seem alike; namely, it appeared 938, 1005 and 1070 times in TACE, SACE and NACE, respectively. TACE and NACE are relatively similar in the distribution of two items (i.e., *English* and *important*) when compared to SACE. In response to this, ten adjectives appeared at parallel frequencies in SACE and NACE in compared to TACE (i.e., *clear, different, lexical, main, new, only, other, previous, social* and *specific*). TACE and SACE seem to be alike in the use of 4 items against NACE, which are *following, further, low* and *significant*. The following are the adjectives they display different tendencies with regard to their use: *main, new* and *similar*. For instance, *main* was preferred 195 times in TACE while it was used 410 and 641 times in SACE and NACE, respectively. It is illustrated in Example 26.

Example 26.

[The **main** objective of these series is to prepare the learners for English Comprehension Examination (ECL), which is based on listening and reading comprehension.]

Extracted from <TACE-GU-2010-ES>

[The reason may be that the verb *can* constitutes the centre of one of the **main** grammatical points in this course: the ability to do something.]

Extracted from <SACE-UM-2011-GAM>

[This chapter summarises the **main** claims of the dissertation and explores their theoretical repercussions, also suggesting solutions for some problems raised.]

Extracted from <NACE-UU-2008-AA>

In a similar vein, *new* was underused by NAEs and overused by SAEs occurring 213, 406 and 513 times in TACE, NACE and SACE, respectively. Example 27 provides the statements taken from each corpus.

Example 27.

[As a result of the related research revision, some common criteria have been chosen and used in this study in order to identify the grammatical category of a **new** word.]

Extracted from <TACE-CU-2005-FA>

[This thesis also suggests that intonation may be a **new** application for a combined measure of phonetic distance for incorporation with phylogenetic network methods like NeighborNet.]

Extracted from <NACE-UE-2010-JS>

[This **new** counter-evidence seemed to be reinforced when the production rate of Head Act internal upgrading devices was calculated.]

Extracted from <SACE-UG-2008-NILS>

Lastly, the three corpora are ironically different in the use of *similar*. That is, it was underused in TACE, appearing 264 times while it was overused in SACE, occurring 495 against NACE where it was found 341 times. The following extracts were taken from the corpora in question.

Example 28.
[It is **similar** to limited-production tasks in that students are required to construct new knowledge and not just select a response.]

Extracted from <TACE-CU-2010-KB>

[However, it must be acknowledged that a **similar** effect is achieved in the Boer War text through the inclusion of extracts from the editorials of other newspapers.]

Extracted from <NACE-MU-2009-CES>

[If they are supplied in **similar** rates, there will be reason to believe that the phonological challenge is a general constraint against consonant clusters (i.e., at the syllable level).]

Extracted from <SACE-UI-2011-GSCD>

The procedure administered to three corpora in order to investigate whether they significantly differ in the use of adjectives was repeated for the categories *adverbs, nouns* and *verbs*.

4.1.5.2. Adverbs in Three Corpora

Adverbs were the least frequented lexical category in three corpora. Namely, they were used about 31000 times in NACE, approximately 24000 times in TACE and over 38000 times in SACE. As for their distribution across three corpora, it is better to provide a figure showing the proportion of epistemic adverbs in all adverbs found across each corpus, as in the following.

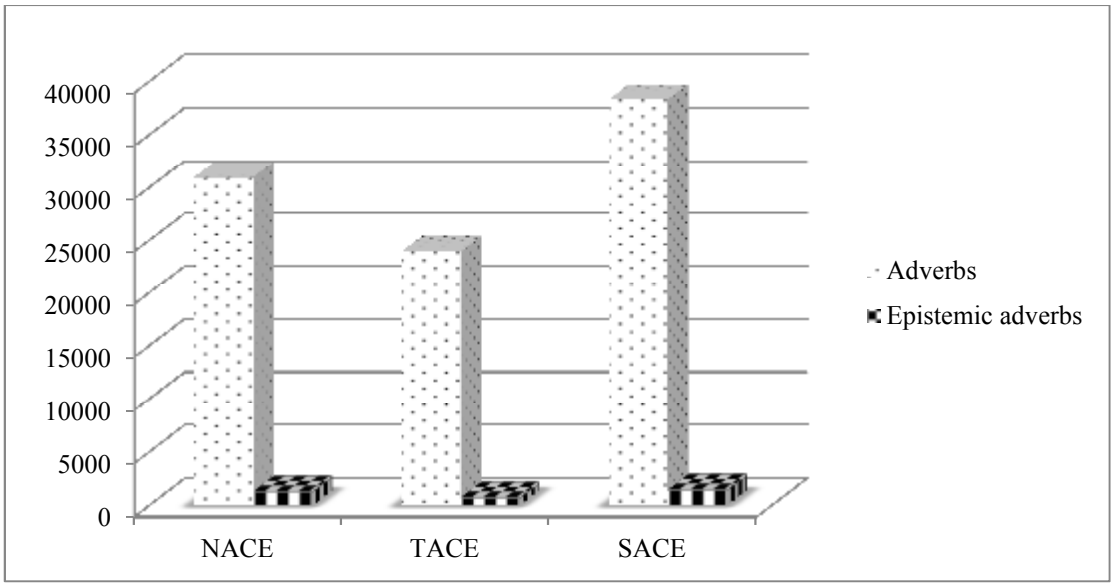


Figure 19. Epistemic adverbs and all adverbs in three corpora

As Figure 19 suggests, the highest proportion of epistemic adverbs to all adverbs was observed in NACE (4.2%) while the lowest was measured in TACE, where epistemic adverbs constitute 3.1% of all adverbs. These items comprise approximately 4% of all adverbs in SACE.

Top fifty adverbs in NACE, TACE and SACE have been obtained and those found in NACE are presented in Table 51.

Table 51

Most Frequently Used 50 Adverbs in NACE

Adverbs	n	T/t %	n per 10.000
More	1975	0.29	29.41
Also	1760	0.26	26.21
However	1246	0.19	18.56
Only	935	0.14	13.92
Both	560	0.08	8.34
Then	518	0.08	7.71
Therefore	508	0.08	7.57
Most	494	0.07	7.36
Very	492	0.07	7.33
So	469	0.07	6.98
Significantly	457	0.07	6.81
Here	430	0.06	6.40
For example	401	0.06	5.97
Less	385	0.06	5.73
Thus	384	0.06	5.72
Often	371	0.06	5.53
Either	369	0.05	5.50
Again	332	0.05	4.94
How	328	0.05	4.88
Still	301	0.04	4.48
Even	294	0.04	4.38
Above	285	0.04	4.24
Further	263	0.04	3.92
As	244	0.04	3.63
Just	242	0.04	3.60
Below	233	0.03	3.47
i.e.	221	0.03	3.29
Perhaps	216	0.03	3.22
Particularly	215	0.03	3.20
Clearly	203	0.03	3.02
Frequently	188	0.03	2.80
E.G.	182	0.03	2.71
Much	181	0.03	2.70
At Least	179	0.03	2.67
Finally-Previously	169	0.03	2.52
Indeed	162	0.02	2.41
Hence	156	0.02	2.32
Generally	152	0.02	2.26
Always	151	0.02	2.25
That	147	0.02	2.19
Now	146	0.02	2.17
Furthermore - Simply	144	0.02	2.14
In addition-Statistically-Yet	134	0.02	2.00
Already - Relatively	133	0.02	1.98
Is	132	0.02	1.97
Too	128	0.02	1.91
Highly	126	0.02	1.88
Consistently - Overall	125	0.02	1.86
In	123	0.02	1.83
Almost	121	0.02	1.80
Total	18110	2.69	269.70

n= raw frequency of adverbs in corpus

% T/t= Type/token ratio, percentage of the adverbs in overall tokens in corpus

As illustrated in Table 51, the most frequented 50 adverbs in NACE comprise 58.5 % of total adverbs found in the corpus at stake. They were seen approximately 269 times in every 10.000 words in NACE. *More*, which appeared 29.41 times in every 10.000 words, revealed to be the most preferred adverb by NAEs. *Also* is the second and *however* is the third mostly used adverbs in NACE occurring 26.21 and 18.56 times in every 10.000 words in the corpus, respectively. Table 52 includes the most used 50 items in TACE.

Table 52

Most Frequently Used 50 Adverbs in TACE

Adverbs	n	T/t %	n per 10.000
Also	1649	0.24	24.43
More	1234	0.18	18.28
However	1066	0.16	15.79
Most	733	0.11	10.86
Only	646	0.10	9.57
Below	519	0.08	7.69
Thus	514	0.08	7.61
Very	492	0.07	7.29
How	463	0.07	6.86
Both	413	0.06	6.12
Therefore	406	0.06	6.01
Above	383	0.06	5.67
Statistically	368	0.05	5.45
Significantly	309	0.05	4.58
Then	268	0.04	3.97
Moreover	244	0.04	3.61
So	238	0.04	3.53
Here	216	0.03	3.20
Yet	211	0.03	3.13
Quite	207	0.03	3.07
Furthermore	195	0.03	2.89
Even	193	0.03	2.86
As well	191	0.03	2.83
That	178	0.03	2.64
Hence	176	0.03	2.61
Correctly - Especially	175	0.03	2.59
As	174	0.03	2.58
Clearly	172	0.03	2.55
Mostly - Still	171	0.03	2.53
Is	168	0.02	2.49
Again	163	0.02	2.41
Either - Less	160	0.02	2.37
Easily	151	0.02	2.24
Conventionally	150	0.02	2.22

(Table 52 Continued)

Almost	147	0.02	2.18
Just	146	0.02	2.16
Generally	144	0.02	2.13
In addition	143	0.02	2.12
When	142	0.02	2.10
Better	138	0.02	2.04
Where	136	0.02	2.01
Similarly	133	0.02	1.97
Frequently	132	0.02	1.96
Why	131	0.02	1.94
Besides	130	0.02	1.93
Often	123	0.02	1.82
For example	122	0.02	1.81
In general	116	0.02	1.72
Finally	112	0.02	1.66
Respectively	111	0.02	1.64
Total	15102	2.27	223.71

n= raw frequency of adverbs in corpus

% T/t= Type/token ratio, percentage of the adverbs in overall tokens in corpus

As illustrated in Table 52, the most frequented 50 adverbs in TACE compose 63.15 % of total adverbs found in the corpus at stake. They were seen approximately 224 times in every 10.000 words in TACE whereas they occurred 270 times per 10.000 words in NACE. In other words, it is seen that 37 adverbs in the lists are identical in TACE and NACE. Followings are the ones that do not appear in the top 50 adverb list of NACE: *moreover, quite, as well, correctly, especially, mostly, easily, conventionally, when, better, where, similarly, why, besides, in general* and *respectively*. Lastly, the most employed adverbs by SAEs are given in Table 53.

Table 53

Most Frequently Used 50 Adverbs in SACE

Adverbs	n	T/t %	n per 10.000
Also	3099	0.46	46.37
More	1608	0.24	24.06
However	1247	0.19	18.66
Thus	1116	0.17	16.70
Only	1114	0.17	16.67
Most	923	0.14	13.81
How	558	0.08	8.35
Both	528	0.08	7.90
Very	525	0.08	7.86
So	489	0.07	7.32
Finally	480	0.07	7.18
Often	433	0.06	6.48
Even - Therefore	432	0.06	6.46
Here	403	0.06	6.03
Frequently	396	0.06	5.93
Below	346	0.05	5.18
Then	324	0.05	4.85
Especially	315	0.05	4.71
Either	314	0.05	4.70
Above	310	0.05	4.64
Still	301	0.05	4.50
Already - Less	292	0.04	4.37
Mainly	282	0.04	4.22
Clearly	273	0.04	4.09
As	264	0.04	3.95
Quite	260	0.04	3.89
Usually	253	0.04	3.79
Nevertheless	245	0.04	3.67
Significantly	242	0.04	3.62
Almost	236	0.04	3.53
Indeed	234	0.04	3.50
Much	232	0.03	3.47
When	224	0.03	3.35
Rather	222	0.03	3.32
i.e.	221	0.03	3.31
Just	220	0.03	3.29
Moreover	211	0.03	3.16
Actually - Probably	209	0.03	3.13
In general - That	207	0.03	3.10
Always	199	0.03	2.98
Respectively	197	0.03	2.95
Sometimes	196	0.03	2.93
Where	195	0.03	2.92
Generally	194	0.03	2.90
E.G. - Mostly	184	0.03	2.75
Again - Slightly	181	0.03	2.71
As well - Commonly	180	0.03	2.69
Normally	177	0.03	2.65
Highly - Particularly	176	0.03	2.63
Together	174	0.03	2.60
Total	21641	3.26	323.84

n= raw frequency of adverbs in corpus

% T/t= Type/token ratio, percentage of the adverbs in overall tokens in corpus

The most used 50 adverbs in SACE constitute 56.5 % of total adverbs found in it. They were found nearly 324 times in every 10.000 words in SACE. It is observed that 38 adverbs in the lists of top 50 in SACE and NACE are identical and 18 found in the list of SACE do not appear in that of NACE. Following are the adverbs in concern: *actually, as well, commonly, in general, mainly, moreover, mostly, quite, nevertheless, probably, rather, respectively, slightly, sometimes, together, usually, when* and *where*. *Also* occurred 46.37 times in every 10.000 words in SACE. *More* revealed to be the second most frequently used adverb in SACE. The third mostly used adverb in SACE is *however* appearing 18.66 times in every 10.000 words in the corpus.

A log likelihood test was administered to all corpora to investigate whether they significantly differ with respect to the identical adverbs in the top 50 lists for each. The first analysis was administered between TACE and NACE and the results are displayed in Table 54.

Table 54

LL Ratio of Identical Adverbs in TACE and NACE (Top 50)

Adverbs	TACE n	NACE n	LL Ratio (*p< 0.05)
Statistically	368	134	+112.17
Below	519	233	+110.04
Most	733	494	+45.58
How	463	328	+22.44
Thus	514	384	+18.20
Yet	211	134	+16.92
Above	383	285	+13.91
Furthermore	195	144	+7.43
Is	168	132	+4.14
That	178	147	+2.80
Almost	147	121	+2.39
Hence	176	156	+1.10
In addition	143	134	+0.25
Very	492	492	-0.00
Generally	144	152	-0.26
Clearly	172	203	-2.73
Also	1649	1760	-4.23
Frequently	132	188	-10.15
Finally	112	169	-11.95
Therefore	406	508	-11.96
As	174	244	-12.15
However	1066	1246	-15.01
Even	193	294	-21.64
Both	413	560	-23.09

(Table 54 Continued)

Just	146	242	-24.52
Significantly	309	457	-29.57
Still	171	301	-36.97
Only	646	935	-54.68
Again	163	332	-59.78
Here	216	430	-73.4
So	238	469	-78.12
Then	268	518	-82.26
Either	160	369	-85.99
Less	160	385	-96.93
Often	123	371	-131.67
For example	122	401	-158.34
More	1234	1975	-176.64

n=raw frequency of identical adverb in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As can be seen in Table 54, 13 out of 37 identical adverbs in TACE and NACE are the ones overused by TAEs. *Statistically* and *below* have the highest LL value in the list (*Statistically*: +112.17; *Below*: +110.04). More specifically, the former was used 368 times by TAEs while it was preferred by 134 times by NAEs and the latter was employed 519 times by TAEs whereas it appeared 233 times in the works of NAEs. The following examples were extracted from a single work in TACE, across which it appeared 10 times:

Example 29.

[It is clearly seen in Table 5.20 that the difference between participants' pre-test-post-test beliefs about the assertion that it is easier to speak than understand a foreign language is **statistically** significant.]

[No **statistically** significant change was observed in the participants' beliefs about items, which posit that nothing should be said in English until it can be said correctly and that the meaning of an unknown word can be guessed.]

[Concerning self-efficacy perceptions about reading, Wilcoxon test results point to **statistically** significant change for all items but one.]

Extracted from <TACE-CU-2010-SB>

In a similar vein, the following examples are provided to exhibit another overused token *below* which was employed 12 times in a single work by a TAE:

Example 30.
[The paired samples t-test table **below** displays the mean as M = 5.8095.]

[That is, stating as a fifth finding, there is not a significant difference between participants' anxiety levels and their attitudes towards L1 as seen **below** in correlation Table 33.]

[In-class observations supported this assumption as seen in the excerptions **below**.]
Extracted from <TACE-HU-2009-AK>

Most, how, thus and *yet* are the other adverbs preferred by TAEs significantly more often than NAEs. *Most* appeared 733 times in TACE whereas it was found 494 times in NACE with a +45.58 LL value. Below are the sample statements taken from a work in TACE.

Example 31.
[These anxiety creator views were **most** effective during their preparation to university entrance exam the year they started to university.]

[Additionally, these students **most** likely do not consider the tests in their courses at university as threatening as university entrance exam because they are not life changing exams.]

[For example, if a category includes three tasks, the highest score might be 15 at **most**.]
Extracted from <TACE-CU-2010-KB>

Similarly, how was employed 463 times by TAEs while it was preferred 368 times by NAEs, which corresponds to +22.44 LL value. It can be shown in the following examples extracted from TACE and NACE.

Example 32.

[The participants commented on **how** they would react to such students with various handling strategies.]

[At the end of the study, the participants seemed to have more knowledge of **how** to control disruptions before they occur and to employ various and more strategies to interfere with misbehaviours.]

[The students usually evaluate their teacher on **how** persistent and motivated s/he is about behaviour management.]

Extracted from <TACE-CU-2007-AK>

[Questions were asked to probe for what purpose and **how** frequently bilingual staff felt they used their mother tongue skills.]

[Although there are many published assessments to measure speech in English speaking children their authors rarely give any insight into **how** the word list was developed.]

Extracted from <NACE-NU-2006-CS>

Thus was also observed to have been employed by TAEs relatively more often when compared to NAEs. Namely, it was found 514 times in TACE while it occurred 384 times in NACE. Following extracts were drawn from the corpora in concern.

Example 33.

[**Thus**, besides the interpretation and exemplification of all errors listed above, a separate analysis of the gradual decrease of case marker and word order errors during the learning process within the period of 10 months, starting from 25.04.2005 to 28.02.2006, have also been presented in this chapter.]

[The reason can be interpreted to be due to the fact that the German alphabet does not include the letter "î" and **thus** is very difficult to pronounce for German native speakers.]

[The noun “zaman” belongs to the verb “geçirmek” (spend time) and **thus** has to be in preceding position of the verb.]

Extracted from <TACE-CU-2007-EB>

[All of them can **thus** be seen as double offers in which different opening styles are combined.]

[Although GPs are independent contractors, their funding comes through the state rather than directly from patients and the relationship between payment and receipt of services is **thus** obscured.]

Extracted from <NACE-EU-2006-HH>

As for *yet*, it was found in NACE less frequently than TACE; namely, it appeared 134 times in the former and 211 times in the latter, which indicates an overuse in TACE with a +16. 92 LL value. It is illustrated in the extracts obtained from each corpus, as in Example 34.

Example 34.

[**Yet**, they mainly used simple vocabulary strategies that are for short-term memory such as memorisation or repetition.]

[Apart from talking about pros and cons of a topic, or causes and consequences of a situation, which are relatively better **yet** not at a desired level, most students feel that they cannot do these and set them as learning objectives.

[Language learning requires a certain degree of risk taking, **yet** the rewards are great and worth it.]

Extracted from <TACE-GU-2008-GE>

[As they approached the last week of the trimester, they were faced once again with chasing down students who had **yet** to make up late work due to sickness.]

[Students in the row closest to me have not said one word in Spanish **yet** and we are 40 minutes into this class period.]

Extracted from <NACE-UI-2012-EJE>

In accordance with LL ratio calculated for the list of top 50 adverbs in TACE, a total number of 14 adverbs were underused by TAEs. *More* is the most underused adverb in TACE with a -176.64 LL value. The following are the expressions found in the two corpora.

Example 35.

[As will be discussed at length in chapters 7 and 8, the phonological cycle crucially predicts that phonological characteristics should only be inherited from immediately embedded words: a stress preservation relationship can only exist where the embedding and embedded words are separated by no **more** than one suffix.]

[There is a positive relationship between embedding frequency and preservation failure: the greater the embedding frequency, the **more** likely relative prominence preservation is to fail.]

[A brief chapter summary is presented in 10.1, before the findings of the thesis and their implications are discussed in **more** detail in 10.2.]

Extracted from <NACE-UE-2007-CS>

[In comparison to her first grid data obtained at the beginning of the study, at the end of the study she has incorporated three **more** constructs into her existing construct cluster.]

[At the end of the study, Teacher 4 seems to have now **more** precise opinions regarding the features of the ineffective teachers, as there is a tightening in the match levels of these.]

[The **more** points a teacher accumulates, the more advantaged she will be in some issues such as nomination to a better work place or to join some seminars.]

Extracted from <TACE-CU-2007-HS>

There is a significant difference between TACE and NACE in the use of *for example*. It appeared 401 times in NACE while it was seen 123 times in TACE. Statements including it are given in Example 36.

Example 36.

[Clearly the present position of genitive in (42) could not be correct, since Hungarian, **for example**, has dative, locative, ablative/instrumental and many others, but no genitive.]

[**For example** the morphological paradigm in (5) relates to several different sets of minimal pairs of changes in syntax.]

[Genitive, **for example**, is normally visibly present (at least in the form of 's or of) where it is analysed as being syntactically present in terms of Case features.]

