

KARADENİZ TECHNICAL UNIVERSITY * THE INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

APPLIED LINGUISTICS MASTER'S PROGRAM

**CORPUS ANALYSIS OF EPISTEMIC MODAL AUXILIARIES AND
LEXICAL VERBS IN ARGUMENTATIVE WRITING**

MASTER'S THESIS

Tuğba ÇITLAK

JUNE-2023

TRABZON

KARADENİZ TECHNICAL UNIVERSITY * THE INSTITUTE OF SOCIAL SCIENCES

DEPARTMENT OF WESTERN LANGUAGES AND LITERATURE

APPLIED LINGUISTICS MASTER'S PROGRAM

**CORPUS ANALYSIS OF EPISTEMIC MODAL AUXILIARIES AND
LEXICAL VERBS IN ARGUMENTATIVE WRITING**

MASTER'S THESIS

Tuğba ÇITLAK

Supervisor: Asst. Prof. Dr. Ali Şükrü ÖZBAY

JUNE-2023

TRABZON

APPROVAL

Upon the submission of the dissertation, **Tuğba ÇITLAK** has defended the study “**Corpus Analysis of Epistemic Modal Auxiliaries and Lexical Verbs in Argumentative Writing**” in partial fulfillment of the requirements for the degree of Master of Arts in Applied Linguistics at Karadeniz Technical University, and the study has been found fully adequate in scope and quality as a thesis by **unanimous/ majority** vote on .../.../2023.

Committee Member		Decision		Signature
Title - Name and SURNAME	Mission	Accept	Refuse	
Assoc. Prof. Dr. Ali Şükrü ÖZBAY	Supervisor	☐	☐	
Assoc. Prof. Dr. Nazan Yıldız	Member	☐	☐	
Assoc. Prof. Dr. Erdem Akbaş	Member	☐	☐	

Approval of the Graduate School of Social Sciences

Prof. Dr. Tülay İLHAN NAS
Director

DECLARATION OF ORIGINALITY

I, **Tuğba ÇITLAK**, hereby confirm and certify that;

- I am the sole author of this work and I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution,
- This work contains no material that has been submitted or accepted for a degree or diploma in any other university or institution,
- All data and findings in the work have not been falsified or embellished,
- This is a true copy of the work approved by my advisor and thesis committee at Karadeniz Technical University, including final revisions required by them.
- I understand that my work may be electronically checked for plagiarism by the use of plagiarism detection software and stored on a third party's server for eventual future comparison,
- I take full responsibility in case of non-compliance with the above statements in respect of this work.

Signature.....

23.06.2023

ACKNOWLEDGEMENT

I would like to express my sincere thanks and gratitude to the following people for their continuous support and contributions throughout the completion of this thesis.

First and foremost, I would like to thank my supervisor Assoc. Prof. Dr. Ali Şükrü ÖZBAY for his continuous support and invaluable feedback throughout each process of my thesis.

I am deeply grateful to the committee members of my thesis involving Assoc. Prof. Dr. Nazan Yıldız and Assoc. Prof. Dr. Erdem Akbaş for their useful feedback concerning the design and content of my thesis.

I am also grateful to Mgr. Kateřina Šteklová for her academic suggestions and guidance when I was at Jana Evangelisty Purkyně University in Ústí nad Labem, Czechia in 2019 and 2023. The beginning of my MA process owes much to her profound knowledge of the study area back in 2019.

I am also deeply grateful to the ICLE learner corpus project initiated by the Centre for English Corpus Linguistics at the University of Louvain (UCL), Belgium for enabling the researchers the opportunity to work with a great variety of learner corpora.

I would also like to express my gratitude to the Scientific and Technological Research Council of Turkey (TUBITAK) since I was funded by the Scientific and Technological Research Council of Turkey (TUBITAK) 2210-A National MSc/MA Scholarship Program for my entire MA process.