Extracted from <NACE-UU-2008-AA>

[Sometimes they are asked to write assuming the role of some other person; **for example**, in Unit 3 students are asked to write a paragraph by looking at a picture of a girl at about the same age with them.]

[**For example**, the students could be asked to imagine themselves as people working for the Ministry of Education or so, and then asked to write about their plans to improve conditions of schooling as a report.]

Extracted from <TACE-IU-2011-GK>

Another adverb significantly underused by TAEs is *often*, which was used in NACE over twice as frequently in TACE with a -131.67 LL value. It can be seen in the extracts taken from the corpora in question.

Example 37.

[Most interestingly, while every player in the High-frequency Cooperative conditions was paired twice as **often** with one other player than with either of the others, this frequency bias did not lead these pairs to diverge significantly from the other players.]

[In three out of the five games, the team that most **often** distinguished correctly between teammates and opponents also won the game.]

Extracted from <NACE-UE-2010-GR>

[For this study, as the researcher was the instructor, we had a chance to meet very **often**.]

[The second question was about self-confidence, which was a term we **often** came up with during the analysis of reflection sheets, cover letters and interviews.]

Extracted from <TACE-CU-2009-SDU>

Subsequently, an LL test was applied between SACE and NACE to see whether they significantly differ regarding adverbs listed as top 50 for each. The results are presented in Table 55.

Table 55

LL Ratio of Identical Adverbs in SACE and NACE (Top 50)

Adverbs	SACE n	NACE n	LL Ratio (*p< 0.05)
Also	3099	1760	+380.27
Thus	1116	384	+376.47
That	207	147	+287.96
Finally	480	169	+156.83
Most	923	494	+134.01
Frequently	396	188	+76.74
Already	292	133	+61.72
How	558	328	+61.51
Almost	236	121	+38.27
Even	432	294	+27.06
Below	346	233	+22.74
Only	1114	935	+16.53
Indeed	234	162	+13.51
Clearly	273	203	+10.67
Highly	176	126	+8.56
Always	199	151	+6.84
Much	232	181	+6.56
Generally	194	152	+5.31
Often	433	371	+5.09
Very	525	492	+1.24
Above	310	285	+1.17
As	264	244	+0.89
So	489	569	+0.52
E.G.	184	182	+0.02
However	1247	1246	+0.02
Still	301	301	+0.00
I.E.	221	221	+0.00

(Table 55 Continued)

Here	403	430	-0.75
Both	528	560	-0.79
Just	220	242	-0.94
Particularly	176	215	-3.71
Either	314	369	-4.17
Therefore	432	508	-5.79
Less	292	385	-12.37
More	1608	1975	-35.91
Then	324	518	-44.18
Again	181	332	-44.39
Significantly	242	457	-66.19

n=raw frequency of identical adverbs in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

As outlined in Table 66, the majority of the identical adverbs in SACE and NACE are overused by SAEs. *Also, thus* and *that* are the three most overused adverbs among them. *Also* was used approximately 3100 times in SACE while it was found much less frequently in NACE. There is +380.27 LL value between two corpora as an overuse in SACE against NACE. Below are a couple of extracts drawn from each corpus.

Example 38.

[The student **also** sent a message pointing out what he considered to be a problem with the feedback message.]

[When researchers decide their research agendas, they have to make choices not only about their areas of contribution, but **also** the issues that are not central to their approaches.]

[However, we **also** believe that it is time for ICALL research to pay more attention to the gap between the NLP technology that has been developed in the last three decades and the use of this technology in real life language teaching.]

Extracted from <SACE-OSU-2007-LAMA>

[However, it is impossible to confirm that this would **also** be the case for Weskott et al's data without full reanalysis of their data.]

[There was **also** no significant preference for either NP in the Active + pronoun condition, unlike in Experiment 5.]

Extracted from <NACE-UE-2009-FW>

Thus reveals to be the second mostly overused adverbs by SAEs against NAEs with a +376.47 LL value. The following are the examples drawn from a work in SACE, where it appeared 1116 times and one in NACE, across which it was used 384 times:

Example 39.
[This strict code, as well as its European counterparts, was imposed on almost any kind of comic-book production, **thus** limiting them to Manichean representations that fit the patriarchal ideology of the dominating culture.]

[**Thus**, the simple question that these images were just drawings meant to offer a comment on the world today was never raised.]

[According to this theory, "high" culture emanates from the "central" system and is **thus** related to the official strata.]

Extracted from <SACE-UZ-2012-ARJ>

[Corsican, **thus**, is understood by many who have no other formal abilities in the language, which explains the spike in percentages for this skill.]

[Question 32 **thus** asked if more awareness of other minority groups had come about as a result of the EU.]

Extracted from <NACE-UE-2007-MK>

The third frequently overused adverb is *that* which appeared 207 and 147 times in SACE and NACE, respectively. A few statements taken from the corpora at stake are provided below.

Example 40.
[The main conclusion drawn from this chapter is **that** the different manifestations of stress assignment in Catalan, Portuguese and Spanish consider weight differently.]

Catalan and Portuguese display a QS non-verbal regular stress system and a QI prosodic morphology.]

[The analysis of the data concludes **that** quantity sensitive (QS) and quantity insensitive (QI) patterns can coexist within the same language: Valencian Catalan and Portuguese display a QS non-verbal regular stress system and a QI prosodic morphology.]

[The analysis on truncation relied on the view **that** disyllabicity was the most productive shape in all truncation patterns.]

Extracted from <SACE-UT-2006-AGS>

[The non-detectors were expected to demonstrate implicit knowledge, as their correction test performance suggested **that** they had no relevant explicit knowledge with which the former could be masked.]

[As the group's performance mirrored the statistical distribution of the input, it is possible **that** they relied on implicit knowledge.]

Extracted from <NACE-UE-2009-ALC>

Finally and *most* are the other adverbs overused by SAEs in comparison to NAEs. Namely, *finally* was found 480 times in SACE while it was observed 169 times in NACE (p= +156.83). Examples taken from the corpora are displayed below.

Example 41.

[**Finally**, in section 1.4, I provided a summary of cleft constructions and I focused on the syntactic differences between pseudo-clefts and the FS.]

[This **finally** led me to conclude that the FS should not be examined as a pseudo-cleft since it is part of the extended projection of the VP.]

Extracted from <SACE-IU-2009-DCMV>

[Students' scores indicate that questions addressing narrator point of view were the easiest to answer, followed by questions regarding narrator tone, the stories' predominant theme and, **finally**, the mood of the characters or characters.]

[**Finally**, the length of the recall per text version is shown by a percentage, in comparison to the length of the text that was read.]

Extracted from <NACE-UI-2005-MEO>

Likewise, *most* was found in SACE approximately twice as in NACE with a +134.01 LL value. Thus, it was used 923 by SAEs while it was preferred 494 times by NAEs, some of whose statements are presented below to exemplify the adverb in concern.

Example 42.

[In 32 out of the 44 cases (72%), both parents are in employment, with **most** of them being self-employed or public sector employees, which, for these families, permits a certain degree of economic stability.]

[The **most** popular topics among participants are animals and sports with 25% each.]

[A large proportion of the vocabulary that appears in the course book corresponds to low levels of frequency, that is, above the 1,000 or 2,000 **most** frequent words.]

Extracted from <SACE-UM-2011-GAM>

[Two or three show signs of the counterargument and rebuttal strategy of the challenge; non-CLIL Text 3 has perhaps the **most** sophisticated example of this approach.]

[This is **most** obviously the case with text organisation, which will therefore not be included here, but also affects other results.]

Extracted from <NACE-UE-20011-CMR>

In the list of top 50 adverbs in SACE, 11 items seem to have been underused by SAEs. For instance, *significantly* occurred 242 times in SACE while it appeared 457 times in NACE with a -66.19 LL value.

Example 43.

[The TD children were **significantly** affected by PP such that low PP non-words were less accurately repeated than high PP non-words.]

[There was a significant main effect of group at both Data Points I and 4, demonstrating that, as has been found in many previous studies, the LI group were **significantly** poorer in their ability to repeat non-words than TD children when compared with respect to age.]

[In summary therefore the LI group were **significantly** poorer than the TD group with respect to age in their abilities to repeat both high and low PP non-words.]

Extracted from <NACE-NU-2009-CM>

[Nevertheless, the number of behaviour-based and colour-based metaphor terms designating body parts of marine organisms is **significantly** lower than that of behaviour-based and colour-based metaphor terms designating marine organisms.]

[The total number of 45,173 non-resemblance metaphor occurrences is **significantly** higher than that of resemblance metaphor terms, which was 32,247.]

Extracted from <SACE-UG-2011-JMUG>

Again, on the other hand, was used 180 times by SAEs whereas it was preferred 332 times by NAEs, corresponding to a -44.39 LL value. It is exemplified in the following statements drawn from both corpora.

Example 44.

[Almost half the patients who applied for appointments at the desk were asked to call **again** later whereas appointments were always given at B and C, unless patients asked for dates and times well into the future36.

[It is also **again** the receptionist who resumes the task focus when she asks for the patient's address after providing relief.]

[The exception **again** proves the rule.]

Extracted from <NACE-UE-2006-HH>

[This type of joke once **again** exploits the linguistic similarities between Danish and Norwegian to create its punch line.]

[As suggested above, this may **again** be due to the importance of AIDS awareness in South African society.]

Extracted from <SACE-UPF-2012-JFMB>

Another adverb considerably underused by SAEs is *then,* which was found 324 and 518 times in SACE and NACE, respectively. The following extracts are intended to exemplify it.

Example 45.
[**Then**, the students will be compared against each other in order to see if there is any significant difference between the EPE and IIEP students, or between the Korean and the Chinese/Taiwanese students.]

[All the participants had only 30 minutes in which to read two essay questions presented to them, select one, and **then** compose and finalize their essay.]

[In the case of Student 7, **then**, the following figure shows that two long co-reference chains co-exist with several shorter chains.]

Extracted from <NACE-UI-2010-JEG>

[Her figure could **then** be perceived as link between an earlier praise for women's intervention in creativity and prosperity, and a growing male conscience of social construction.]

[The authors‟ use of time framing **then** enables them to interrogate the substantiality of the actions that are portrayed.]

Extracted from <SACE-UH-2010-APV>

Seen 1608 times in SACE, *more* reveals to be underused by SAEs against NAEs, who preferred it 1975 times in their works. It is included in the statements taken from the corpora in concern.

Example 46.

[The peripheral, **more** semantically contentful cases are situated in P, the genitive and partitive in D, accusative variably in D and -, and nominative is the bare noun itself, in these languages.]

[P-cases tend **more** often to be the cases of adverbials and idiosyncratically selected arguments, whereas D-cases appear more regularly on arguments.]

[In directional examples the PPs are **more** like direct objects than those in the locative examples.]

Extracted from <NACE-UU-2008-AA>

[It is in this last group of legends where women are **more** strongly related to the idea of sovereignty, probably because they focus on kingship and the power of rulers.]

[But given the general content of the book, it is definitely **more** intended to promote this behaviour in women than in men.]

Extracted from <SACE-UH-2010-APV>

As in the case of adverbs, SACE and TACE seem to be significantly different from each other. That is, SACE displays an apparent overuse against NACE with respect to the use of adverbs that were distinctively underused in TACE. It is seen that a total number of 29 adverbs in the top 50 adverb lists of three corpora are identical. Here, it might be convenient to put an overall picture of three corpora with regard to the distribution of these adverbs. The only adverb they appear to be similar in frequency is *very*, which was used 492 times in TACE and SACE, and 525 times in NACE. The following are the examples taken from each.

Example 47.

[Although the participants used non-conventionally indirect strategies more often when compared to the written discourse completion tests, the percentage of non-conventionally indirect strategies was still **very** limited.]

Extracted from <TACE-AU-2008-AZG>

[The relevance of the mountain is emphasised several times in the text from the **very** first page, where we see that an unknown man has decided to camp there.]

Extracted from <SACE-UZ-2012-MPBA>

[Although spelling out words may occur in classrooms, particularly language classrooms, it would appear to be **very** uncommon in other interactional settings.]

Extracted from <NACE-NU-2011-AB>

It is would not be convenient to mention a similarity among three corpora in the use of certain adverbs such as *below, even, how, less, more, most, often, only, significantly* and *thus*, as illustrated in Figure 20.

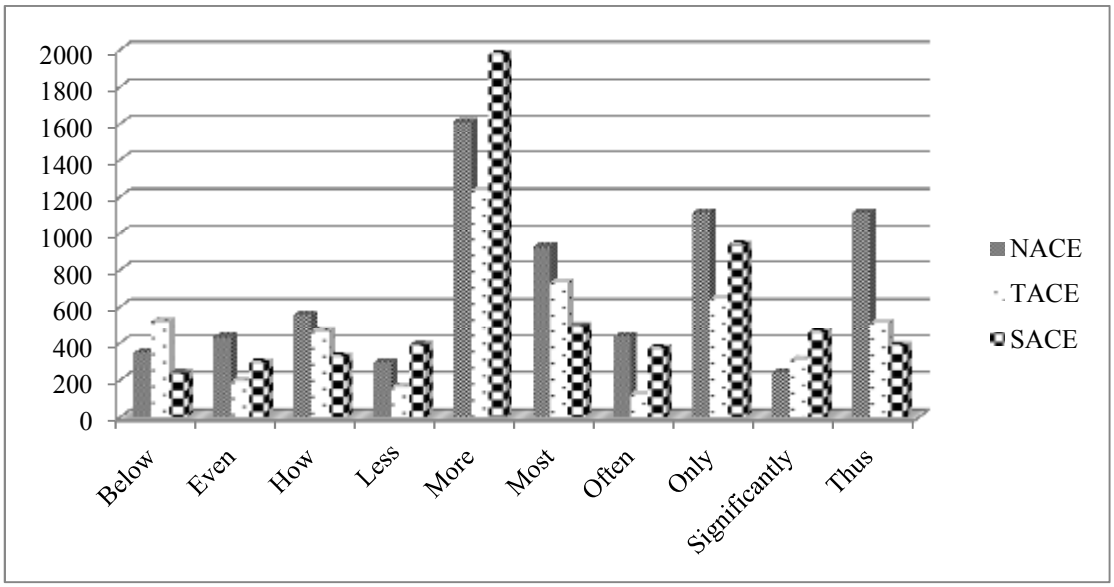


Figure 20. Identical adverbs three corpora significantly differ in use

As suggested in Figure 20, an apparent underuse was observed in both TACE and SACE against NACE with respect to the use of *even, how, most, often, only* and *thus*. *Significantly* seems to be overused in TACE and SACE, whereby it appeared 309 and 457 times, respectively. The following are the examples taken from each corpus.

Example 48.

[Interesting differences also emerged. For example, the findings provided in table 8.10, show that the subjects **significantly** increased their use of addressivity.]

Extracted from <NACE-UE-2007-MP>

[Form revisions and organization revisions, on the other hand, were found to be **significantly** affecting the average essay score and accounting for 35 per cent of its variation with a combined effect.]

Extracted from <TACE-METU-2009-ETD>

[When it comes to overall performance, only the Japanese group shows target-like behaviour, while the L1 Mandarin and the L1 Spanish group differ **significantly** from the native controls.]

Extracted from <SACE-UI-2011-GSCD>

More and *less*, on the other hand, are the ones that were underused by TAEs and overused by SAEs. Namely, they occurred 1234 and 160 times in TACE, respectively and 1975 and 385 times in SACE, respectively. In response, it is seen that *below* was underused by SAEs and overused by TAEs. As for the other items in the list, certain adverbs such as *above, as, both, either, here, just, so* and *still* were found at similar frequencies in NACE and SACE when compared to TACE, which can be seen in Figure 21.

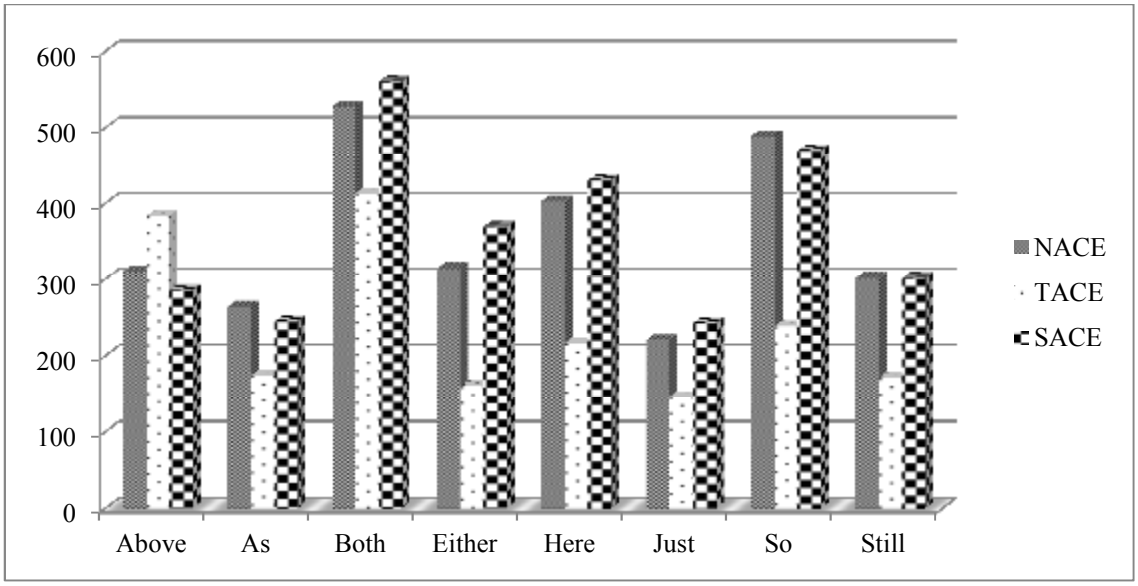


Figure 21. Identical adverbs in SACE and NACE against TACE

As illustrated in Figure 21, *both* was used 528 and 560 times in NACE and SACE, respectively whereas it was found 413 times in TACE. The item in concern is exemplified in the statements extracted from three corpora.

Example 49.

[The first three texts use past-in-present tense in **both** projected and non-projected material, whereas the last three texts use it only in projected material.]

Extracted from <NACE-UM-2009-CES>

[In **both** cases, the majority of the participants constitute twice as much of the remaining people in their categories.]

Extracted from <TACE-CU-2012-HT>

[Finally, I have shown that **both** wh-extraction and embedded coordination present further evidence for the analysis proposed.]

Extracted from <SACE-GU-2009-FJFR>

Likewise, *above* appeared 285 and 310 times in SACE and NACE, respectively while it occurred 383 times in TACE. In return for this, TACE and NACE are more alike in compared to SACE with respect to the use of *therefore, then* and *again*. For instance, *therefore* was employed 406 and 432 times by TAEs and NAEs, respectively while it was preferred 508 times by SAEs. Example 49 illustrates it in the statements drawn from the corpora in concern.

Example 50.

[I **therefore** follow suit and combine these two microfactors into a single 'other' category.]

Extracted from <NACE-UE-2011-RJA>

[**Therefore**, the following strategy is substitution in which learners were taught that they could explain words they do not remember or know in different ways.]

Extracted from <TACE-CU-2011-DIA>

[**Therefore**, I was particularly interested in determining whether students would explore the dynamic layers of identity and the multiple factors of language and power.]

Extracted from <SACE-TAMU-2011-HLGF>

So far, adjectives and adverbs have been described and outlined across NACE, TACE and SACE as well as top fifty lists of the corpora in concern have been covered for each lexical category. The following section provides findings related to the use of nouns across three corpora.

4.1.5.3. Nouns in Three Corpora

Not surprisingly, nouns were found to be the most frequently used lexical items in NACE, TACE and SACE (See Figure 16 and Table 40). Figure 22 shows epistemic nouns as well as all nouns found in three corpora.

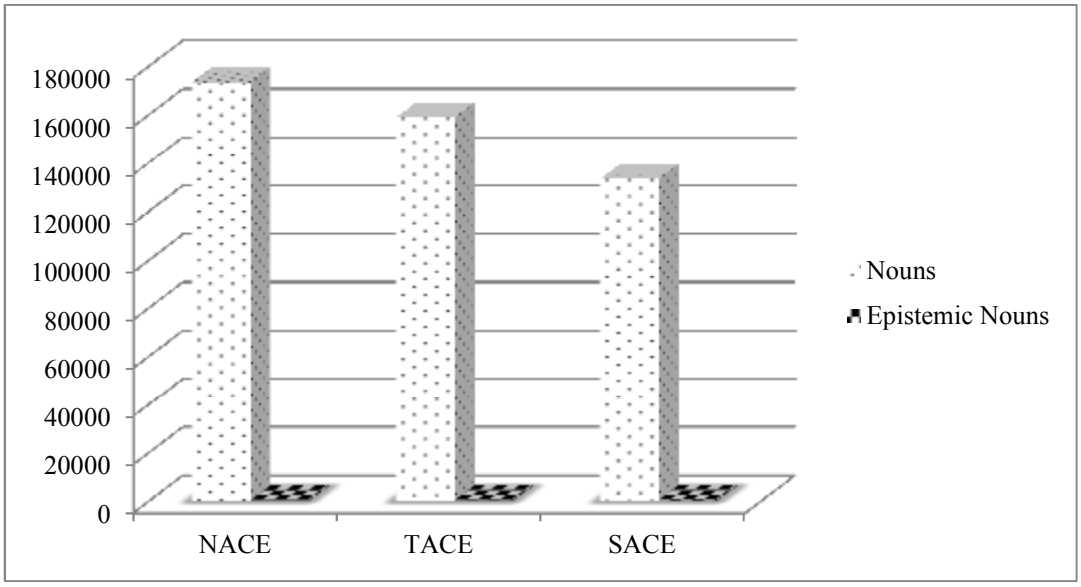


Figure 22. Epistemic nouns and all nouns in three corpora

As seen in Figure 22, epistemic nouns appeared at very low frequencies in all corpora. However, in parallel with the difference observed among three corpora with respect to the use of nouns, proportions of epistemic nouns to all nouns slightly differ across three corpora. For instance, epistemic nouns constitute 0.26% of the nouns in NACE while they form 0.34% and 0.47% of the nouns in TACE and SACE, respectively.

Three corpora were analysed to investigate top fifty noun lists. The one consists of top fifty nouns in NACE are presented in Table 56.

Table 56

Most Frequently Used 50 Nouns in NACE

Nouns	n	T/t %	n per 10.000
Group	2605	0.39	38.80
Student	2459	0.37	36.62
Text	2386	0.36	35.53
Child	2077	0.31	30.93
Language	2021	0.30	30.10
Word	1793	0.27	26.70
Result	1468	0.22	21.86
Difference	1458	0.22	21.71
Effect	1371	0.20	20.42
Level	1307	0.19	19.46
Speaker	1253	0.19	18.66
Learner	1231	0.18	18.33
Study	1192	0.18	17.75
Analysis	1127	0.17	16.78
Chapter	1098	0.16	16.35
Participant	987	0.15	14.70
Use	985	0.15	14.67
Verb	971	0.14	14.46
Type	960	0.14	14.30
English	958	0.14	14.27
Table	921	0.14	13.72
Number	916	0.14	13.64
Context	893	0.13	13.30
Question	872	0.13	12.99
Section	832	0.12	12.39
Experiment	788	0.12	11.74
Speech	787	0.12	11.72
Task	775	0.12	11.54
Time	770	0.11	11.47
Clause	756	0.11	11.26
Teacher	748	0.11	11.14
Subject	735	0.11	10.95
Research	724	0.11	10.78
Interaction	697	0.10	10.38
Pattern	667	0.10	9.93
Response	634	0.09	9.44
Finding	629	0.09	9.37
Change	603	0.09	8.98
Class	596	0.09	8.88
Figure	585	0.09	8.71
Passive	579	0.09	8.62
Case	578	0.09	8.61
Form – Point	559	0.08	8.32
Example	556	0.08	8.28
Structure	552	0.08	8.22
Knowledge	550	0.08	8.19
Feedback	538	0.08	8.01
Condition	532	0.08	7.92
Test	511	0.08	7.61
Reader	511	0.08	7.61
Evidence	505	0.08	7.52
Total	51316	7.66	764.23

n= raw frequency of nouns in corpus

% T/t= Type/token ratio, percentage of the nouns in overall tokens in corpus

As illustrated in Table 56, the most frequented 50 nouns in NACE compose approximately 30 % of total nouns seen in it. They were found 764.23 times in every 10.000 words in NACE. *Group*, which occurred 38.80 times in every 10.000 words, is the most frequently used noun by NAEs. *Student* is the second and *text* is the third mostly used ones in NACE occurring 36.62 and 35.53 times in every 10.000 words in the corpus, respectively. Table 57 displays the most used 50 nouns in TACE.

Table 57

Most Frequently Used 50 Nouns in TACE

Nouns	n	T/t %	n per 10.000
Participant	4375	0.65	64.81
Student	3910	0.58	57.92
Teacher	3283	0.49	48.63
Group	2224	0.33	32.94
Study	1924	0.29	28.50
Language	1871	0.28	27.72
Result	1787	0.26	26.47
Learner	1782	0.26	26.40
Strategy	1515	0.22	22.44
Word	1416	0.21	20.98
Verb	1395	0.21	20.66
Sentence	1330	0.20	19.70
Table	1283	0.19	19.01
Test	1086	0.16	16.09
Analysis	1083	0.16	16.04
Level	1009	0.15	14.95
Difference	1004	0.15	14.87
Question	1002	0.15	14.84
English	928	0.14	13.75
Type	897	0.13	13.29
Score	891	0.13	13.20
Use	875	0.13	12.96
Task	856	0.13	12.68
Learning	798	0.12	11.82
Finding	789	0.12	11.69
Category	744	0.11	11.02
Writing	742	0.11	10.99
Vocabulary	738	0.11	10.93
Error	719	0.11	10.65
Activity	718	0.11	10.64
Noun	707	0.10	10.47
Time	696	0.10	10.31
Example	675	0.10	10.00
Class	667	0.10	9.88
Research	662	0.10	9.81
Item	657	0.10	9.73

(Table 57 Continued)

Form	600	0.09	8.89
Subject	589	0.09	8.72
Skill	565	0.08	8.37
Frequency	564	0.08	8.35
End	562	0.08	8.33
Extraction	561	0.08	8.31
Change	552	0.08	8.18
Extract	517	0.08	7.66
Way	512	0.08	7.58
Teaching	507	0.08	7.51
Clause	502	0.07	7.44
Reason	499	0.07	7.39
Part	486	0.07	7.20
Total	54837	8.14	812.31

n= raw frequency of nouns in corpus

% T/t= Type/token ratio, percentage of the nouns in overall tokens in corpus

As can be seen in Table 57, the most frequented 50 nouns in TACE constitute 34.48 % of total nouns found in the corpus at stake. They occurred over 812 times in every 10.000 words in TACE whereas they appeared 749 times per 10.000 words in NACE. It is seen that 29 nouns in the lists are identical in TACE and NACE. 19 nouns in the list of TACE do not appear in that of NACE. It is particularly interesting that *strategy, sentence* and *score*, which were found 1515, 1330 and 891 times across TACE, respectively do not exist in the top 50 list of NACE. The other nouns which are among the top 50 nouns in TACE but do not appear in the list of NACE are the following: *activity, category, end, error, extraction, frequency, item, learning, noun, part, reason, skill, teaching, vocabulary, way* and *writing*. Finally, the ones mostly employed by SAEs have been determined and presented in Table 58.

Table 58

Most Frequently Used 50 Nouns in SACE

Nouns	n	T/t %	n per 10.000
Language	1828	0.27	27.35
Word	1586	0.24	23.73
Student	1388	0.21	20.77
Result	1080	0.16	16.16
Group	1031	0.15	15.43
Study	1010	0.15	15.11
Learner	981	0.15	14.68
English	951	0.14	14.23
Verb	933	0.14	13.96
Strategy	909	0.14	13.60
Text	801	0.12	11.99
Use	799	0.12	11.96
Case	790	0.12	11.82
Form	776	0.12	11.61
Woman	769	0.12	11.51
Example	761	0.11	11.39
Analysis	729	0.11	10.91
Data	725	0.11	10.85
Corpus	673	0.10	10.07
Structure	630	0.09	9.43
Difference	622	0.09	9.31
Term	607	0.09	9.08
Face	594	0.09	8.89
Type	588	0.09	8.80
Category	583	0.09	8.72
Meaning	550	0.08	8.23
Number	517	0.08	7.74
Research	504	0.08	7.54
Noun	504	0.08	7.54
Way	503	0.08	7.53
Vocabulary	481	0.07	7.20
Teacher	474	0.07	7.09
Move	467	0.07	6.99
Context	461	0.07	6.90
Speaker	457	0.07	6.84
Image	453	0.07	6.78
Novel	439	0.07	6.57
Participant	432	0.06	6.46
Adjective	431	0.06	6.45
Discourse - Sequence	428	0.06	6.40
Fact	418	0.06	6.26
Sentence	416	0.06	6.23
Level	408	0.06	6.11
Figure - Question	403	0.06	6.03
Part	399	0.06	5.97
Table	394	0.06	5.90
Section	381	0.06	5.70
Metaphor	375	0.06	5.61
Instance - Chapter	367	0.05	5.49
Disagreement	365	0.05	27.35
Total	33171	4.97	490.92

n= raw frequency of nouns in corpus

% T/t= Type/token ratio, percentage of the nouns in overall tokens in corpus

The most used 50 nouns in SACE compose approximately 25% of total nouns found in it. They appeared over 490 times in every 10.000 words in SACE. It is observed that 30 adverbs in the top 50 lists of SACE and NACE are identical and 23 found in the list of SACE do not exist in that of NACE. An interesting finding might be that 7 of the ones not shared by NACE regarding the lists in concern exist in the counterpart list of TACE (*category, noun, part, sentence, strategy, vocabulary* and *way*). Following are the other items that are not identical in SACE and NACE: *adjective, corpus, data, disagreement, discourse, face, fact, image, instance, meaning, metaphor, move, novel, sequence, term,* and *woman*. The most frequently used noun in SACE is *language* occurring 27.35 times in every 10.000 words in it. *Word* is the second and *student* is the third most common nouns in the list. They were used 23.73 and 20.77 times per 10.000 words in SACE.

A log likelihood test was run over all corpora to find out whether they significantly differ with respect to the identical nouns in the top 50 lists for each. The first analysis was administered between TACE and NACE and the results are displayed in Table 59.

Table 59

LL Ratio of Identical Nouns in TACE and NACE (Top 50)

Nouns	TACE n	NACE n	LL Ratio (*p< 0.05)
Participant	4375	987	+2294.49
Teacher	3283	748	+1707.18
Student	3910	2459	+325.83
Test	1086	511	+208.71
Study	1924	1192	+169.71
Learner	1782	1231	+98.43
Verb	1395	971	+76.16
Table	1283	921	+57.82
Result	1787	1468	+26.94
Finding	789	629	+17.25
Example	675	556	+10.9
Question	1002	872	+8.35
Class	667	596	+3.63
Task	856	775	+3.61
Form	600	559	+1.24
English	928	958	-0.65
Analysis	1083	1127	-1.13
Type	897	960	-2.49
Change	552	603	-2.53
Time	696	770	-4.14
Language	1871	2021	-6.61
Use	875	985	-7.11
Research	662	724	-9.35
Subject	589	735	-16.92
Group	2224	2605	-32.15
Level	1009	1307	-40.05
Word	1416	1793	-46.42
Clause	502	756	-53.00
Difference	1004	1458	-86.63

n=raw frequency of nouns in corpus

+ indicates overuse in TACE relative to NACE

- indicates underuse in TACE relative to NACE

As indicated in Table 59, identical nouns were relatively overused in TACE against NACE. *Participant* is the most overused noun by TAEs with a +2294.49 LL value. Thus, it was employed 4375 times by TAEs whereas it was preferred 987 times by NAEs, which means it was used in TACE over 4.4 times as much in NACE. The following are the examples extracted from a single work by a TAE, where it was employed a total of 54 times and a work by a NAE.

Example 51.

[The data gathered from **participant** teachers' weekly discussions and journals was analyzed using content analysis conducted so as to specify emergent patterns and themes relevant to the participants' conceptualization of TEYL.]

[The **participants** also expressed the change in their construing different aspects of TEYL such as teaching four skills, lesson planning, assessment, assigning homework, etc.]

Extracted from <TACE-CU-2007-MS>

[The studies included in that review did not include any information about whether the **participants** were from a Pakistani heritage background.]

Extracted from <NACE-NU-2006-CS>

The second most overused noun in TACE is *teacher*, which were 3283 and 748 times in TACE and NACE, respectively. *Student* is the third in this category with a +325.83 LL value. Namely, it occurred 3910 times in TACE whereas it was found 2459 times in NACE. Similarly, *test* and *study* are the other identical nouns significantly overused by TAEs against NAEs.

As for the underused nouns in TACE, *difference* seems to be at the top of the list with a -86.63 LL value. It was employed 1458 times in NACE while it was seen 1004 times in TACE. The following are the expressions drawn from a single work in NACE, whereby it was used 38 times and from one in TACE.

Example 52.

[In no game was there a significant between-team **difference** in within-team convergence (based on a Monte Carlo test).

[It should be noted that there remain important **differences** between language evolution and biological evolution, even if phonetic tokens are treated as evolutionary replicators.]

[As can be seen Section 7.1.7, however, there were small **differences** between the two experiments beyond frequency of interaction.]

Extracted from <NACE-UE-2010-GR>

[The same questionnaire was given before the training started but there were some slight **differences** in the statements.]

Extracted from <TACE-GU-2012-DBA>

Clause is the second and *word* is the third mostly underused nouns in TACE with -53.00 and -46.42 LL values, respectively.

The second analysis was carried out between SACE and NACE and the results are provided in Table 60.

Table 60

LL Ratio of Identical Nouns in SACE and NACE (Top 50)

Nouns	SACE n	NACE n	LL Ratio (*p< 0.05)
Case	790	578	+34.01
Form	776	559	+33.47
Example	761	556	+33.03
Structure	630	552	+5.53
English	951	958	+0.00
Verb	933	971	-0.59
Language	1828	2021	-8.78
Word	1586	1793	-11.71
Study	1010	1192	-14.20
Use	799	985	-18.54
Learner	981	1231	-27.13
Figure	403	585	-32.85
Research	504	724	-38.58
Result	1080	1468	-57.46
Teacher	474	748	-60.65
Analysis	729	1127	-84.11
Type	588	960	-88.50
Number	517	916	-110.67
Context	461	893	-138.20
Section	381	832	-169.62
Question	403	872	-174.39
Table	394	921	-214.73
Participant	432	987	-220.32
Student	1388	2459	-297.02
Difference	622	1458	-341.69
Chapter	367	1098	-378.13
Speaker	457	1253	-381.42
Level	408	1307	-491.32
Group	1031	2605	-696.89
Text	801	2386	-816.89

n=raw frequency of nouns in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

As shown in Table 60, the identical nouns were widely underused in SACE against NACE. That is, only five items were overused while 25 nouns were considerably underused by SAEs. *Case* is the most frequently overused noun by SAEs with a +34.01 LL value. Example 53 provides statements taken from a single work by a SAE, who employed it 32 times and from one in NACE, where it appeared 12 times.

Example 53.
[By contrast, in the **case** of Spanish doctors who write first in Spanish, only half of them perform the translation themselves.]

[As with other nouns in the list, "grupo/s" is sometimes accompanied by a past participle forming in this **case** "grupo tratado".]

[As it was the **case** with "high", the English use of this adjective is wide and the Spanish researcher will find no difficulty in learning collocation patterns with this adjective in medicine.]

Extracted from <SACE-UPV-2008-ARV>

[However, this is not the **case** in ARCHER, where the frequency of phrasal verbs in British English and American English is very similar in each period.]

Extracted from <NACE-UG-20010-CW>

Form and *example* are the second and third ones in this category. Namely, the former was employed 776 and 578 times in SACE and NACE, respectively. The latter occurred 761 in SACE and 556 times in NACE, which entails a +33.03 LL value.

The most underused noun in SACE is *text*, which was used 801 times by SAEs and 2386 times by NAEs, corresponding to a -816.89 LL value. It is exemplified in the following statements drawn from a single work in NACE, across which it was employed 78 times and from one in SACE, whereby it occurred 7 times.

Example 54.
[These statistics reveal differences between the surgical and sales **texts** on the one hand and the casual conversations and interviews on the other.]

[That is, in this text as well, one may consider that the patterns of selection found in this excerpt might differ from those of the succeeding **text**.]

Extracted from <NACE-UE-2008-BAS>

[Rather, the teacher encourages meaning-making to occur between the **text** being studied and students' lived experiences.]

Extracted from <SACE-TAMU-2011-HLGF>

The second and the third items in this list are *group* and *level*. For instance, *group* was employed 1301 times across SACE while it was found 2605 times in NACE. Likewise, *level* appeared 408 times in SACE whereas it occurred 1307 times in NACE, which corresponds to a -491.32 LL value.

All in all, as in the case of adjectives, SACE and TACE, significantly differ from each other in the use of nouns that appear in the top 50 lists. What is new and interesting is that TACE exhibits a remarkable overuse against NACE regarding the use of nouns, which were notably underused in SACE. At this point, it is considered useful to discuss three corpora with respect to the distribution of these nouns. A total number of 21 nouns seem to be identical in three corpora; however, this does not mean that they were found at similar proportions over the corpora in question. The only noun of which frequency seems relatively similar in three corpora is *English*. It was used 928, 951 and 958 times in TACE, SACE and NACE, respectively. Figure 23 displays the distribution of the items three corpora significantly differ from each other.

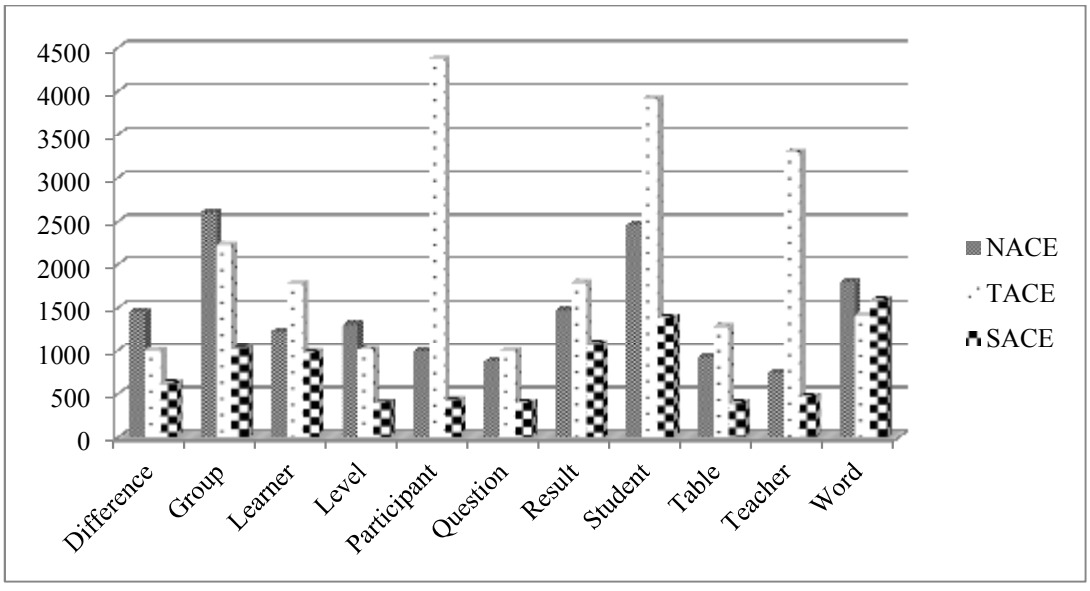


Figure 23. Identical nouns three corpora significantly differ in use

As indicated in Figure 23, there is a significant difference across three corpora in the use of certain nouns such as *difference, group, learner, level, participant, question, result, student, table, teacher* and *word*. For example, *difference* was remarkably underused in TACE and SACE against NACE; namely, it appeared 1004 times in TACE and 622 times in SACE whereas it occurred 1458 times in NACE. It is given in the following extracts taken from each corpus.

Example 55.
[The possession style is more common than the presence one, although there are again **differences** between the three practices, as shown in Table 5.7.]

Extracted from <NACE-UE-2006-HH>

[This means that if there is a statistically significant **difference** between the groups on the post-tests, it will be due to the instructional intervention.]

Extracted from <TACE-CU-2011-EA>

[My analysis will reveal that Jane Eyre is a clear intertext in Tooth and Claw, exploring the coincidences and **differences** between the two novels.]

Extracted from <SACE-UZ-2012-MPBA>

In a similar vein, they significantly differ in the use of *group*. It was underused both in TACE and SACE; namely, it appeared 2224 and 1031 times in TACE and SACE, respectively while it was found 2605 times in NACE. The following statements were drawn from each.

Example 56.
[The participants in the L2 **group** were native speakers of English, who had learned German after the age of 11.]

Extracted from <NACE-UE-2009-FW>

[Those who were instructed the grouping strategy could easily identify the **groups** in the list.]

Extracted from <TACE-CU-2011-DIA>

[Similar to the case of age, the fact that all children belong to the same school, and to the same place, favours the homogeneity of the **group**.]

Extracted from <SACE-UM-2011-GAM>

The other nouns three corpora considerably differ are *participant, student* and *teacher*. *Participant* was found over three times as its total frequency in NACE and SACE; namely, it occurred 4375 times in TACE, 987 in NACE and 432 in SACE. The statements in the following examples were extracted from the corpora in question.

Example 57.

[Three **participants** were excluded from the study because they chose inappropriate control items over 40% of their all responses.]

Extracted from <TACE-AU-2008-AZG>

[It also includes a detailed description of the retrospective interviews consisting of the **participants'** views on the quantitative findings, and stimulated recall sessions as well as the discussions of the hypotheses.]

Extracted from <TACE-GU-2010-CB>

[The following chapter, detailing the word learning abilities of the study **participants** may shed some light on this explanation.]

Extracted from <NACE-NU-2009-CM>

[The **participants** of our experiment found them very useful but demanded more illustrative examples of sentences that should be produced by them from CQs.]

Extracted from <SACE-UPM-2011-EMP>

Likewise, *student* was used 3910 times in TACE, 2459 times in NACE and 1388 times in SACE, which means that it was significantly overused by TAEs and underused by SAEs. The following expressions were taken from three corpora.

Example 58.