Finally, my profound thanks go to my family and my friend Ertürk Akyüz for their continuous encouragement and moral and material support throughout the whole process.

June, 2023

Tuğba ÇITLAK

TABLE OF CONTENTS

ACKNOWLEDGEMENT	IV
TABLE OF CONTENTS.....	V
ÖZET.....	VIII
ABSTRACT	IX
LIST OF TABLES	X
LIST OF FIGURES	XII
LIST OF ABBREVIATIONS.....	XIII
INTRODUCTION.....	1-2

CHAPTER ONE

1. FRAMEWORK OF THE STUDY.....	3-7
1.1. Background of the Study.....	3
1.2. Statement of the Problem	4
1.3. Purpose of the Study	5
1.4. Significance of the Study	5
1.5. Research Questions	6
1.6. Definition of Terms.....	6
1.7. Organization of Thesis.....	7

CHAPTER TWO

2. LITERATURE REVIEW.....	8-27
2.1. Modality	8
2.2. Epistemic Modality	9
2.2.1. Epistemic Modal Auxiliary.....	12
2.2.2. Epistemic Lexical Verbs.....	13
2.3. Corpus Linguistics	15
2.4. Contrastive Interlanguage Analysis	18
2.5. Corpus Research on Epistemic Modality.....	20
2.6. Using Corpus in the Teaching and Learning of English	22
2.7. Evidentiality and Modality.....	23
2.8. Evidentiality in English.....	24

2.9. Evidentiality in Argumentative Discourse	26
2.10. The Relationship between Epistemic Modality and Evidentiality	26
2.11. Native and Non-native Speakers in Applied Linguistics	27

CHAPTER THREE

3. METHODOLOGY.....	28-46
3.1. Introduction.....	28
3.2. Rationale for the Study.....	28
3.3. Selection Criteria for Epistemic Modality	29
3.4. Recording of Epistemic Devices.....	29
3.5. Development of the Criteria for Epistemic Classification	29
3.5.1. Some Aspects Regarding Modal Verbs	30
3.5.2. Some Aspects Regarding Lexical Verbs.....	32
3.6. Adopted Criteria for Epistemic Classification	33
3.7. Participants.....	34
3.8. Corpora.....	35
3.9. Data Collection	37
3.10. Data Analysis	37
3.11. Corpus Tool Used in This Study.....	38
3.12. Procedure	38
3.12.1. Preparation of the List of Epistemic Devices.....	38
3.12.2. Classification of Epistemic Devices Regarding Their Epistemic Strengths	39
3.13. The Extraction and Examination of Epistemic Devices in Four Corpora.....	41
3.14. Raw and Normalized Frequencies	45

CHAPTER FOUR

4. RESULTS AND DISCUSSION.....	47-116
4.1. Introduction.....	47
4.2. Frequency Distribution of Epistemic Lexical Verbs and Modal Auxiliaries.....	47
4.3. Epistemic Devices: Modal Auxiliaries.....	49
4.3.1. Frequency of Epistemic Modal Auxiliaries	49
4.3.2. Epistemic Strengths of Epistemic Modal Auxiliaries	53
4.3.3. Qualitative Analysis of Modal Auxiliaries' Elimination of Epistemic Uses from the Non-Epistemic Uses.....	58
4.3.4. Cases of Epistemic Modal Auxiliaries in Four Corpora	61
4.3.4.1. The Case of Epistemic 'May' and 'Might'	61
4.3.4.2. The Case of Epistemic 'Will' and 'Would'	67
4.3.4.3. The Case of Epistemic 'Must' and 'Should'	73