[Language teachers should be given in-service teacher training programs on how to use humorous language in the classrooms to develop target language skills of **students** more effectively.]

Extracted from <TACE-GU-2010-CY>

[Nevertheless, unlike the first teacher, he is found to be challenging and quite understanding and friendly by his **students**.]

Extracted from <TACE-GU-2012-DBA>

[**Students** in this particular district may begin language study formally in the 7th grade; the language program culminates with Advanced Placement (AP) offerings in all three languages.]

Extracted from <NACE-UI-2012-EJE>

[As discussed in Chapter 3, her subjects were university **students** taking intermediate Spanish lessons.]

Extracted from <SACE-DU-2006-CGMA>

As for *teacher*, it was overused by TAEs and underused by SAEs. It was employed 3283 times in TACE whereas it was found much less frequently in NACE and SACE (NACE: 748; SACE: 474). Below are the extracts drawn from each corpus.

Example 59.

[In order to eliminate the errors, **teachers** may focus on the differences between two languages.]

Extracted from <TACE-CU-2006-DE>

[The present study guides **teachers** and material developers since the results of the study give an idea about what should be taught about connectives in classrooms, in textbooks, and in other teaching materials.

Extracted from <TACE-AU-2009-DA>

[The IIEP students have **teacher** ratings of their essays that can be examined, as well as the teachers' decision regarding what level in which to place the students.]

Extracted from <NACE-UI-2010-JEG>

[The **teachers**' output and an early introduction to meaning differences between L1 and L2 appear to be effective.]

Extracted from <SACE-USC-2012-LRV>

Another finding obtained from the analysis is that TACE and NACE display similarities in the use of certain nouns against SACE such as *analysis, example, form, research* and *type*. For example, *analysis* was found 1083 and 1127 times in TACE and NACE, respectively while it was used 729 times in SACE. It is exemplified in the following extracts.

Example 60.

[The results of this **analysis** reveal the fact that t-test results of student and teacher questionnaire have significant difference.]

Extracted from <TACE-HU-2008-IFA>

[The interpretation of differences between aspirational statements and those referring to perceived current dynamics is woven into **analysis**.]

Extracted from <NACE-UL-2011-CWM>

Likewise, *type* was employed 897 and 960 times in TACE and NACE, respectively whereas it appeared 588 times in SACE. On the other hand, there are some nouns that had a relatively balanced distribution over SACE and NACE in compared to TACE. For instance, *study* was used 1192 and 1110 times in NACE and SACE, respectively while it occurred 1924 times in TACE. The following extracts were taken from SACE and NACE.

Example 61.

[I collected the data for this **study** from individual interviews with teachers during the spring semester of 2008.]

Extracted from <NACE-UI-2010-AH>

[What is of crucial importance for this **study** is the fact that in early Irish belief the equation woman-nature appeared as a remarkable issue.]

Extracted from <SACE-UH-2010-APV>

In addition, *verb* was seen 971 and 933 times in NACE and SACE whereas it was found approximately 1400 times in TACE. The only noun TACE and SACE seem identical in use against NACE is *language*, which was preferred 1871 and 1828 times

by TAEs and SAEs while it was used 2021 times in NACE. It is exemplified in the following statements drawn from both.

Example 62.
[Improvement in different **language** skills like writing or speaking is one of the main concerns in this study.]

Extracted from <TACE-CU-2006-NK>

[They have learned and used English as a second **language** on a daily basis as they have lived and were currently living in an English-speaking context.]

Extracted from <SACE-UI-2012-JMPR>

In this section, nouns have been identified across NACE, TACE and SACE and the top fifty list of each corpus has been discussed. In the following section, findings of the analysis administered to three corpora with respect to the use of verbs are presented.

4.1.5.4. Verbs in Three Corpora

As mentioned previously, verbs revealed to be the second most frequently used lexical items in TACE and third in both NACE and SACE. However, the corpora in concern do not vaguely differ from each other in the use of these items when the frequency table of lexical categories is considered (See Figure 16 and Table 41). Namely, verbs were found 55288 times in TACE, 58491 times in NACE and 60971 times in SACE. They significantly differ in the proportion of epistemic verbs in all verbs seen across them, as outlined in Figure 24.

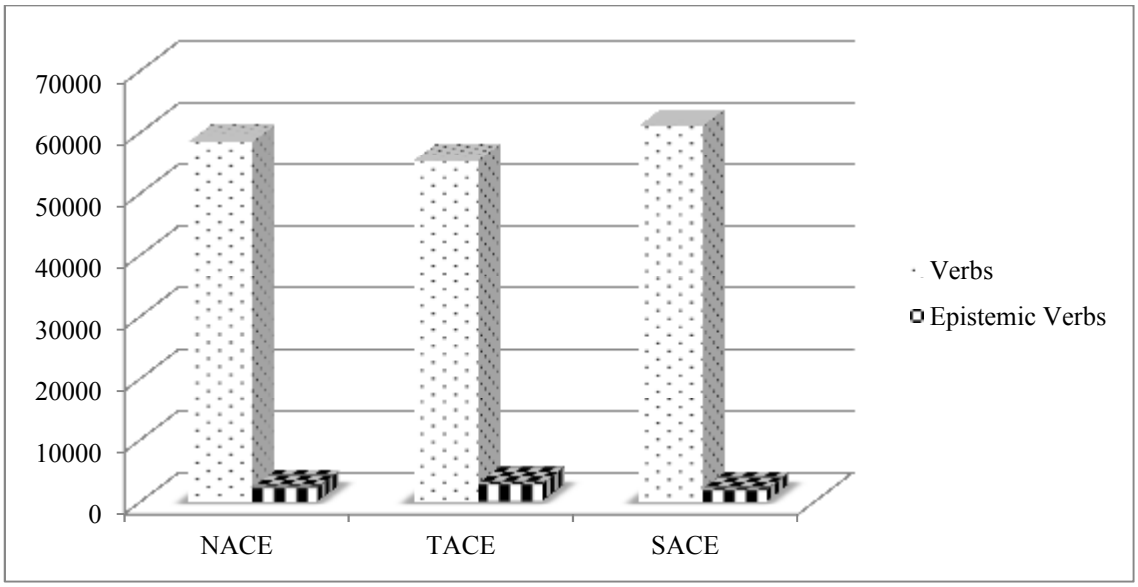


Figure 24. Epistemic verbs and all verbs in three corpora

The proportion of epistemic verbs in all verbs was measured highest in TACE, where epistemic verbs constitute approximately 5.3% of all verbs. In NACE, 4% of all verbs are comprised of epistemic verbs while this proportion is realized 3.3% in SACE. Three corpora were analysed to identify top fifty lists of nouns seen across each. Table 61 displays the results obtained from the analysis of NACE.

Table 61

Most Frequently Used 50 Verbs in NACE

Verbs	n	T/t %	n per 10.000
Show	1750	0.26	26.06
Use	1587	0.24	23.63
Make	1002	0.15	14.92
Find	917	0.14	13.66
Suggest	849	0.13	12.64
See	814	0.12	12.12
Produce	788	0.12	11.74
Give	716	0.11	10.66
Indicate	693	0.10	10.32
Seem	672	0.10	10.01
Present - Follow	623	0.09	9.28
Discuss	593	0.09	8.83
Occur	554	0.08	8.25
Take	551	0.08	8.21
Appear	536	0.08	7.98
Provide	531	0.08	7.91
Compare	523	0.08	7.79
Consider	501	0.07	7.46

(Table 61 Continued)

Identify	394	0.06	5.87
Demonstrate	366	0.05	5.45
Relate	363	0.05	5.41
Report	360	0.05	5.36
Include - Work	326	0.05	4.85
Describe	319	0.05	4.75
Learn - Write	317	0.05	4.72
Ask	304	0.05	4.53
Need - Speak	293	0.04	4.36
Look	278	0.04	4.14
Support - Reflect	275	0.04	4.10
Involve	273	0.04	4.07
Allow	271	0.04	4.04
Examine	270	0.04	4.02
Mean	268	0.04	3.99
Represent	263	0.04	3.92
Become	262	0.04	3.90
Analyse/ Analyze	261	0.04	3.89
Teach	261	0.04	3.89
Explain	260	0.04	3.87
Investigate	256	0.04	3.81
Understand	255	0.04	3.80
Affect - Refer	242	0.04	3.60
Develop	240	0.04	3.57
Require - Create	239	0.04	3.56
Tend	234	0.03	3.48
Observe	232	0.03	3.46
Expect - Revealed	223	0.03	3.32
Acquire	222	0.03	3.31
Read	212	0.03	3.16
Note	210	0.03	3.13
Increase	207	0.03	3.08
Total	22956	3.42	341.87

n= raw frequency of verbs in corpus

% T/t= Type/token ratio, percentage of the verbs in overall tokens in corpus

As shown in Table 61, the most frequented 50 verbs in NACE were found approximately 342 times in every 10.000 words in it and they constitute over 39% of all verbs seen across it. Found 26.06 times every 10.000 words in NACE, *show* reveals to be the most frequently used verb in it. *Use* and *make* are the second and third mostly used verbs in NACE appearing 23.63 and 14.92 times per 10.000 words, respectively across it. The most used 50 verbs in TACE are given in Table 62.

Table 62

Most Frequently Used 50 Verbs in TACE

Verbs	n	T/t %	n per 10.000
Use	2235	0.33	33.11
Show	1130	0.17	16.74
Make	1048	0.16	15.52
Give	1007	0.15	14.92
See	1001	0.15	14.83
Find	840	0.12	12.44
Teach	836	0.12	12.38
Learn	737	0.11	10.92
Write	705	0.10	10.44
Indicate	696	0.10	10.31
Present	647	0.10	9.58
Seem	605	0.09	8.96
State	587	0.09	8.70
Take	566	0.08	8.38
Ask	553	0.08	8.19
Produce	545	0.08	8.07
Follow	528	0.08	7.82
Observe	524	0.08	7.76
Reveal	495	0.07	7.33
Relate	477	0.07	7.07
Read	462	0.07	6.84
Think	447	0.07	6.62
Choose	422	0.06	6.25
Mean	393	0.06	5.82
Explain	381	0.06	5.64
Provide	362	0.05	5.36
Compare	349	0.05	5.17
Understand	347	0.05	5.14
Analyze / Analyse	338	0.05	5.01
Become	337	0.05	4.99
Know	333	0.05	4.93
Prefer	331	0.05	4.90
Get	324	0.05	4.80
Start	322	0.05	4.77
Say	318	0.05	4.71
Need	315	0.05	4.67
Display	311	0.05	4.61
Express - Suggest	300	0.04	4.44
Mention	299	0.04	4.43
Consider	294	0.04	4.36
Discuss	290	0.04	4.30
Help	263	0.04	3.90
Focus	256	0.04	3.79
Require	255	0.04	3.78
Believe	253	0.04	3.75
Feel	251	0.04	3.72
Include	245	0.04	3.63
Speak	240	0.04	3.56
Develop	238	0.04	3.53
Change	237	0.04	3.51
Total	25275	3.77	374.40

n= raw frequency of verbs in corpus

% T/t= Type/token ratio, percentage of the verbs in overall tokens in corpus

As illustrated in Table 62, the most frequented 50 verbs in TACE constitute approximately 46% of total verbs found in it. They appeared over 374 times in every 10.000 words in TACE whereas they occurred nearly 342 times per 10.000 words in NACE. It is observed that 35 nouns in the lists are identical in TACE and NACE. 17 verbs in the list of TACE do not exist in that of NACE. It is noteworthy that *choose, state* and *think*, which occurred 422, 587 and 447 times across TACE, do not appear in the top 50 list of NACE. The following are the other verbs which exist in the top 50 verb list of TACE and which do not appear in that of NACE: *believe, change, display, express, feel, focus, get, help, know, mention, prefer, say, speak* and *start*. Lastly, top 50 verbs in SACE have been identified and provided in Table 63.

Table 63

Most Frequently Used 50 Verbs in SACE

Verbs	n	T/t %	n per 10.000
Use	1501	0.22	22.46
Show	1236	0.18	18.50
Find	1108	0.17	16.58
Make	926	0.14	13.86
Seem	825	0.12	12.35
See	644	0.10	9.64
Appear	622	0.09	9.31
Take	607	0.09	9.08
Follow	593	0.09	8.87
Provide	522	0.08	7.81
Consider	511	0.08	7.65
Present	489	0.07	7.32
Give	479	0.07	7.17
Learn	456	0.07	6.82
Observe	398	0.06	5.96
Obtain	383	0.06	5.73
Suggest	381	0.06	5.70
Indicate	374	0.06	5.60
Refer	367	0.05	5.49
Mention	345	0.05	5.16
Write	340	0.05	5.09
Create	336	0.05	5.03
Include	334	0.05	5.00
Analyse/ Analyze	324	0.05	4.85
Favour/ Favor	319	0.05	4.77
Report - Mean	311	0.05	4.65
Say	307	0.05	4.59
Compare	304	0.05	4.55
Explain - Occur	297	0.04	4.44
Reveal	292	0.04	4.37
Describe	280	0.04	4.19

(Table 63 Continued)

Need	276	0.04	4.13
Produce	260	0.04	3.89
Discuss	256	0.04	3.83
Understand	251	0.04	3.76
Imply	247	0.04	3.70
Perform	245	0.04	3.67
Relate	236	0.04	3.53
Illustrate	235	0.04	3.52
Examine	234	0.04	3.50
Focus - Establish	233	0.03	3.49
Contain	231	0.03	3.46
Employ	230	0.03	3.44
Support	220	0.03	3.29
Confirm	219	0.03	3.28
Involve	210	0.03	3.14
Know	208	0.03	3.11
Get	205	0.03	3.07
Point	202	0.03	3.02
Identify	196	0.03	2.93
Total	20605	3.09	308.35

n= raw frequency of verbs in corpus

% T/t= Type/token ratio, percentage of the verbs in overall tokens in corpus

As can be concluded from Table 63, the most frequented 50 verbs in SACE occurred more than 308 times per 10.000 words across it and they compose approximately 34% of all verbs found in the entire corpus. The first four verbs in the list are observed to be identical in those appearing in the list of NACE. Namely, *use*, which is the mostly employed verb in SACE, is the second mostly used one in NACE. Similarly, *show* is the second most frequented verb in SACE and first in NACE. Occupying the third rank in the list of SACE, *find* reveals to be fourth mostly used verb in NACE. Finally, two corpora are relatively similar in the use of *make*, which was found 13.86 and 14.92 times per 10.000 words in SACE and NACE, respectively. Overall, a total number of 15 verbs found in the top 50-verb list of SACE do not appear in the counterpart list of NACE. It is noteworthy that five of them reveal to exist in the list of TACE and do not appear in that of NACE, which are *focus, get, know, mention* and *say*. For instance, *mention* was seen 299 and 345 times in TACE and SACE, respectively. Likewise, *say* appeared 318 times in TACE and 307 times in SACE. The other verbs not identical in the lists of SACE and NACE are the following: *confirm, contain, employ, establish, favour/ favour, focus, illustrate, imply, obtain, perform* and *point*.

In order to investigate whether TACE and SACE significantly differ from NACE regarding the use of identical verbs found in the top fifty list of each, the three corpora were administered a log likelihood test. The initial analysis was made between TACE and NACE. The related results are exhibited in Table 64.

Table 64

LL Ratio of Identical Verbs in TACE and NACE (Top 50)

Verbs	TACE n	NACE n	LL Ratio (*p< 0.05)
Teach	836	261	+319.75
Learn	737	317	+174.12
Write	705	317	+152.93
Observe	524	232	+117.18
Use	2235	1587	+113.53
Reveal	495	223	+106.97
Read	462	212	+96.19
Ask	553	304	+74.60
Give	1007	716	+50.79
Mean	393	268	+24.39
Explain	381	260	+23.56
See	1001	814	+20.21
Relate	477	363	+16.07
Understand	347	255	+14.56
Analyse / Analyze	338	261	+10.3
Become	337	262	+9.78
Make	1048	1002	+1.27
Need	315	293	+0.91
Require	255	239	+0.6
Present	647	623	+0.58
Take	566	551	+0.28
Indicate	696	693	+0.03
Develop	238	240	+0.00
Find	840	917	-3.02
Seem	605	672	-3.20
Speak	240	293	-5.03
Follow	528	623	-7.40
Include	245	326	-11.14
Provide	362	531	-31.37
Compare	349	523	-34.12
Produce	545	788	-43.39
Consider	294	501	-53.53
Discuss	290	593	-104.67
Show	1130	1750	-131.56
Suggest	300	849	-270.71

n=raw frequency of verbs in corpus

+ indicates overuse in TACE relative to NACE.

- indicates underuse in TACE relative to NACE

As indicated in Table 64, 23 out of 35 identical verbs in TACE and NACE were overused by TAEs. In accordance with the test results, *teach* is the most frequently overused item in TACE against NACE with a +319.75 LL value. That is, it was employed 836 times in TACE while it was used 261 times in NACE. It is exemplified in the following extracts drawn from both corpora.

Example 63.

[The overall analysis of the responses indicates that students of the participant teachers liked English lessons since they found the lessons enjoyable and effectively **taught**.]

Extracted from <TACE-CU-2007-MS>

[As another passive language skill that is **taught** in all Galician schools, it is not surprising that such a high number claim high competence.]

Extracted from <NACE-UE-2007-MK>

Learn seems to be the second in the list occurring 737 times in TACE and 317 times in NACE. *Write* is another verb significantly overused by TAEs with a +152.93 LL value. The following examples taken from both TACE and NACE are intended to illustrate it.

Example 64.

[On the other hand, the tenth excerpt presented above which was **written** by a native speaker obtained 44.494 and 12 respectively.]

Extracted from <TACE-CU-2011-IU>

[If they choose the right interpretation, they are told so but that the sentence they **wrote** conveys the wrong interpretation.]

Extracted from <NACE-UE-2007-RMH>

The other verbs more frequented in TACE are *observe* and *use* with +117.18 and +113.53 LL values, respectively. The most underused verb in TACE against NACE is *suggest*, which was found 849 times in NACE and 300 times in TACE. The following are the expressions taken from works in TACE and NACE.

Example 65.

[The qualitative data analysis **suggested** that classroom management was an important contributing factor to foreign language student teacher anxiety.]

Extracted from <TACE-AU-2010-MA>

[This chapter also **suggested** that the negative campaign theme, particularly prominent of the third week of the election, arose out of the conflict paradigm.]

Extracted from <NACE-UG-2006-DPS>

Show is the second in this category with a -131.56 LL value and *discuss* is the third one with a -104.67 LL values, respectively. They are illustrated in the following examples.

Example 66.

[Type, number, and percentages of the connectives used by the participants are **shown** in Table 6.]

Extracted from <TACE-AU-2009-DA>

[In discussing their ideal situations the groups **showed** a desire for structure of a specific nature.]

Extracted from <NACE-UL-2011-CWM>

Example 67.

[It is noteworthy to **discuss** these differences and the potential sources of these differences among various grades on the basis of several categories.]

Extracted from <TACE-AU-2010-GS>

[Here I **discuss** a few of these, looking at the way in which the main claims of the dissertation fit in with them.]

Extracted from <NACE-UU-2008-AA>

Subsequently, the same test was run between SACE and NACE to find out whether they significantly differ in terms of verbs listed as top 50 for each. The results are provided in Table 65.

Table 65

LL Ratio of Identical Verbs in SACE and NACE (Top 50)

Verbs	SACE n	NACE n	LL Ratio (*p< 0.05)
Observe	398	232	+45.06
Refer	367	242	+26.44
Learn	456	317	+25.80
Find	1108	917	+18.97
Create	336	239	+16.91
Seem	825	672	+16.41
Reveal	292	223	+9.61
Analyse/ Analyze	324	261	+6.90
Appear	622	536	+6.81
Mean	311	268	+3.41
Take	607	551	+2.98
Explain	297	260	+2.64
Write	340	317	+0.92
Consider	511	501	+0.15
Include	334	326	+0.14
Understand	251	255	-0.02
Provide	522	531	-0.04
Need	276	293	-0.43
Follow	593	623	-0.60
Use	1501	1587	-2.00
Describe	280	319	-2.36
Examine	234	270	-2.40
Make	926	1002	-2.64
Report	311	360	-3.35
Support	220	275	-5.86
Involve	210	273	-7.94
Present	489	623	-15.55
See	644	814	-19.06
Relate	236	363	-26.53
Give	479	716	-46.18
Compare	304	523	-57.64
Occur	297	554	-77.61
Show	1236	1750	-86.47
Indicate	374	693	-95.32
Discuss	256	593	-135.91
Suggest	381	849	-180.39
Produce	260	788	-276.07

n=raw frequency of verbs in corpus

+ indicates overuse in SACE relative to NACE

- indicates underuse in SACE relative to NACE

In accordance with the findings in Table 65, it might be convenient to suggest most of the identical verbs in the top 50 list of SACE were underused against NACE.

Produce is the most significantly underused verb in SACE with a -276.07 LL value. It appeared 788 times in the works of NAEs while it was found 260 times in SACE. The statements including it are presented in the following examples drawn from both corpora.

Example 68.
[One–tailed t-tests for the individual blocks **produced** a significant effect of condition in block one, but not in block two.]

Extracted from <NACE-UE-2009-ALC>

[For instance, there are several references to milk ó the traditional favourite of cats or hisses ó a typical sound **produced** by cats.]

Extracted from <SACE-UZ-2012-MPBA>

Used 381 times in SACE and 849 times in NACE, *suggest* is another verb considerably underused by SAEs. The following are the statements extracted from both corpora.

Example 69.
[This is not a shortcoming of the experimental design, but it at least gave more reason to replicate the experiment, incorporating improvements as **suggested** above.]

Extracted from <NACE-UE-2010-GR>

[As was **suggested** previously in this chapter, this national agency for the assessment of English could be created afresh or by modifying the current role and membership of ACLES.]

Extracted from <SACE-UPV-2012-DLN>

Discuss was also underused by SAEs against NAEs with a -135.91 LL value. It was seen 256 times in SACE and 593 times in NACE. The following extracts were taken from the corpora in concern.

Example 70.

[In the following sections we address the research questions and **discuss** how our data can influence current theories of L2 acquisition.]

Extracted from <NACE-UC-2006-JSG>

[Firstly, we **discuss** the results concerning resemblance metaphor, and secondly, those concerning non-resemblance metaphors.]

Extracted from <SACE-UG-2011-JMUG>

Observe, learn and *refer*, on the other hand, are the most significantly overused verbs in SACE. *Observe* was found 398 times in SACE while it was seen 232 times in NACE, which corresponds to a +45.06 LL value. It is exemplified in the following extracts.

Example 71.

[However, in that scenario, groups 1 and 3 would be the most different, which is not the case **observed** here.]

[With regard to the variable ANIMACY, the disfavoring of overt pronominal subjects with inanimate referents is parallel to the one **observed** for the monolingual Spanish data.]

Extracted from <SACE-PSU-2009-APP>

[Layout of the page and screen can be **observed** through the choice of headings and sub-headings.]

Extracted from <NACE-US-2006-JJ>

Refer was employed 367 times in the works of SAEs whereas it was preferred 242 times in those of NAEs.

Example 72.

[By 'ritual,' I **refer** to the enactment of certain behavioral patterns encoded and established by a certain cultural or sub-cultural group.]

Extracted from <SACE-UCM-2012-CAR>

[This is also evident from examples of participant reference in 7.2 below insofar as different lexical items are used to **refer** to the same participant.]

Extracted from <NACE-UE-2008-OS>

Learn was found 456 times in SACE and 317 times in NACE, which means an overuse with a +25.80 LL value. The statements including it have been extracted from both corpora and given in the following example.

Example 73.

[Due to that, the first hypothesis is corroborated, since students with high level of auditory skill have **learned** better.]

Extracted from <SACE-UH-2011-CMTF>

[It is through the fox's words that his nature is made clear to the audience, when he claims that his two children have **learned** from his example.]

Extracted from <NACE-UG-2012-JG>

To sum up, different portraits were drawn for each corpus with respect to the use of verbs. In particular, TACE and SACE significantly differ from each other in this concern when taking into consideration the facts that most of the identical verbs in the top 50 lists were overused by the former and underused by the latter. Namely, 23 out of 35 identical verbs in TACE and NACE were overused by TAEs whereas 23 out of 38 in SACE and NACE were underused by SAEs.

It would be practical to describe and outline the verbs identical in three corpora. A total number of 28 verbs are shared in the top 50 lists of three corpora; however, there is a balance across them in the distribution of only three of them, which are illustrated in Figure 25.

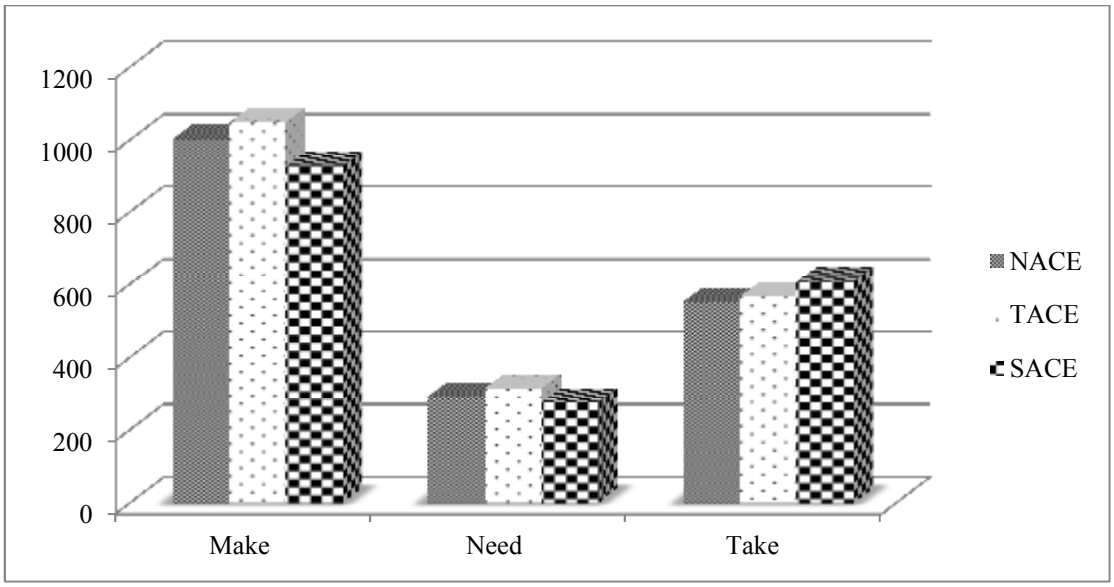


Figure 25. Identical verbs of similar frequencies over three corpora

As shown in Figure 25, *make* was used 1048, 1002 and 926 times in TACE, NACE and SACE, respectively. The following are the statements taken from each corpus.

Example 74.

[The following table (Table 2) shows these frames and their definitions **made** by the researcher relying on the data.]

Extracted from <TACE-CU-2011-AA>

[Similarly, no attempt was **made** to categorise children according to the amount or length of their exposure to English.]

Extracted from <NACE-NU-2006-CS>

[From an intra-lexical standpoint, the compound nature of the words introduced could have **made** their acquisition more difficult.]

Extracted from <SACE-UM-2011-GAM>

Similarly, *need* overused in TACE, whereby it appeared 315 times and underused in SACE, across which it was found 276 times against NACE, in which it occurred 293 times. It is also illustrated in the following examples drawn from each corpus.

Example 75.

[A comparative study among language types is **needed** to evaluate their functional content.]

Extracted from <TACE-HU-2008-AAB>

[In other words, the smaller the bias in the coin, the greater the number of tosses **needed** to detect that bias.]

Extracted from <NACE-UE-2011-RJA>

[As in previous examples, people from other cultures would **need** an explanation in order to decode the meaning of this joke.]

Extracted from <SACE-UPF-2012-JFMB>

Lastly, take was overused in TACE and NACE against NACE appearing 566 and 607 times, respectively. However, this is not a significant difference across them. The following are the extracts drawn from three corpora.

Example 76.

[In this respect, is worth **taking** into account that the appearance of focus within IP and DP could call for a different interpretation.]

Extracted from <SACE-GU-2007-JLC>

[The content for our experimental questions was **taken** directly from authentic texts found in books and magazines.]

Extracted from <NACE-UE-2007-RMH>

Three corpora significantly differ from each other in the use of six verbs, which are *give, learn, observe, produce, relate* and *see*. The distribution of the items in concern over three corpora is displayed in Figure 26.

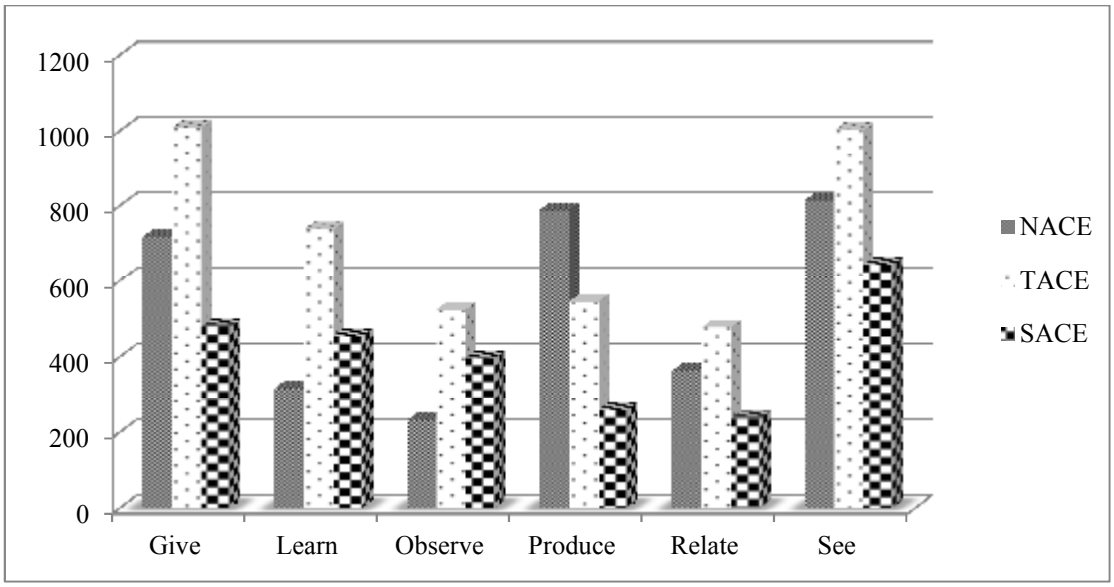


Figure 26. Identical verbs three corpora significantly differ in use

As can be seen in Figure 26, there is not a consistency among three corpora with respect the frequency of six items. For instance, *give* was used 1007 times in TACE, 716 times in NACE and 479 times in SACE. The following extracts are presented to exemplify it.

Example 77.

[Furthermore, they tend not to be distinguished from other types of group verb, and there are few attempts to **give** them a name.]

Extracted from <NACE-UG-20010-CW>

[After the video recording session, the participants were **given** appointments to watch and discuss the disruptive behaviours they experienced and their handling strategies.]

Extracted from <TACE-CU-2007-AK>

[It is my belief that as L2 writing teachers, we need to **give** students freedom to use all the resources they have during the L2 writing process including the students' L1s.

Extracted from <SACE-UI-2012-JMPR>

In the use of *learn*, three corpora significantly differ once again. It occurred 737 times in TACE, 317 times in NACE and 456 times in SACE. Thus, it was significantly

overused by TAEs and SAEs in compared to NAEs. The following statements were taken from three corpora.

Example 78.

[They were also aware of the fact that they **learned** something from their writings, which were shown in the following examples.]

Extracted from <TACE-CU-2009-SDU>

[The participants in the L2 group were native speakers of English, who had **learned** German after the age of 11.]

Extracted from <NACE-UE-2009-FW>

[Age group 2 includes people who **learned** to read and write only in Spanish and attended school when instruction was mandatorily in Spanish.]

Extracted from <SACE-PSU-2009-APP>

It might be an interesting finding that SACE and NACE display similarities against TACE in the use of certain verbs such as *consider, explain, follow, include, mean, provide, reveal, understand, use* and *write*, which can be seen in Figure 27.

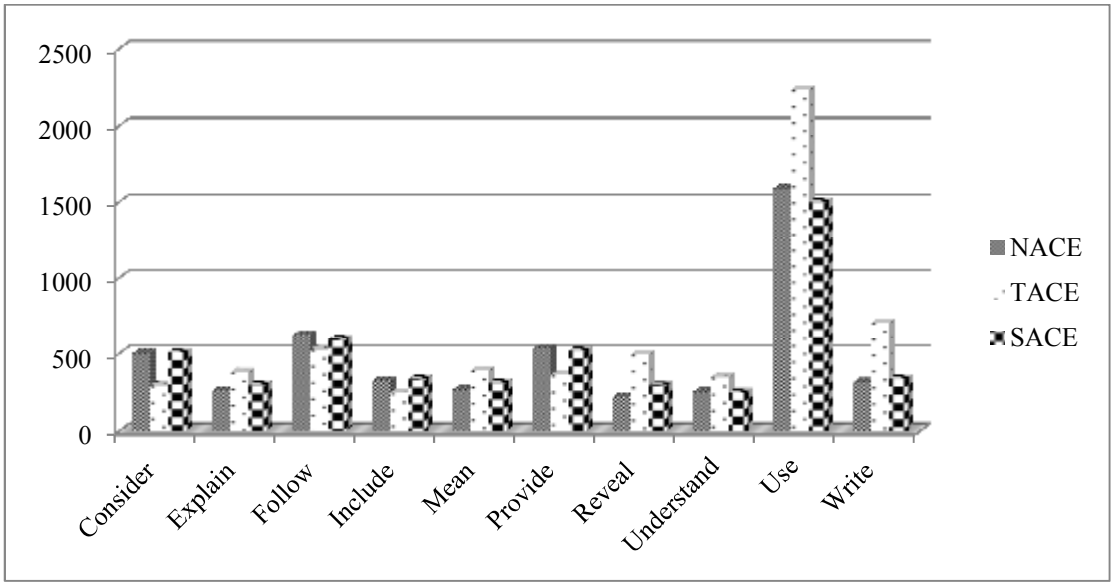


Figure 27. Identical verbs in SACE and NACE against TACE

As illustrated in Figure 27, there is not a significant difference between SACE and NACE in the use of 10 particular verbs as the one between TACE and NACE. For example, *consider* was used approximately with the same frequency in SACE and NACE while it was employed significantly underused in TACE occurring less than 300 times in it. It is exemplified in the following statements.

Example 79.

[In this section, the observations made in the analytic chapters will be **considered** in terms of orientations to linguistic identities.]

Extracted from <NACE-NU-2011-AB>

[**Considering** the detailed descriptions of the cases, I presume that EFL teachers and managers will be in a position to examine the practical implications of these findings by themselves.

Extracted from <SACE-UW-2010-HSS>

[The inter-rate reliability, which is referred to eliminate the inconsistency among the scores from two or more scorers, needed to be **considered**.]

Extracted from <TACE-GU-2008-DE>

Include is another verb which seems to have been distributed relatively at similar frequencies in SACE and NACE when compared to TACE. Namely, it appeared 326 times in NACE and 334 times in SACE whereas it was found 345 times in TACE. The following are the expressions taken from three corpora.

Example 80.

[In this graph I **include** all the selections, so as to get a sense also of the range of selections and their proportions.]

Extracted from <NACE-UE-2008-BAS>

[Music should also be **included** in other subjects and other levels since, if there is an improvement in primary, students of secondary and university could also benefit from them.]

Extracted from <SACE-UH-2011-CMTF>

[Furthermore, since the omission in Turkish Broca's aphasic patients **includes** only the inflections but not the verb or noun stem, no telegraphic speech is formed as it is the case in English.]

Extracted from <TACE-HU-2008-AAB>

One of the most prominent items in this category might be *use*, which was distinctively overused in TACE against NACE. That is, it occurred 2235 times in TACE while it was employed 1587 and 1501 times in NACE and SACE, respectively. The following sentences including it were drawn from each corpus.

Example 81.

[While the participants **used** plural-s correctly in 45 sentences, they only did seven errors among the total of the sentences produced.]

Extracted from <TACE-CU-2006-DE>

[The questionnaire asked respondents to identify (from four options) how often they **used** their languages other than English.]

Extracted from <NACE-NU-2006-CS>

[Finally, in section 3.4.2, I **used** binding tests to illustrate that contrastive FS does not involve syntactic movement.]

Extracted from <SACE-IU-2009-DCMV>

Another interesting finding might be that TACE and NACE display similarities against SACE concerning the use of such verbs as *find, indicate, present* and *seem*, which is illustrated in Figure 28.

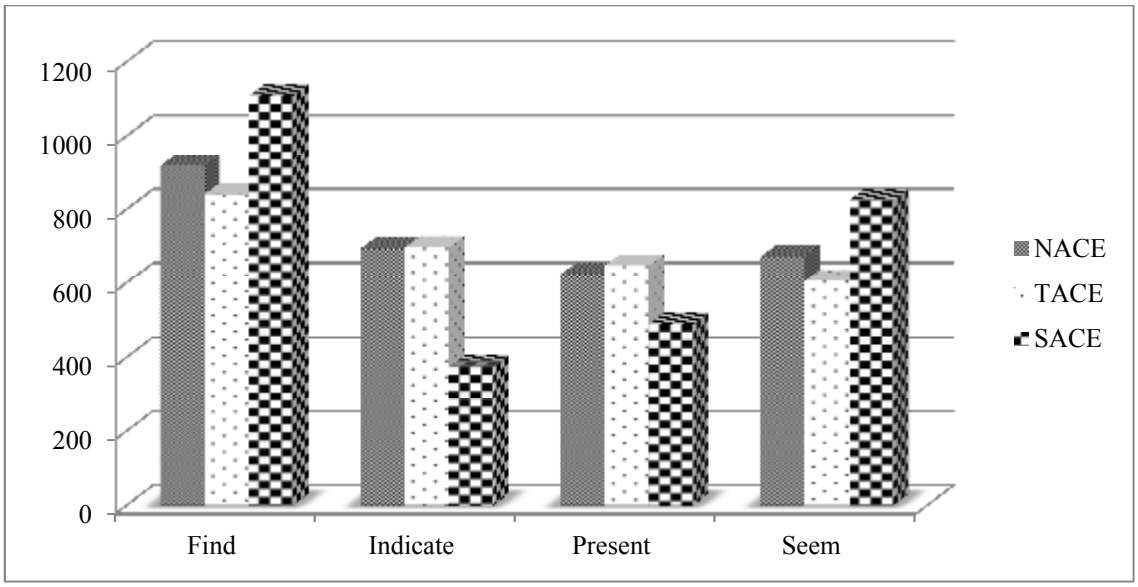


Figure 28. Identical verbs in TACE and NACE against SACE

As seen in Figure 28, SACE differs from TACE and NACE regarding the use of find *indicate*, which was found 840 and 917 times in TACE and NACE. It is observed to have been significantly overused in SACE with a frequency of 1108. Statements extracted from three corpora are given below to exemplify the item in concern.

Example 82.
[Furthermore, we **found** that explicit instruction leads to gains in comprehension tasks but not in production tasks.]

Extracted from <TACE-CU-2011-EA>

[Significant differences for these tenses in perfective contexts are also **found** for low group learners.]

Extracted from <NACE-NU-2011-KM>

[We see that this tendency is **found** in formal written English as illustrated by the English section of our corpus.]

Extracted from <SACE-USC-2009-NVC>

Indicate is another verb, which was found 374 times in SACE, and approximately equal times in TACE and NACE. It is exemplified in the following statements taken from three corpora.

Example 83.

[The outcomes **indicate** that the course needs improvement in terms of listening skill.]

Extracted from <TACE-GU-2010-ES>

[The majority of LEP students must take the GQE, even if test preparations **indicate** the likelihood that these students may not pass the exam.]

Extracted from <NACE-UI-2010-AH>

[These numbers strongly **indicate** that none of the L2 groups is excluded from attaining native-like overall performance.]

Extracted from <SACE-UI-2011-GSCD>

Present was observed to be employed 647 and 623 times in TACE and NACE, respectively while it occurred less than 500 times in SACE. The following are the statements taken from three corpora.

Example 84.

[For a comparative analysis, we will **present** disruptive behaviours examples of both transition and induction phases.]

Extracted from <TACE-CU-2007-AK>

[The position of these genres in terms of the dimensions written/speech-based and formal/informal is **presented** in table 9-3.]

Extracted from <NACE-UG-20010-CW>

[In this thesis we have **presented** two approaches to deal with multilingualism at two different stages of the ontology development process.]

Extracted from <SACE-UPM-2011-EMP>

As for *seem*, it was found overused in SACE with a frequency of 825 whereas it appeared 605 and 825 times in TACE and NACE, respectively. It is exemplified in the following expressions drawn from the corpora in question.

Example 85.

[It **seems** that they are more relaxed and at ease in their writing paragraphs since they have learnt the main paragraph patterns and written six different paragraphs.]

Extracted from <TACE-HU-2009-AK>

[These meanings thus **seem** to be a semantic feature of the register as part of the point of departure for meanings both at clause rank and in discourse structure.]

Extracted from <NACE-UM-2009-CES>

[Again, the Spanish speakers not only **seem** to disagree more than the other two groups, but they also do so in a direct fashion more frequently than the rest.]

Extracted from <SACE-UG-2008-NILS>

Finally, the verbs used at similar frequencies in SACE and TACE against NACE could be discussed. Figure 29 depicts five verbs of such kind.