4.3.4.4. The Case of Epistemic ‘Could’	77
4.4. Epistemic Devices: Lexical Verbs	80
4.4.1. Frequency of Epistemic Lexical Verbs	80
4.4.2. Epistemic Strengths of Epistemic Lexical Verbs.....	83
4.4.3. Qualitative Analysis of Lexical Verbs’ Elimination of Epistemic Uses from the Non-Epistemic Uses.....	90
4.4.4. Cases of Epistemic Lexical Verbs in Four Corpora.....	93
4.4.4.1. The Case of Epistemic ‘Think’	93
4.4.4.2. The Case of Epistemic ‘Believe’	96
4.4.4.3. The Case of Epistemic ‘Show’	98
4.4.4.4. The Case of Epistemic ‘Seem’	100
4.4.4.5. The Case of Epistemic ‘Know/n’	101
4.4.4.6. The Case of Epistemic ‘Claim’	103
4.4.4.7. The Case of Epistemic ‘Feel’	103
4.4.4.8. The Case of Epistemic ‘Argue’	104
4.4.4.9. The Case of Epistemic ‘Consider’	106
4.5. Discussion	108
CONCLUSION AND RECOMMENDATIONS	117
REFERENCES	120
APPENDICES	127
CURRICULUM VITAE	149

ÖZET

Bu tez, yerli ve yabancı yazarların, tartışmacı yazılarında bilgisel kip olarak kullanılan yardımcı fiilleri ve ana fiilleri nasıl kullandıklarını araştırmaktadır. Diller arası karşılaştırma yapmayı sağlayan karşılaştırmalı diller arası analiz, farklı anadillerine sahip olan ve anadili İngilizce olan yazarların bu bilgisel kiplik belirteçlerini tartışmacı yazılarda argüman oluşturmak amacıyla nasıl kullandıklarını görmek için derlem tabanlı yöntem ile birlikte kullanıldı. Bununla ilgili olarak, yazılarında bilgisel kip olarak kullanılan yardımcı fiiller ve ana fiiller, sıklık, çeşitlilik, gramer kategorileri, en çok kullanılan bilgisel kiplik belirteçlerinin eşdizimliliklerini ve bilgisel kipliğinin güç düzeyleri açısından incelenmiştir. Hem nicel hem de nitel analizler yapılmıştır. Nicel analiz için, frekanslar bulundu ve dört derlem arasında istatistiksel olarak anlamlı bir fark olup olmadığını belirlemek için Log olabilirlik testleri uygulandı. Çalışmanın niteliksel yönünde ise, derin bir karşılaştırma yapmak ve yapıları ayrıntılı olarak görmek için dizin analizleri incelenmiştir.

Sonuçlara göre, Norveççe, Japonca, Türkçe dillerini içeren farklı anadillere sahip kişilerin ve anadili İngilizce olan kişilerin tartışmacı yazılarında bilgisel kip olarak kullanılan yardımcı fiillerin ve ana fiillerin kullanımında farklılıklar bulundu. LOCNESS ve NICLE arasında bilgisel kiplik fiillerin kullanımı ve bilgisel kip *would* ve *will* kullanımı açısından benzerlik olduğu tespit edilmiştir. Farklı olarak, JPICLE ve TRICLE toplamda bilgisel kiplik fiillerini ve bilgisel kip *would*'u yeterince kullanmadı. Bunun yerine bilgisel kip *will*'i kullanmayı diğerlerinden daha çok kullandılar. Bilgisel kipliklerin güç düzeyleri söz konusu olduğunda, LOCNESS en çok orta düzeyde gücü kullanırken, JPICLE ve TRICLE en çok güçlü bilgisel gücü ve en az orta düzeyde bilgisel gücü kullanıldığı görülmüştür. NICLE, bilgisel güçleri neredeyse aynı şekilde kullanmıştır; ancak en çok zayıf bilgisel gücü kullanmıştır.

Bilgisel sözcüksel fiiller için, dört derlem, bilgisel kiplik fiillerini epistemik sözcüksel fiillerden daha fazla kullanmıştır. Diğer taraftan bilgisel fiilleri en çok JPICLE, en az ise TRICLE kullanmıştır. LOCNESS en çok bilgisel fiil *show*'u tercih ederken, diğer üç derlem bilgisel *think*'i fazlasıyla kullanmıştır. JPICLE ve TRICLE, bilgisel fiil *believe* ve *seem*'i az kullanırken, NICLE ve LOCNESS bunları neredeyse aynı şekilde kullanmıştır. Bilgisel kipliklerin güç düzeyleriyle ilgili olarak, dört derlem en çok orta düzeyde bilgisel kiplik gücü kullanmıştır; ancak, JPICLE orta düzeyde bilgisel gücü fazlasıyla kullanmıştır. LOCNESS, JPICLE'den daha güçlü bilgisel kiplik gücü ve TRICLE'den daha fazla zayıf düzeyde bilgisel kiplik gücü kullanmıştır.

Anahtar Kelimeler: Öğrenen Derlemi, Bilgisel Kiplik, Tartışmacı Yazı

ABSTRACT

This thesis investigates how native and non-native writers employ epistemic modality by using epistemic modal auxiliaries and lexical verbs in their argumentative essays. Corpus-based contrastive interlanguage analysis, which enables a comparison between the languages, was employed to see how the writers from different L1 backgrounds and native speakers of English use epistemic modal auxiliaries and lexical verbs to make an argumentation in argumentative essays. The epistemic modal auxiliaries and lexical verbs were investigated regarding frequency, diversity, grammatical categories, co-occurrence patterns of the mostly used epistemic devices, and epistemic strength. Both quantitative and qualitative analyses were employed. In quantitative analysis, frequencies were counted along with the log-likelihood tests to determine whether there is a statistically significant difference across four corpora or not. As for the qualitative analysis, concordance lines were examined to observe the structural patterns for comparison.

According to the results, there were differences in the use of epistemic modal auxiliaries and lexical verbs in argumentative essays of Norwegian, Japanese, Turkish writers, and their counterparts. It was found there is a similarity between LOCNESS and NICLE in terms of the use of epistemic modal auxiliaries and the use of epistemic *would* and *will*. Differently, JPICLE and TRICLE underrepresented epistemic modal auxiliaries in total, especially the epistemic *would*. On the other hand, JPICLE and TRICLE preferred to use epistemic *will* more than the other two corpora. When it comes to epistemic strengths, LOCNESS used the medium strength the most whereas JPICLE and TRICLE used the strong epistemic strength the most and the medium epistemic strength the least. NICLE used the epistemic strengths nearly the same; however, it used the weak epistemic strength the most.

For epistemic lexical verbs, four corpora used epistemic modal auxiliaries more than epistemic lexical verbs. On the other hand, JPICLE used it the most while TRICLE used it the least. LOCNESS preferred to use epistemic lexical verb *show* the most whereas the other three corpora overrepresented epistemic *think*. JPICLE and TRICLE underrepresented epistemic lexical verbs *believe* and *seem* while NICLE and LOCNESS used them nearly the same. Regarding the epistemic strengths, four corpora used the medium epistemic strength the most; however, JPICLE overrepresented the medium epistemic strength. LOCNESS used more strong epistemic strength than JPICLE and weak epistemic strength more than TRICLE.