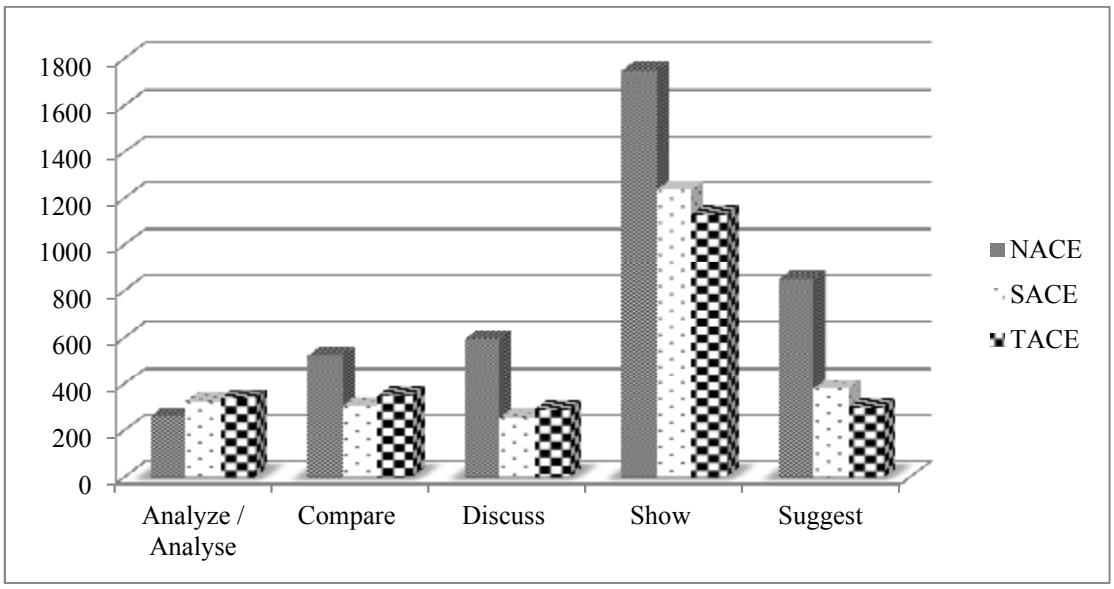


Figure 29. Identical verbs in TACE and NACE against SACE

It is seen that *analyse/ analyze* was overused in TACE and SACE in compared to NACE, whereby it occurred 261 times. It is illustrated in the following extracts taken from three corpora.

Example 86.

[As the first index of the lexical competence, the participants' sizes of vocabulary knowledge were **analysed** in terms of both overall knowledge and frequency bands.]

Extracted from <TACE-AU-2012-IY>

[The relational implications of these variations will be **analysed** in detail in the next chapter and their transactional use in the next three sections of this one.]

Extracted from <NACE-UE-2006-HH>

[Once the data were coded, processed and **analysed**, the third and fourth phases of the present study began.]

Extracted from <SACE-USC-2009-RFC>

Compare and *discuss* were significantly underused by TAEs and SAEs against NAEs. Namely, it was used 523 times in NACE while it appeared 349 and 304 times in TACE and SACE, respectively. *Discuss*, on the other hand, was employed in NACE over its total frequencies in SACE and TACE across which it was found 290 and 256 times, respectively. NACE remarkably differs from TACE and SACE in the use of *show* and *suggest*. *Show*, for instance, is observed to have been underused by TAEs and SAEs in compared to NAEs. Below are the extracts drawn from three corpora.

Example 87.

[As **shown** in Table 17, the native English speakers' English ARSPAN scores are significantly correlated with their AOSPAN scores.]

Extracted from <TACE-BU-2010-FC>

[The clear articulation of vision, policy and willingness to integrate learning produced by internal groups is currently not **shown** as present.]

Extracted from <NACE-UL-2011-CWM>

[In this chapter, I have **shown** that pre- and postverbal clitic alternations arise uniformly in the matrix context in Austrian, Galician and European Portuguese.]

Extracted from <SACE-GU-2009-FJFR>

Suggest was seen 300 and 381 times in TACE and SACE whereas it was found approximately 850 times in NACE. It is exemplified in the expressions below.

Example 88.
[Among the 46 main topics, the most favourite ones (above 200 pts) are **suggested** for the syllabus.]

Extracted from <TACE-AU-2008-GE>

[This is not a shortcoming of the experimental design, but it at least gave more reason to replicate the experiment, incorporating improvements as **suggested** above.]

Extracted from <NACE-UE-2010-GR>

[This confirms what previous studies have **suggested**, that the L2 writing process is a bilingual event.]

Extracted from <SACE-UI-2012-JMPR>

4.2. Discussion

4.2.1. Epistemic Stance Devices in Three Corpora

A total number of 113 ESDs, identified by Biber (2006), were searched across three corpora including dissertations of native and non-native academic authors of English. Each type was identified and counted via WordSmith Tools (Version 6.0) to reveal its proportion in the corpora at stake. It is noteworthy that 13 ESDs were not found in any corpora (See Figure 11). In addition, three corpora differ in the proportion of ESD types across them; namely, 97 types were found in NACE while 90 and 86 types were seen in SACE and TACE, respectively. Roughly expressing, ESDs were mostly found in NACE with a frequency of 4719 while they were seen 4455 and 4333 times in TACE and SACE, respectively. In other words, they were underused by TAEs and SAEs against NAEs. This finding contradicts with what was reported by Hyland and Milton (1997) who investigated epistemic modals and epistemic adverbs across exam scripts of native university students in UK and those of non-native students with higher and lower proficiency levels of English in Hong Kong, China. They revealed that both groups displayed similar tendencies in the way they deliver their assertions;

namely, they used epistemic items in similar frequencies whereas our study indicated the native and non-native groups significantly differ from each other in the use of ESDs. ESD types were analysed across three corpora within the categories of *epistemic adjectives, epistemic adverbs, epistemic nouns* and *epistemic verbs*. The results do not allow us to propose a general tendency in accordance with the categories of frequency and proportion. Namely, epistemic adverbs and epistemic nouns occurred in SACE with the highest frequency whereas epistemic verbs were mostly preferred in TACE. In other words, epistemic adverbs and epistemic nouns were overused by SAEs and epistemic verbs were overused by TAEs against NAEs. Epistemic adjectives, on the other hand, were used in higher proportion in NACE than both non-native corpora indicating that they were underused by TAEs and SAEs in comparison to NAEs. The latter finding is in parallel with Gabrielatos and McEnery's (2005) study which showed that epistemic adjectives were used more frequently in MA theses written by native speakers of English than those by non-native speakers of English attending Lancaster University, UK. However, the finding related to the overuse of epistemic adverbs by SAEs against NAEs in our study contradicts with the finding that they were underused in MA theses of non-native students in Gabrielatos and McEnery's (2005) study. Considering TACE into account, the finding in question seems to be confirmed by the finding of our study demonstrating epistemic adverbs were underused by TAEs in their doctoral dissertations. However, it is disapproving the finding of the study carried out by Can (2012), which indicated Turkish university students use stance adverbs more frequently than native speaker university students despite using them with less variety.

Subsequently, values of overuse and underuse were counted via LL calculation to investigate whether they are statistically significant or not. The underuse in SACE and TACE against NACE with respect to the overall ESDs was approved by LL results with values of -9.07 and -14.66, respectively. As for epistemic categories, epistemic adverbs and epistemic adjectives in TACE were of low frequency when compared to those in NACE, which was confirmed with -160.26 and -139.71 LL values. On the other hand, an overuse was observed in TACE against NACE in the use of epistemic verbs and epistemic nouns, which was once again approved by LL results (Epistemic Verbs: +56.55; Epistemic Nouns: +6.05). The results also indicate that SAEs underused epistemic adjectives and epistemic verbs and overused epistemic adverbs and epistemic nouns against NAEs. Likewise, these categorical frequencies were confirmed by LL

results (Epistemic Adjectives: -207.19; Epistemic Adverbs: +12.55; Epistemic Nouns: +27.10; Epistemic Verbs: -24.50).

As a third step, ESDs were examined in two groups as *certainty* and *likelihood* devices in order to evaluate the native and non-native groups in accordance with the choice of these items. The findings have revealed that certainty devices outnumber likelihood devices in all corpora. For instance, certainty devices were mostly used in TACE with a frequency of 2829 while they appeared 2611 and 2389 times in SACE and NACE, respectively. In return, likelihood devices were found most frequently in NACE appearing 2330 times in it. The LL test administered to three corpora confirmed the related results. This finding is in line with what Biber (2006) previously found. It partially endorses with the finding of Kotrc's (2012) study, in which stance adverbials in the language of press were analysed across randomly selected articles published in the newspapers in UK. He concluded that certainty adverbials were preferred significantly more often than likelihood adverbials in the articles published during the General Election 2010 in the country. Our finding also confirms the results of Ventola and Maureanen's (1990) study, which indicated that safe expressions of epistemic possibility were underused by Finnish-speaking learners of English against the native speakers, and those of Hyland and Milton's (1997), which indicated students with Chinese L1 are more assertive and committed with their statements than native speakers of English. Finally, it is disapproving with the finding of the study conducted by McEnery and Kiffle (2002), who revealed epistemic modals in weak forms occurred more frequently in Eritrean secondary school students' written productions in English than those of native students.

TACE and NACE seem identical in the most frequently used certainty categories. That is, certainty verbs were found with the highest frequency across two corpora occurring 1700 in TACE and 1095 times in NACE, conforming to the findings of Biber's study (2006) which displayed stance adjectives and stance nouns were less frequented than stance verbs in both spoken and written registers. Interestingly, the mostly preferred category in SACE is that of certainty adverbs with a frequency of 1143. Another finding of the analysis is that all groups seem identical in the preferences for the rest of the categories. Namely, the third mostly used and least used certainty devices in three corpora are epistemic nouns and epistemic adjectives.

Three corpora also display similarities in the use of likelihood categories, which are likelihood adjectives, likelihood adverbs, likelihood nouns and likelihood verbs. The

results have indicated that three corpora are identical in their choice of the most and second most frequently used likelihood categories. Thus, likelihood verbs are the most frequently employed category and likelihood adverbs are the second most frequented devices in three corpora. As for the third mostly used likelihood categories, native corpus differs from the non-native corpora in that likelihood adjectives fulfil this position in it whereas it is occupied by likelihood nouns in both TACE and SACE, which entails that three corpora also differ in the use of least preferred likelihood category. Namely, this category is constituted by likelihood adjectives in the non-native corpora and by likelihood nouns in the native corpus. It is noteworthy that the Log Likelihood Test results have approved all the differences between TACE and NACE as well as those found between SACE and NACE.

Finally, individual ESDs were surveyed via LL analysis in order to test the significance of frequently used ones in TACE and SACE against NACE. Seeing that LL values were consistent with the results of the previous analyses, we decided to make a further analysis to reveal whether it is the case through corpora concerning the use of main lexical categories (*adverbs, verbs, adjectives* and *nouns*). Each category was identified and counted via Wmatrix, a computer software package commonly employed in the field of corpus linguistics, and a log likelihood test was administered to compare the related results. The following section summarizes the results obtained from the analysis in concern.

4.2.2. Lexical Categories in Three Corpora

Three corpora were investigated via two software package programmes, Wmatwix and Log-Likelihood Calculator, with respect to four lexical categories as *adjectives, adverbs, nouns* and *verbs*. The overall results have shown that approximately half of each corpus consists of items belonging to these categories. Not surprisingly, nouns reveal to be most frequented lexical category in all corpora (NACE: 25.86%; TACE: 23.56%; SACE: 19.97%), which is approving Biber (2006) who found that nouns were used in written registers much more extensively than any other word class. It is also not surprising that adverbs occurred with lowest frequency in three corpora (NACE: 4.61%; TACE: 3.54%; SACE: 5.73%), which is again in consistent with Biber (2006) who indicated adverbs were employed in written registers much less frequently than adjectives, nouns and verbs. On the other hand, three corpora significantly differ

from each other in the use of adjectives and verbs. For instance, in NACE, adjectives were relatively more frequented than verbs (Adjectives: 10.54%; Verbs: 8.71%), which rather confirms Biber (2006) who reported verbs were used slightly more often than adjectives in written registers. The opposite is the case in TACE; that is, verbs were employed much more frequently than adjectives (Adjectives: 5.10%; Verbs: 8.27%). As for SACE, verbs were slightly less frequented than adjectives (Adjectives: 9.70%; Verbs: 9.12%).

Subsequently, each corpus was analysed to calculate percentages of lexical categories across it. In line with the above-mentioned results, it has been revealed that adverbs occurred with lowest frequency and nouns with the highest frequency in all corpora. However, it is interesting to see that adjectives constituted approximately 22% of the lexical items in SACE while they were slightly less frequented in NACE, whereby nouns appeared as the largest lexical category when compared to TACE and SACE. It is also noteworthy that nouns constituted approximately 60% of the lexical items in TACE.

Another finding is that adverbs constituted nearly 13% of the lexical items in SACE, 9.27% of those in NACE and 8.77% of those in TACE.

Adjectives seem to be employed less frequently than verbs in NACE (Adjectives: 10.54%; Verbs: 8.71%). Another finding related to verbs is that they had relatively a higher percentage than adjectives within lexical tokens in TACE (Adjectives: 12.62%; Verbs: 20.28%). It is also observed that they had similar proportions over the lexical categories in SACE (Adjectives: 21.79%; Verbs: 20.49%).

Finally, native and non-native corpora were compared with respect to the frequency ratio of each lexical category for 10.000 in order to find out whether there is an overuse or underuse in TACE and SACE against NACE and a Log-Likelihood Test was administered to see whether the obtained values are statistically significant or not. First of all, adjectives seem to be generally underused in both non-native corpora against the native corpus; however, the difference between SACE and NACE was not so significant as the one between TACE and NACE. Namely, in accordance with the test results, adjectives were underused in TACE with -13040.32 and in SACE with -234.07 LL values. As for adverbs, non-native corpora significantly differ from each other. That is, they were overused in SACE appearing 573.27 times per 10.000 words in it whereas they were underused in TACE occurring 354.24 times in every 10.000 words in it, which was confirmed with +818.01 and -943.81 LL values, respectively. Nouns,

on the other hand, constituted another lexical category that was considerably overused in both non-native corpora against the native corpus. For instance, they occurred approximately 590 and 230 fewer times per 10.000 words in SACE and TACE, respectively than NACE, which were approved with -5093.28 and -719.73 LL values, respectively. Lastly, it is revealed that verbs were significantly underused in TACE and overused in SACE against NACE. Results of the test have confirmed the findings in concern, with -108.01 and +64.10 LL values.

In this section, main findings of our study were summarized and related to findings of previous studies conducted with similar foci to its. Possible reasons for the findings in concern as well as their implications to language instruction will be discussed in the Conclusion section.

CHAPTER V

CONCLUSION

5.0. Introduction

In this study, Epistemic Stance Devices in the doctoral dissertations produced by NAEs (Native Academic Authors of English), TAEs (Turkish-Speaking Academic Authors of English), and SAEs (Academic Authors of English) were investigated through quantitative analysis. Three corpora were compared in terms of the results obtained from the analyses. The study, overall, revealed that 13 out of 117 types were not found in any corpora. In this section, the research questions are revisited and implications of our findings to language teaching are presented. Finally, suggestions are provided for further research.

5.1. Evaluation of Research Questions

RQ 1: Do native academic authors of English and Turkish-speaking academic authors of English significantly differ in the use of English ESDs?

The study shows that TAEs significantly differ from the native academic authors in the use of ESDs. First of all, 86 out of 117 ESD types were found in their doctoral dissertations while 97 types occurred in those of native authors. In line with this finding, the study also shows that ESDs were more frequented in the native corpus than the Turkish corpus, which means they were underused by TAEs in comparison to NAEs. As for the categories, epistemic adjectives and epistemic adverbs were significantly underused in TACE whereas epistemic nouns and epistemic verbs were considerably overused in it against NACE. Finally, it is observed that TAEs display a stronger tendency to use certainty devices than NAEs, who used likelihood devices much more frequently than TAEs, which leads us to the conclusion that the former are more confident with their predictions and the latter are more tentative in their academic writing.

RQ 2. Do native academic authors of English and Spanish-speaking academic authors of English significantly differ in the use of English ESDs?

The study indicates that 90 ESD types were found in the Spanish corpus and that the devices in question were significantly underused by SAEs against NAEs. Concerning the epistemic categories, it has been found that epistemic adjectives and epistemic verbs were significantly underused in SACE while epistemic nouns and epistemic adverbs were overused in it against NACE. The most striking difference between two corpora was found in the category of epistemic adjectives. Finally, as in the case of TAEs, it is observed that SAEs employed certainty devices more frequently and likelihood devices less frequently than NAEs, which once again means that the non-native group prefer confident predictions while the native group are more cautious in their academic writing.

RQ 3. Do Turkish-speaking academic authors of English and Spanish-speaking academic authors of English significantly differ in the use of ESDs?

The Spanish corpus revealed richer than the Turkish corpus in terms of ESD types. However, it is observed that the devices in concern were preferred less frequently by SAEs than TAEs. Regarding epistemic categories, it is seen that epistemic nouns and epistemic adverbs were more frequented in the doctoral dissertations of SAEs whereas epistemic adjectives and epistemic verbs were employed more often in those of TAEs. Lastly, it has been found that SAEs employed certainty devices less frequently and likelihood devices more frequently than TAEs, which leads to the conclusion that the Turkish-speaking academic writers have a predisposition to make predictions with relatively more confidence and their Spanish colleagues are more cautious in their writing.

All in all, it is seen that certainty devices were preferred more frequently than likelihood devices in the doctoral dissertations of both native and non-native speakers of English. However, the difference observed between these categories in the non-native corpora were measured much higher than the one in the native corpus. Considering that the non-native corpora included the doctoral dissertations of advanced learners of English, it would not be wise to claim this result has to do with proficiency level of their writers. As proposed in previous studies that indicated similar findings (e.g. Kaplan,

1966; Hinds, 1987; Matalene, 1985; Hyland, 2002; McCrostie, 2008; Yue, 2009 and Nasiri, 2012), it, rather, might be contributed to the cultural differences or thought patterns between the community they were born and raised and the one of which spoken language they have learned as a second/ foreign language. Yet, in order to come up with such a reason, we need to know whether the writers whose dissertations were examined in this study displayed similar behaviours while writing in their L1, as well. If so, the reason might be considered culture-related or as some factors concerning thought patterns of their L1 community. If not, some other reasons should be sought to account for the case such as instruction of academic writing. Nga (2009), who reported English-major students in Vietnam showed poor performance in their academic writing, claims this might be stemmed from the fact that those students are not introduced courses focusing on academic writing or academic writing style in a second language earlier than their seventh semester at university or that academic writing is not properly emphasised during the courses in concern. Considering that this particular finding of our study might also be caused by similar factors, we made a small-scale research on the official websites of the universities to which most of the dissertations examined in this study were submitted in order to find out whether Academic Writing Course is offered in their degree programmes. Furthermore, we tried to elicit this information via email when the programmes in concern are not available online. In consequence, to the best of our knowledge, Academic Reading and/or Academic Writing courses are not included in the undergraduate programmes in most of the universities in Turkey, Spain and the UK. They appear only in the graduate programmes of the universities in concern and it is noteworthy that they are rarely seen in the category of compulsory courses, which means some of the dissertations investigated in our study might have been written by an "academic" author who had never taken Academic Writing course during their higher education.

5.2. Implications of the Study

The present study has shown that TAEs and SAEs incline to use certainty devices while NAEs mostly prefer likelihood ones, entailing that the non-native groups are more confident with their predictions whereas the native group are more cautious in their academic writing. In other words, the native group is more tentative in expressing themselves when compared to the non-native groups. Baring in mind that academic

writing requires likelihood expressions rather than certainty ones, we could conclude that the non-native speakers have not adopted this attitude so satisfactorily as the native speakers of English. As in the case observed by Nga (2009), this situation might be resulted either from that academic writing is not properly emphasised during higher education or that it is not officially introduced at earlier levels of higher education in Spain and Turkey. It should be included in the higher education curricula as early as possible so that students could have the opportunity to develop their academic writing skills extensively. Alternatively, it might be offered that Academic Writing courses should become a compulsory part of the graduate/ postgraduate programmes; namely, students attending the programmes in question might be required to take these courses at least one semester before they commence writing their theses. This, of course, would not be sufficient for them to overcome all problems they are likely to face in their attempt to construct written productions in accordance with the principles of academic language; consequently, it is believed that the instructors assigned with these courses play the leading role in the process. In other words, no matter how we are aware of the fact graduate students are expected to take the responsibility of their own writing, these instructors might be provided some suggestions so as to help their students improve their academic writing. For instance, Carlson (1984) has revealed that in postgraduate programmes, engineers are more expected to describe charts whereas students of business studies are more required to compare ideas and take a position in their written assignments, which means the latter are more supposed to adopt a certain stance while conveying their ideas. At this point, Adel (2010) posits instructors should be both descriptive linguists and language policy makers. Namely, as suggested by Nga (2009), they should teach the essential features of academic writing, which are *complexity, formality, objectivity, explicitness, hedging* and *responsibility*, and set the acceptable writing behaviours –especially the ones concerning *hedging* in academic writing. In this case, for instance, graduate students studying social sciences should be made aware of the principle that academic writing is expected to include likelihood expressions rather than certainty ones in "taking a position". In other words, they might be advised to be cautious in delivering their opinions and interpreting the results of their works. Namely, they might be suggested to avoid using expressions like *X is the result of Y* or *Y caused X* and to prefer *X might have stemmed from Y* or *Y might have triggered X,* respectively.

Spack (1988) argues that graduate students, especially non-native graduate students, can learn academic writing by observing the process during which academic

texts are produced by professional writers or by attending programmes designed with the aim of teaching them how to write a research paper. Adel (2010) maintains that instructors should use complementary materials while teaching academic writing to these students. Therefore, a second suggestion might be that instructors teaching academic writing should employ corpora including academic works written by professional native and non-native authors of English. Adel (2010) also contends that non-native speakers lack native-speaker intuitions and have a predisposition to transfer writing patterns in their L1 to a rhetorically different L2 in an appropriate manner. For this reason, she believes corpus-based courses on academic writing are strongly needed for non-native graduate students. Using BAWE (British Academic Corpus of Written English) could be useful in this concern as it consists of a large number of academic works produced by native speakers of English. The instructor might choose an article in BAWE and analyse it with students in accordance with all or some selected principles of academic writing. In our case, it could be analysed in terms of certainty and likelihood expressions in order to comment upon the stance adopted by its author.

Adel (2010) reports highly positive results on a small-scale experiment she conducted with her non-native graduate students with high proficiency level of English. During a writing class, she asked those students to examine MICUSP, a corpus consisting of highly proficient student writing, via WS Tools (Scott, 2004) to find out what academic writers use when they (a) give an example, (b) refer to other texts or researchers, (c) introduce the topic and (d) start the Conclusion section. She observed that students were enthusiastic about inductive learning about the use of computer and that the corpus session was nicely supported by instruction from the regular writing instructors, during which note-taking on hand-outs was encouraged. In parallel with this, students might be assigned with analysing some articles in BAWE with regard to certain features of academic writing. In doing so, they are likely to become more conscious about the process in concern and to improve their own writing.

Hyland and Milton (1997) investigated the degree of caution and certainty in EAP writers' works in a corpus of one million words including 1700 UK and Hong Kong students' exam scripts. They also wanted to discover the ways students attending the University of Hong Kong present assertions in their writing. They found that confident predictions were more favoured by L2 writers while tentative expressions were preferred by native writers more frequently. The study also shows that non-native writers with higher proficiency of English approximate native speaker usage more

closely than those with lower proficiency of the same language. So, thirdly, graduate students might be asked to examine the essays they wrote prior to taking academic writing course or their own exam papers and to report on them taking into consideration what they have learned about author stance or other concerns during the course. In this case, they might also be asked to compare their papers with those produced by native and/or other non-native graduate students in terms of certainty and likelihood items used across the papers in concern and to write a report on the similarities and/ or differences between the papers at stake. In addition, they might be asked to revise their essays accordingly. Likewise, this process might be changed into a pair or group work activity. That is, students might be asked to form groups and each group might be assigned with an essay or exam paper produced by students. Then, the groups might be expected to analyse and revise the papers by reflecting their knowledge about author stance or other concerns of academic writing.

5.3. Suggestions for Further Research

This particular study is limited to the investigation of doctoral dissertations written by Turkish-speaking, Spanish-speaking and native academic authors of English between 2005 and 2012. It is also limited to the field of English Language Teaching, English Language and Literature, Applied Linguistics and Modern Languages. Therefore, its results cannot be generalized to any other settings. So, further research might involve building and investigating much broader corpora which include works of Turkish-speaking, Spanish-speaking and native academic authors of English produced prior to 2005 within various fields. Alternatively, similar sets of corpora might be built with the collection of dissertations produced by non-native academic authors of English from L1 backgrounds other than Turkish and Spanish (e.g. Chinese, German, Japanese, and etc.), and analysed with respect to author stance. In addition, author stance could be investigated concerning the use of certainty and likelihood expressions through spoken productions of academics in such events as conferences or symposiums. Finally, the reasons why certainty items are used more frequently than likelihood items in academic writing might be explored. Namely, the reasons why certainty items were significantly overused in the non-native corpora against the native corpus analysed in this study might be searched. More specifically, counterpart sets of two corpora that include doctoral dissertations written in native languages of the writers (Turkish and Spanish, in

our case) within similar subject areas (e.g., Spanish / Turkish Language and Literature, Spanish / Turkish Language Teaching, Applied Linguistics in Spanish / Turkish, and etc.) could be analysed in order to see whether the writers' L1 or culture or thought patterns of the society they were raised triggered this situation.

REFERENCES

Academic Writing (2013). University of Technology Sydney. Retrieved from http://www.lib.uts.edu.au/help/study-skills/writing-reading-speaking

Ädel, A. (2010). Using corpora to teach academic writing: Challenges for the direct approach. In M. C. Compoy-Cubillo, B. Belles-Fortuno & M. L. Gea-Valor (Eds.), *Corpus based approach to English language teaching* (pp. 39-54).

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburg: Edinburg.

Bawarshi, A. S. & Reiff, M. J. (2010). Genre in linguistic traditions: English for specific purposes. In A. S. Bawarshi and M. J. Reiff (Eds.), *Genre: An introduction to history, theory, research, and pedagogy* (pp. 41-56).

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

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

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

Biber D., Conrad, S., Reppen, R., Byrd, P. & Helt, M. (2002). Speaking and writing in the university: A multi-dimensional comparison. *TESOL Quarterly,* 36, 9-48.

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

Biber, D & Gray, B. (2010). Challenging stereotypes about academic writing: complexity, elaboration, explicitness. *Journal of English for Academic Purposes,* 9, 2-20.

Brinton, L. J. (2000). *The structure of modern English: A linguistic introduction*. John Benjamins.

Burke, S. B. (2010). *The construction of writer identity in the academic writing of Korean ESL students: A qualitative study of six Korean students in the U.S.* Doctoral Dissertation. Indiana University of Pennsylvania.

Campoy-Cubillo, M.C., Belles-Furtuno, B. & Gea-Valor, M.L. (2010). Introduction to corpus linguistics and ELT. *Corpus-based approaches to English language teaching*. Continuum (pp. 3-17).

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

Chalker, S. & Weiner, E. (1994). *Oxford dictionary of English grammar*. Oxford University Press.

Chang, Y. Y. & Swales, J. (1999). Informal Elements in English Academic Writing: Threats or Opportunities for Advanced Non-native Speakers? In C. Candlin & K. Hyland (Eds.), *Writing: Text, processes and practices*. London: Longman.

Clyne, M. (1994). *Intercultural communication at work: Cultural values in discourse.* Cambridge: Cambridge University Press.

Clyne, M. (1999). Writing, testing and culture. In B. Deocke (Ed.), *Responding to students' writing: Continuing conversations* (pp. 165-176). Norwood: AATE.

Conrad, S. & Biber, D. (2000). Adverbial making of stance in speech and writing. In S. Hunston and G. Tompson (Eds.), *Evaluation in Text*, 56-73. Oxford: Oxford University Press.

Crawford-Cemiciottoli, B. (2010). *A* corpus-informed approach to teaching lecture comprehension skills in English for business studies. *Corpus-based approaches to English language teaching*. Continuum (pp. 95-106).

Faerch, C., Haastrup, K. & Phillipson, R. (1984). Learner language and language learning. *Multilingual matters*, Clevedon.

Fairclough, N. (1992) *Discourse and social change*. Cambridge: Polity Press.

Conrad, S. & Biber, D. (2000). Adverbial Marking of Stance in Speech and Writing. In Hunston, S. & Thompson, G. (Eds.), *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second-language writing* (2nd Ed). London: Cambridge University Press.

Dontcheva-Navratilova, Olga. (2009). Evaluation in non-native writer's academic discourse: Stance devices. In *Research in English Language Teacher Education*. Brno: Masarykova Univerzita. Spisy Pedagogické fakulty Masarykovy univerzity. ISBN 978-80-210-4967-3. 33-42.

Flowerdew, L. (2004), The argument for using English specialized corpora. In Connor, U. & Upton T. A. (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 11–33). Amsterdam/Philadelphia: John Benjamins.

Gabrielatos & McEnery (2005). Epistemic modality in MA dissertations. In Fuertes Olivera, P.A. (Ed.), *Lengua y Sociedad: Investigaciones recientes en lingüística*

aplicada (pp. 311-331). Lingüística y Filología, 61. Valladolid: Universidad de Valladolid.

Gavioli, L. (2005). *Exploring corpora for ESP learning*. University of Modena and Reggio Emilia. John Benjamins Publishing Company.

Gersten, R., Baker, S.K., Shanahan, T., Linan-Thompson, S., Collins, P. & Scarcella, R. (2007). *Effective literacy and language instruction for English learners in the elementary grades, An IES practice guide*, DC: IES, Department of Education, Washington, 2007.

Gilquin, G. (2001). The Integrated Contrastive Model. Spicing up your data. *Languages in Contrast*, *3*(1), 95-123.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In: Aijmer, K., Altenberg, B./Johansson, M. (Eds.), *Languages in Contrast* (pp. 37-51). Lund: Lund University Press.

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

Granger, S. & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data driven learning. *Learner English on computer* (pp. 199-209). London and New York: Longman.

Gregg, J. (1983). *Contrastive Rhetoric: An xxploration of Chinese and American expository patterns*. New York City Technical College.

Hinds, J. (1987). Reader versus writer responsibility: A new typology. In T. J. Silva & P. K. Matsuda (Eds.), *Landmark Essays on ESL Writing* (pp. 63-85). Philadelphia: Lawrence Erlbaum Associates.

Hunston, S. & Thompson, G. (Eds.) (2000). *Evaluation in Text*. Oxford, UK: Oxford University Press.

Hyland, K. (1995). *The author in text: Hedging scientific writing*. Hong Kong Papers in Linguistics and Language Teaching, 18, 33-42.

Hyland , K. (1998). *Hedging in scientific research articles*. Amsterdam/ Philadelphia: John Benjamins.

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *TEXT 18*(3), p. 349-382.

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow: Pearson Education.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20, 207-226.

Hyland, K. (2002). Specificity revisited: How far should we go now? *English for Specific Purposes*, 21, 385-395.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics,* 34, 1091-1112.

Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. London: New York.

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

Hyland, K. & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6*(2), 185-205.

Hyland, K. & Tse, P. (2005). Hooking the reader: A corpus study of *evaluative that* in abstracts. *English for Specific Purposes,* 24, 123–139.

Irvin, L. L. (2010). What is "academic" writing? *Writing spaces: Readings on Writing*, Volume 1.

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

Johns, A. & Paz, D. (1997). Text analysis and pedagogical summaries: Revisiting Johns and Davies. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications* (pp. 33-49). English Language Programs, Washington, D.C.: United States Information Agency.

Johns, T. & Davies, F. (1983). Text as a vehicle for information: The classroom use of written texts in teaching reading in a foreign language. *Reading in a Foreign Language, 1*(1), 1-20.

Kafes, H. (2009). *Authorial stance in academic English: Native and non-native academic speaker writers' use of stance devices (modal verbs) in research articles*. PhD Thesis, Anadolu University.

Kaplan, R. (1997). Contrastive rhetoric. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications*. English Language Programs (pp. 18-32), Washington, D.C.: United States Information Agency.

Kaplan, R. (2001). Cultural thought patterns in inter-cultural education. In T. Silva & P. K. Matsuda. *Landmark Essays on ESL Writing* (pp. 11-25). Mahwah, NJ: Lawrance Erlbaum Associates.

Kramsch, C. (1997). Rhetorical models of understanding. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications*. English Language Programs (pp. 50-63), Washington, D.C.: United States Information Agency.

Krieger, D. (2003). Corpus linguistics: What it is and how it can be applied to teaching. *The Internet TESL Journal*, *9*(3). Retrieved 20 April 2007, from: http://iteslj.org/

Leki, L. & Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly,* 28, p. 81.

Maroko, G. M. (2013). Learning about author positioning in written academic discourse. *Argentinian Journal of Applied Linguistics, 1*(2), 47-60.

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.

MacIntyre, R. (2010). *How do they stand? Research into the representation of stance in academic writing and its implications for pedagogy*. Sophia University Faculty of Liberal Arts (p. 19-32).

McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal*, 32, 99. 97-114.

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

McEnery, T.& Kifle, N. A. (2002). Epistemic modality in argumentative essays of second-language writers. In Flowerdew, J. (Ed.), *Academic Discourse*, (pp. 182-195). London: Longman.

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

Nasiri, S. (2012). Academic writing: The role of culture, language and identity in writing for community. *International Journal of Learning and Development, 2*(3), 1-8.

Nga, N. T. H. (2009). Academic English at tertiary level: What, why and how. *VNU Journal of Science, Foreign Languages* 25, 112-117.

Ochs, E. (1990). Cultural universals in the acquisition of language: Keynote address. *Papers and reports on child language development*, Stanford University, 29, 1–19.

Odlin, T. (1989). *Language transfer*. Cambridge: Cambridge University Press.

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

Oshima, A. & Hague, A. (1994). *Writing academic English*. Addison-Wesley Publishing.

Pisanski-Peterlin, A. (2010). Hedging Devices in Slovene-English Translation: A corpus-Based Study. *Nordic Journal of English Studies 9*(2), 171-193.

Rayson, P. (2004). Keywords are not enough. *Invited talk for JAECS (Japan Association for English Corpus Studies)*, November , 2004, Chuo University, Tokyo, Japan.

Rayson, P. (2013). Vmatrix: *A Web-based corpus processing environment*. Lancaster University. Retrieved from: http://ucrel.lancs.ac.uk/vmatrix.

Rayson, P. & Garside, R. (2000). Comparing corpora using frequency profiling. Proceedings of the workshop on *Comparing Corpora*, 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000), October, 2000, 1-6, Hong Kong.

Römer, U. (2010). *Corpus based approaches to English language teaching*. MPG Books.

Salager-Meyer, F. (1997). I think that perhaps you should: A study of hedges in written scientific discourse. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications*. English Language Programs (pp. 18-32), Washington, D.C.: United States Information Agency.

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

Scott, M. (2012). *WordSmith Tools.* Version 6, Liverpool: Lexical Analysis Software.

Skelton, J. (1988). Comments in academic articles. In P. Grunwell (Ed.), *Applied Linguistics in Society: British Studies in Applied Linguistics* 3, (pp. 98-108). London: Centre for International Language.

Spack, R. (1988). Initiating ESL students into the academic discourse community: How far should we go? *TESOL Quarterly*, *22*(1), 29-52.

Stubbs, M. (2004). Language corpora. In A. Davies & C. Alder (Eds.), *The Handbook of Applied Linguistics,* (pp. 106-132). Blackwell Publishing.

Sultan, A.H.B.J. (2011). A Contrastive Study of metadiscourse in English and Arabic Linguistics Research Articles. *Acta Linguistica*, *5*(1).

Thaiss, C. & Zawacki, T. (2006). *Engaged writers & Dynamic disciplines: Research on the academic writing life*. Portsmouth: Boynton/Cook.

Thonney, T. (2011). Teaching the conventions of academic discourse. *Teaching English*

in the Two-Year College, 38(4), 347-362.

Uribe, D. (2008). Characteristics of academic English in the ESL classroom. *The Internet TESL Journal*, *14*(3). Retrieved from http://iteslj.org/

Ventola, E. & Maureanen, A. (1990). *Research and Writing in English*. Helsinki: Yliopistopaino.

Yue, Y. (2009). Rhetoric: A tale of two cultures. *The International Journal of Language, Society and Culture* 28, 92-99.

APPENDICES

APPENDIX A. DISSERTATIONS in NACE

Alcorn, R. J. *Pronouns, Prepositions and Probabilities: A Multivariate Study of Old English Word Order.* Unpublished doctoral dissertation. The University of Edinburgh.

Asbury, A. (2008). *The Morphosyntax of Case and Adpositions.* Unpublished doctoral dissertation. Universiteit Utrecht.

Barras, W.S. (2010). *The Sociophonology of Rhoticity and r-sandhi in East Lancashire English.* Unpublished doctoral dissertation. The University of Edinburgh.

Brandth, A. (2011). *The Maintenance of Mutual Understanding in Online Second Language Talk.* Unpublished doctoral dissertation. Newcastle University.

Bell, P. (2010). *Full Covariance Modelling for Speech Recognition.* Unpublished doctoral dissertation. The University of Edinburgh.

Brown, N. (2010). *Hollywood, the Family Audience and the Family Film, 1930-2010.* Unpublished doctoral dissertation. Newcastle University.

Collie, S. (2007). *English stress preservation and Stratal Optimality Theory*. Unpublished doctoral dissertation. The University of Edinburgh.

Cook-Leonard, A. *An Investigation into the Role of Implicit Knowledge in Adult Second Language Acquisition.* Unpublished doctoral dissertation. The University of Edinburgh.

Couper-Kenney, F. (2008). *Automatic determination of sub-word units for automatic speech recognition.* Unpublished doctoral dissertation. The University of Edinburgh.

Donneky, C. M. (2011). *Language Ideologies and Language Practices in France and Spain: The Case of Breton, Occitan, Catalan and Galician.* Unpublished doctoral dissertation. University of Newcastle-Upon-Tyne.

Evans, E.J. (2012). *Managing the Foreign Language Classroom: Reflections from the Pre-Service Field and Beyond.* Unpublished doctoral dissertation. University of Iowa.

Farrimond, K. (2011). *Beyond Backlash the Femme Fatale in Contemporary American Cinema.* Unpublished doctoral dissertation. Newcastle University.

Foreman, J. O. (2006). *The Morphosyntax of Subjects in Macuiltianguis Zapotec.*

Unpublished doctoral dissertation. University of California.

Gilkerson, J. S. (2006). *Acquiring English Particle Verbs: Age and Transfer Effects in Second Language Acquisition.* Unpublished doctoral dissertation. University of California.

Good, J. (2012). *The Human Presence in Robert Henryson's Fables and William Caxton's The History of Reynard the Fox.* Unpublished doctoral dissertation. The University of Glasgow.

Gressang, J. E. (2010). *A Frequency and Error Analysis of the Use of Determiners, theRelationships between Noun Phrases, and the Structure of Discourse in English Essays by Native English Writers and Native Chinese, Taiwanese, and Korean Learners of English as a Second Language.* Unpublished doctoral dissertation. University of Iowa.

Hanson, R. M. (2007). *Feedback in Intelligent Computer-Assisted Language Learning and Second Language Acquisition: A study of its effect on the acquisition of French past tense aspect using an Intelligent Language Tutoring System.* Unpublished doctoral dissertation. The University of Edinburgh.

Harman, R. (2008). *Systemic Functional Linguistics and the Teaching of Literature in Urban School Classrooms.* Unpublished doctoral dissertation. University of Massachusetts Amherst.

Hawkey, D. J. (2008). *Beyond the Individual in the Evolution of Language.* Unpublished doctoral dissertation. The University of Edinburgh.

Hetler, A. (2010). *Indiana Teachers' Perspectives on testing Accommodations for Limited English Proficient Students Taking the Graduation Qualifying Exam.* Unpublished doctoral dissertation. University of Iowa.

Hewitt, H. (2006). *Front Desk Talk: A Study of Interaction between Receptionists and Patients in General Practice Surgeries.* Unpublished doctoral dissertation. The University of Edinburgh.

Hurd, S. (2008). *Second Language Learning at a Distance: Metacognition, Affect, Learning Strategies and Learner Support in relation to the Development of Autonomy. Volume 1: Introduction to the Published Work.* Unpublished doctoral dissertation. Open University.

Jones, J. (2006). *Multiliteracies for Academic Purposes: A Metafunctional Explorationof Intersemiosis and Multimodality in University Textbook and*

Computer-Based Learning Resources in Science. Unpublished doctoral dissertation. The University of Sydney.

Kronenthal, M. (2007). *Policy, Planning and Perceptions in the European Union: A Comparative Perspective on Minority Language Vitality.* Unpublished doctoral dissertation. The University of Edinburgh.

Mangham, J. W. (2011). *A Combined Framework for Investigating Communities of Practice and the Function of the Learning Organization: A Case Study of an Industrial Training Unit in the United Arab Emirates.* Unpublished doctoral dissertation. University of Leicester.

Maxwell-Reid, C. (2011). *The effect of bilingual education on students" first languagewritten discourse: a contrastive Spanish-English study using systemic functional linguistics.* Unpublished doctoral dissertation. The University of Edinburgh.

McKean, C. (2009). *Investigating the Development of a Developmental Disorder: Mapping the Trajectory of Lexical Development in Specific Language Impairment.* Unpublished doctoral dissertation. Newcastle University.

McKenzie, R.M. (2006). *A Quantitative Study of the Attitudes of Japanese Learners towards Varieties of English Speech: Aspects of the Sociolinguistics of English in Japan.* Unpublished doctoral dissertation. The University of Edinburgh.

McManus, K. (2011). *The Development of Aspect in a Second Language.* Unpublished doctoral dissertation. Newcastle University.

Messenger, K. (2009). *Syntactic Priming and Children's Production and Representation of the Passive.* Unpublished doctoral dissertation. The University of Edinburgh.

Miller, S. (2012). *Shelley's Early Fiction in Relation to his Poetics and his Politics: An Assessment.* Unpublished doctoral dissertation. University of Glasgow.

Mroz, A. P. (2012). *Nature of L2 Negotiation and Co-Construction of Meaning in a Problem-Based Virtual Learning Environment: A Mixed Methods Study.* Unpublished doctoral dissertation. University of Iowa.

Nicholsan, H. B. M. (2009). *Disfluency In Dialogue: Attention, Structure And Function.* Unpublished doctoral dissertation. The University of Edinburgh.

O'Donnell, M. E. (2005). *Use of Textual Elaboration with Literary Texts in Intermediate Spanish.* Unpublished doctoral dissertation. University of Iowa.

Peterson, M. (2007). *An Investigation of Learner Interaction in a MOO-Based Virtual*

Environment. Unpublished doctoral dissertation. The University of Edinburgh.

Roberts, G. (2010). *Cooperation, Social Selection, and Language Change: An Experimental Investigation of Language Divergence.* Unpublished doctoral dissertation. The University of Edinburgh.

Scott, C. E. (2009). *Reporting Armistice: A Diachronic, Functional Perspective.* Unpublished doctoral dissertation. Macquarie University.

Scott-Phillips, T.C. (2008). *The Social Evolution of Pragmatic Behaviour. Unpublished* doctoral dissertation. The University of Edinburgh.

Smith, B. A. (2008). *Intonational Systems and Register: A Multidimensional Exploration*. Unpublished doctoral dissertation. Macquarie University.

Soule, D. P. J. (2006). *The Construction and Negotiation of Meaning in Scottish Political Discourse: A Case Study of the 2003 Scottish Parliament Elections.* Unpublished doctoral dissertation. The University of Glasgow.

Stegen, O. (2008). *In Quest of a Vernacular Writing Style for the Rangi of Tanzania: Assumptions, Processes, Challenges.* Unpublished doctoral dissertation. The University of Edinburgh.

Stow, C. (2006). *The Identification of Speech Disorders in Pakistani Heritage Children.*Unpublished doctoral dissertation. Newcastle University.

Sullivan, J. N. (2010). *Approaching Intonational Distance and Change.* Unpublished doctoral dissertation. The University of Edinburgh.

Wild, C. (2008). *Attitudes towards English Usage in the Late Modern Period: The Case of Phrasal Verbs.* Unpublished doctoral dissertation. The University of Glasgow.

Wilson, F. (2009). *Processing at the Syntax-Discourse Interface in Second Language Acquisition.* Unpublished doctoral dissertation. The University of Edinburgh.

APPENDIX B. DISSERTATIONS in TACE

Açan-Budak, A. (2008). *Agrammatism in Turkish Broca's Aphasic Patients: A Descriptive Case Study*. Unpublished doctoral dissertation. Hacettepe University: Ankara, Turkey.

Akçayoğlu, D. İ. (2011). *Exploring The Role of Strategies-Based Language Instruction in Teaching English to Young Gifted Learners*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Akman, D. B. (2012). *A Suggested Syllabus for the "Effective Communication Skills" Course for the Pre-service English Language Teachers at Gazi University*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Akpınar, K. D. (2009). *Using Theatre Extracts in Teaching Speech Acts to Turkish EFL Students with Special Reference to Apologies and Refusals*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Altay, İ. F. (2008).*A Suggested Syllabus for Advanced Writing Skills at English Language Teaching Departments*. Unpublished doctoral dissertation. Hacettepe University: Ankara, Turkey.

Altınkamış-Türkay, F. (2005). *Children's Early Lexicon in Terms of Noun/Verb Dominance*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Altunay, D. (2009*). Use of Connectives in Written Discourse: A Study at an ELT Department in Turkey*. Unpublished doctoral dissertation. Anadolu University: Eskişehir, Turkey.

Altunkol, E. (2011). *Acquisition of Causative / Inchoative Transitivity Alternations by Turkish Learners of L2 English*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Avsar, A. (2011). *Co-Construction of Learner And Teacher Identities in EFL Classroom Context*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Balçıkanlı, C. (2010). *The Effects of Social Networking on Pre-service English Teachers' Metacognitive Awareness and Teaching Practice*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Bardakçı, M. (2010*). The Impact of Raising Awareness about Reasoning Fallacies on the Development of Critical Reading*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Başaran, S. (2010). *Effects of Podcasts on Language Learning Beliefs and Self- Efficacy Perceptions of First-Year Turkish University Students*.Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Bilge, Z. (2008). *An Alternative Mode of Communication: Songs and Singing in Shakespeare's Tragedies.* Unpublished doctoral dissertation. Istanbul University: Istanbul, Turkey.

Boyno, M. (2011). *An Analysis of the Factors Influencing Learner Autonomy in the Turkish EFL Context.* Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Büyükkarcı, K. (2010). *The Effect of Formative Assessment on Learners' Test Anxiety and Assessment Preferences in EFL Context.* Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Büyüknisan, E. (2007). *Developmental And Transfer Errors of German Adult Learners of Turkish: A Case Study*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Çankaya, M. (2009). *Linguistic and Semiotic Representation of Gender in Turkish TV Commercials.* Unpublished doctoral dissertation. Hacettepe University: Ankara, Turkey.

Çele, F. (2010). *Processing Wh-Dependencies in L2 English: the Role of L1 and Working Memory Capacity.*Unpublished doctoral dissertation. Boğaziçi University: Istanbul, Turkey.

Demiröz, H. (2008). *Goal Orientation and L1 and L2 Reading Comprehension Strategy Use of Turkish ELT Students: A Qualitative Investigation.* Unpublished doctoral dissertation. Atatürk University: Erzurum, Turkey.

Ekşi, G. (2008). *A Suggested Syllabus for the ELT Preparatory Students at Gazi University in accordance with CEF B2 (Vantage) Level*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Erice, D. (2008). *The Impact of E-Portfolio On the Writing Skills of Foreign Language Learners Studying at Abant İzzet Baysal University Basic English Program*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Ertekin, D. (2006). *Acquisition of L2 English Inflectional System by Turkish Native Speakers: Rule-Governed Or Rote-Learned.* Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Genç, B. (2007). *An Analysis of Communication Strategies Employed By Turkish-Speakers of English*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Kafes, H. (2009). *Authorial Stance in Academic English: Native and Non-Native Academic Speaker Writers' Use of Stance Devices (Modal Verbs) in Research Articles.* Unpublished doctoral dissertation. Anadolu University: Eskişehir, Turkey.

Kahraman, A. (2009). *The Role of L1 Use in Improving Affective and Cognitive Factors in English Language Classrooms.* Unpublished doctoral dissertation. Hacettepe University: Ankara, Turkey.

Kıray, G. (2011). *Writing as a Literacy and Communication Skill in the Process of Learning and Teaching English as a Foreign Language at State Schools of Turkey (From Grade 4 To Grade 10).* Unpublished doctoral dissertation. Istanbul University: Istanbul, Turkey.

Kızıldağ, A. (2007). *A Longitudinal Study on English Language Teachers' Beliefs about Disruptive Behaviour in Classrooms: Case Studies From Practicum to the First-Year in the Profession*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Köse, N. (2006). *Effects of Portfolio Implementation and Assessment on Critical Reading and Learner Autonomy of ELT Students.*Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Merç, A. (2010). *Foreign Language Student Teacher Anxiety*. Unpublished doctoral dissertation. Anadolu University: Eskişehir, Turkey.

Oral, Y. (2009). *Classroom Power Relations in an English as a Foreign Language Setting from a Critical Pedagogical Perspective*. Unpublished doctoral dissertation. Istanbul University: Istanbul, Turkey.

Önalan, O. (2009). *A Comparative Study of Formal Complaints by Native Speakers and Turkish Learners of English.*Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Razı, S. (2010). *Effects of a Metacognitive Reading Program on the Reading Achievement and Metacognitive Strategies*.Unpublished doctoral dissertation. 9 Eylül University: İzmir, Turkey.

Solak, E. (2010).*A Suggested Syllabus for the Advanced Level English Course at Gendarmerie Schools Command in Accordance With NATO STANAG 6001 Level 3 Perspective*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Subaşı, G. (2010). *Why Do I Want to Become an English Teacher? Developing a Questionnaire*. Unpublished doctoral dissertation. Anadolu University: Eskişehir, Turkey.

Şanal, F. (2007). *A Learner Corpus Based Study on Second Language Lexicology of Turkish Students of English.*Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Şeker, M. (2007). *Exploring Effects of Collaborative Learning in Enhancing Teachers' Development in Teaching English to Young Learners*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Şimşek, H. (20079. *A Teacher Development Program for Young Learners of English: An Action Research*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Tokdemir-Demirel, E. (2009). *An Investigation of a Complementary Feedback Model for L2 Writing: Peer and Teacher Feedback versus Teacher Feedback*. Unpublished doctoral dissertation. METU: Ankara, Turkey.

Tuncer, H. (2012). *The Role of Speech and Thought Presentation in the Comprehension and Appreciation of English Prose Fiction*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Üğüten, S. D. (2009). *The Use of Writing Portfolio in Preparatory Writing Classes to Foster Learner Autonomy.* Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Ülkersoy, N. (2007). *Markedness Differential Hypothesis and the Phonological Errors of Turkish EFL Learners.* Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Ünaldı, İ. (2011). *A Comparative Investigation of Lexical Networks of Turkish Learners of English as A Foreign Language: A Corpus Based Study*. Unpublished doctoral dissertation. Çukurova University: Adana, Turkey.

Yangın, C. (2010). *Developing Prospective English Language Teachers' Comprehension of Texts with Humorous Elements*. Unpublished doctoral dissertation. Gazi University: Ankara, Turkey.

Yeşilyurt, S. (2008). *A Self-Determination Approach to Teaching Writing in Pre-Service EFL Teacher Education.* Unpublished doctoral dissertation. Atatürk University: Erzurum, Turkey.

Yetkin, N. (2009). *Analysing the Translatability in Subtitled Humour in the Turkish Cultural and Linguistic Context.*Unpublished doctoral dissertation. Hacettepe University: Ankara, Turkey.

Yüksel, İ. (2012). *Cross-Sectional Evaluation of Turkish ELT Majors' General and Academic Lexical Competence and Performance*. Unpublished doctoral dissertation. Anadolu University: Eskişehir, Turkey.

Zingir-Gülten, A. (2008). *Requesting in English: Interlanguage Pragmatics of Turkish Children*. Unpublished doctoral dissertation. Anadolu University: Eskişehir, Turkey.

APPENDIX C. DISSERTATIONS in SACE

Amaral, L. A. M. (2007). *Designing Intelligent Language Tutoring Systems for Integration into Foreign Language Instruction*. Unpublished doctoral dissertation. The Ohio State University: Ohio, USA.

Baines-Alarcos, M. P. (2012). *Gothic Fiction in an Australian Landscape: An Analysis of Gabrielle Lord's Tooth and Claw, Elizabeth Jolley's The Well and Tim Winton's the Winter Dark*. Unpublished doctoral dissertation. Universidad de Zaragoza: Zaragoza Spain.

Basols, J. F. M. (2012*). The Sound of Humor: Translation, Culture and Phonological Jokes.* Unpublished doctoral dissertation. Universitat Pompeu Fabra.

Carrasco, R. C. (2010). *Of Men and Cyborgs: The Construction of Masculinity in Contemporary U.S. Science Fiction Cinema.*Unpublished doctoral dissertation. Universidad de Huelva

Carril, R. F. (2009). *English Vocabulary Teaching and Learning in the Galician EFL Context. The Role and Importance of Vocabulary Learning Strategies*. Unpublished doctoral dissertation. Universidade de Santiago de Compostela.

Cazzoli-Goeta, M. A. (2006). *The L2 Acquisition of Spanish Non-nominative Subjects by Adult L1 English Speakers.* Unpublished doctoral dissertation. Durham University: Durham, UK.

Conde, N. V. (2009). *Aspects of Expressive Strength Reduction in British English and Castilian Spanish: A Corpus-Based Analysis*. Unpublished doctoral dissertation. Universidade de Santiago de Compostela.

Dintrans, G. S. C. (2011). *Acquisition of Morphosyntax in the Adult Second Language:The Phonology Factor*. Unpublished doctoral dissertation. University of Iowa.

Férez, P. C. (2008). *Motion in English and Spanish: A Perspective from Cognitive Linguistics, Typology and Psycholinguistics*. Unpublished doctoral dissertation. Universidad de Murcia.

Durandegui, A. B. (2007). *The Guggenheim Bilbao Museum in the Basque Nationalist Press: A Discursive and Rhetorical Analysis*. Unpublished doctoral dissertation. Loughborough University.

Fernandez-Rubiera, F. J. (2009*). Clitics at the Edge: Clitic Placement in Western Iberian Romance Languages.* Unpublished doctoral dissertation. Georgetown University.

Fernandez, I. P. (2009). *In Search of New Spaces: Contemporary Black British and Asian British Women Writers.* Unpublished doctoral dissertation. Universidad de Oviedo.

Flores, H. L. G. (2011). *Promoting Positive Ethnolinguistic Identity in the Heritage Language Classroom Through Dialect Awareness*. Unpublished doctoral dissertation. Texas A&M University: Texas, USA.

Gomez, J. M. U. (2011). *Metaphor in Specialised Language: An English-Spanish Comparative Study in Marine Biology.* Unpublished doctoral dissertation. Universidad de Granada.

Hernández, I. Z. (2007). *Demonstrative Pronouns in Spanish: A Discourse-Based Study.*Unpublished doctoral dissertation. Ohio State University: Ohio, USA.

Jódar, A. R. (2012). *The Iconical Dıscourse Community, the Narratıve Iconical Subgenres and the Trauma Graphıc Novel in English*. Unpublished doctoral dissertation. Universidad de Zaragoza: Zaragoza Spain.

Litzler, M. F. (2011). *A Corpus of Middle English Medical Prologues in the Sloane Collection of the British Library: An Introduction to the Genre in Prose*. Unpublished doctoral dissertation. Universidad de Palmas de Gran Canaria.

López-Cortina, J. (2007). *The Spanish Left Periphery: Questions and Answers.* Unpublished doctoral dissertation. Georgetown University.

Lopez-Sako, N. I. (2008). *Politeness in American English, Spanish and Japanese: The Case of Disagreements in Conversation*. Unpublished doctoral dissertation. Universidad de Granada.

Madrigal, D. M. (2010).*"Beyond 'Spanglish'": Ideologies of Language and Identity in Bilingual Chicana/O Cultural Production*. Unpublished doctoral dissertation. Stanford University.

Marbol, G. A. (2011). *Quantity and Rate of Vocabulary Acquisition in the Context of Non-systematic Input: Elementary Education Students of English as a Foreign Language*. Unpublished doctoral dissertation. Universidad de Murcia.

Martin, P. A. (2005). *A Genre–Analytic Study of English and Spanish Research Article Abstracts in the Experimental Social Sciences*. Unpublished doctoral dissertation. Universidad de la Laguna.

Mora, L. C. (2006). *How to Make a Mountain out of a Molehill: A Corpus-Based Pragmatic and Conversational Analysis Study of Hyperbole in Interaction*. Unpublished doctoral dissertation. Universitat de Valencia: Valencia, Spain.

Navas, D. L. (2012). *Comparative Study of Two Foreign Language Examinations for University Entry in England and Spain*. Unpublished doctoral dissertation. Universitat de Politecnica de Valencia: Valencia, Spain.

Pastor, M. D. G. (2007). *A Socio-Cognitive Approach to Political Interaction: An Analysis of Candidates Discourses in U.S. Political Campaign Debates*. Unpublished doctoral dissertation. Universitat de Valencia: Valencia, Spain.

Perez, A. P. (2009). *Subject Expression in Minorcan Spanish: Consequences of Contact with Catalan*. Unpublished doctoral dissertation. The Pennsylvania State University: Pennsylvania, USA.

Pomacóndor, S. M. C. (2007). *Representations and Transfer Processes in L2 Speech Production: Evidence from Catalan Learners of English*. Unpublished doctoral dissertation. University of Edinburgh: Edinburgh, Scotland.

Ponsoda, E. M. (2011). *Multilingual Lexico-Syntactic Patterns for Ontology Modelling and Linguistic Information Repository for Ontology Localization*. Unpublished doctoral dissertation. Universidad Politécnica de Madrid: Madrid, Spain.

Ramirez, J. M. P. (2012). *Language Switching: A Qualitative Clinical Study of Four Second Language Learners' Composing Processes*. Unpublished doctoral dissertation. University of Iowa.

Recarte, C. A. (2012). *Jazz and Blues-Idiomatic Mythography and Mythopoeia in Contemporary American Culture*. Unpublished doctoral dissertation. Universidad Complutense de Madrid: Madrid, Spain.

Riaza, B. G. (2012). *Attribution and Thematization Patterns in Science Popularization Articles of the Guardian Newspaper*. Unpublished doctoral dissertation. Universidad de Salamanca.

Sanchez, H. S. (2010). *An Investigation into the Relationships among Experience, Teacher Cognition, Context, and Classroom Practice in EFL Grammar Teaching in Argentina.* Unpublished doctoral dissertation. University of Warwick.

Sanchez, R. C. (2008). *Patterns of Activity Sequencing in the Teaching of English as a Foreign Language and Their Effects on Learning: A Quasi-Experimental Study.* Unpublished doctoral dissertation. Universidad de Murcia.

Sempere, A. G. (2006). *Conflicting Quantity Patterns in Ibero-Romance Prosody*. Unpublished doctoral dissertation. University of Texas: Texas, USA.

Serrano, R. S. (2007). *Time Distribution and the Acquisition of English as a Foreign Language.* Unpublished doctoral dissertation. Universitat de Barcelona: Barcelona, Spain.

Toscano-Fuentes, C. M . (2011). *Empirical Study of the Relationship among the Level of a Second Language Acquisition, Auditory Skill and Musical Intelligence of Students*. Unpublished doctoral dissertation. Universidad de Huelva: Huelva, Spain.

Vallejo, D. C. M. (2009). *Focalizing Ser ('To Be') in Colombian Spanish*. Unpublished doctoral dissertation. Indiana University: Indiana, USA.

Varela, L. R. (2012). *New Insights into the Study of English False Friends: Their Use and Understanding by Spanish Learners of English*. Unpublished doctoral dissertation. Universidade de Santiago de Compostela.

Vayá, A. R. (2008). *An ESP Comparative Analysis in Medical Research Articles: Spanish-English.* Unpublished doctoral dissertation. Universidad Politécnica de Valencia: Valencia, Spain.

CURRICULUM VITAE

A. PERSONAL INFORMATION

Name : Reyhan AĞÇAM

Date of Birth : 29/10/1983

Place of Birth: Adana

E-mail : reyhanagcam@gmail.com

B. EDUCATIONAL BACKGROUND

Date	Degree	Institution
2009-2014	Doctor of Philosophy	Çukurova University, Instıtute of Social Sciences
2006-2008	Master of Arts	Çukurova University, Instıtute of Social Sciences Division of English Language Teaching
2002-2006	Bachelor of Arts	Çukurova University English Language Teaching Department

C. PROFESSIONAL BACKGROUND

Date	Title	Institution
2012-	Instructor	Kahramanmaraş Sütçü İmam University Kahramanmaraş, Turkey
2010-2012	International Affairs Coordinator	Osmaniye Korkut Ata University Osmaniye, Turkey
2007-2010	EFL Teacher	Ministry of National Education Gaziantep, Turkey

D. ACADEMIC BACKGROUND

PAPERS PRESENTED

1. Ağçam, R. (2010). Second Language Acquisition of Any-Type NPIs by Turkish Adult Learners. 22nd International Conference on Second Language Acquisition and Foreign Language Learning. University of Silesia, Szczyrk, Poland, May 27-29, 2010.

2. Kırkgöz, Y. & Ağçam, R. (2011). Investigating Culture in Locally Published Textbooks in Turkish Primary Education. 15th International INGED ELT Conference "Taking It to the Limits", Hacettepe Üniversitesi, Ankara, Turkey, October 20-22, 2011.

3. Ağçam, R., Çoban, M. & Karadeniz, Z. (2012). We Like to Move It: L2 Acquisition of English Syntactic Movement in Noun Clauses by Turkish Adult Learners. 24th International Conference on Second Language Acquisition and Foreign Language Learning. University of Silesia, Szczyrk, Poland, May 17-19, 2012.

4. Ağçam, R. (2014). Author Stance in Doctoral Dissertations of Native and Non-Native Speakers of English: A Corpus-Based Study on Epistemic Verbs. LIF2014 – Language in Focus Exploring the Challenges and Opportunities in Linguistics and English Language Teaching. Antalya, Turkey, March 27-29, 2014.

5. Ağçam, R. (2014). Author Stance in Doctoral Dissertations of Native and Non-Native Speakers of English: A Corpus-Based Study on Epistemic Adverbs. CILC2014 International Conference on Corpus Linguistics. University of Las Palmas de Gran Canaria, Spain, May 22-24, 2014.

6. Ağçam, R. (2014). Author Stance in Doctoral Dissertations of Native and Non-Native Speakers of English: A Corpus-Based Study on Epistemic Nouns. 2nd International Symposium: Learner Corpus Studies in Asia and the World (LCSAW) 2014. University of Kobe, Kobe, Japan, May, 31-June 1, 2014.

7. Ağçam, R. (2014). Author Stance in Doctoral Dissertations of Native and Non-Native Speakers of English: A Corpus-Based Study on Epistemic Adjectives. 3rd International Conference on English Studies (IDEA 2014). Near East University, Nicosia, North Cyprus, June 19-20, 2014.

CHAPTERS IN BOOKS

1. Kırkgöz, Y. & Ağçam, R. (2011). Investigating Culture in Locally Published Textbooks in Turkish Primary Education. CEPS (Center for Educational Policy Studies) Journal. Vol.1., N.1, 2011, 153-167. Ljublijana, Slovenia (Joint Study).

2. Can, C. & Ağçam, R. (2011). Second Language Acquisition of Any-Type NPIs by Turkish Adult Learners. Nauczycielskie Kolegium Jezykow Obych w Zabrzu, 5-63. Uniwersytet Slaski, Zabrze, Poland.

ARTICLES IN JOURNALS

1. Kırkgöz, Y. and Ağçam, R. (2012). Investigating the Written Assessment Practices of Turkish Teachers of English at Primary Education. The Journal of Language and Linguistic Studies, 8 (2), 119-136. http://www.jlls.org/vol8no2/119-136.pdf