Keywords: Learner Corpus, Epistemic Modality, Argumentative Essay

LIST OF TABLES

Table No	Table Name	Page Nr.
1	Types of Modality.....	8
2	Categorization of Modality Types	9
3	Interpersonal Function and its Dimensions	11
4	Lexico-Grammatical Features to Express the Stance	12
5	The Meanings of Deontic and Epistemic Modal Auxiliaries.....	13
6	Complementation Patterns for Perception Verbs Indicating Evidentiality.....	25
7	Overall Comparability of Four Corpora	36
8	The List of Epistemic Modal Auxiliaries and Lexical Verbs	39
9	Epistemic Strengths Expressed by Epistemic Modal Auxiliaries and Lexical Verbs.....	40
10	Inter-Rater Reliability Test for ELVs Resulting Low Agreement in LOCNESS	41
11	Inter-Rater Reliability Test for ELVs Resulting Low Agreement in NICLE	42
12	Inter-Rater Reliability Test for ELVs Resulting Low Agreement in JPICLE	42
13	Inter-Rater Reliability Test for ELVs Resulting Low Agreement in TRICLE.....	43
14	Inter-Rater Reliability Test for Epistemic Modal Auxiliaries in LOCNESS.....	43
15	Inter-Rater Reliability Test for Epistemic Modal Auxiliaries in NICLE.....	44
16	Inter-Rater Reliability Test for Epistemic Modal Auxiliaries in JPICLE.....	44
17	Inter-Rater Reliability Test for Epistemic Modal Auxiliaries in TRICLE	45
18	Frequency Distribution of Epistemic Modal Auxiliaries in Four Corpora	49
19	Epistemic Strengths of Epistemic Modal Auxiliaries.....	53
20	Frequency Distribution of Weak Epistemic Modal Auxiliaries in Four Corpora.....	53
21	Frequency Distribution of Medium Epistemic Modal Auxiliaries in Four Corpora.....	54
22	Frequency Distribution of Strong Epistemic Modal Auxiliaries in Four Corpora	54
23	Frequency Distribution of Epistemic Strength in Four Corpora.....	54
24	Log-Likelihood Tests for Epistemic Strength in NICLE and JPICLE	55
25	Log-Likelihood Tests for Epistemic Strength in NICLE and LOCNESS	56
26	Log-Likelihood Tests for Epistemic Strength in NICLE and TRICLE	56
27	Log-Likelihood Tests for Epistemic Strength in JPICLE and LOCNESS	56
28	Log-Likelihood Tests for Epistemic Strength in JPICLE and TRICLE	57
29	Log-Likelihood Tests for Epistemic Strength in LOCNESS and TRICLE	57
30	Top Ten Epistemic Lexical Verbs in Four Corpora.....	80

31	Epistemic Strengths of Epistemic Lexical Verbs.....	83
32	Frequency Distribution of Weak Epistemic Lexical Verbs in Four Corpora	84
33	Frequency Distribution of Medium Epistemic Lexical Verbs in Four Corpora	85
34	Frequency Distribution of Strong Epistemic Lexical Verbs in Four Corpora	85
35	Frequency Distribution of Epistemic Strength in Four Corpora.....	86
36	Log-Likelihood Tests for Epistemic Strength in NICLE and JPICLE	86
37	Log-Likelihood Tests for Epistemic Strength in NICLE and LOCNESS	87
38	Log-Likelihood Tests for Epistemic Strength in NICLE and TRICLE	87
39	Log-Likelihood Tests for Epistemic Strength in JPICLE and LOCNESS	87
40	Log-Likelihood Tests for Epistemic Strength in JPICLE and TRICLE	88
41	Log-Likelihood Tests for Epistemic Strength in LOCNESS and TRICLE	88
42	Epistemic Lexical Verbs that are not Used in Four Corpora	89



LIST OF FIGURES

Figure No	Figure Name	Page Nr.
1	Research Design	46
2	Comparison of the Number of Words in Four Corpora.....	47
3	Overall Frequencies of EMAs and ELVs in Four Corpora.....	48
4	Top Five Epistemic Modal Auxiliaries in Four Corpora.....	50
5	Top Five Epistemic Modal Auxiliaries in LOCNESS.....	51
6	Top Five Epistemic Modal Auxiliaries in NICLE.....	51
7	Top Five Epistemic Modal Auxiliaries in JPICLE.....	52
8	Top Five Epistemic Modal Auxiliaries in TRICLE.....	52
9	Overall Frequencies of Epistemic Strength of Modal Auxiliaries in Four Corpora	58
10	Top Five Epistemic Lexical Verbs in NICLE	81
11	Top Five Epistemic Lexical Verbs in LOCNESS	82
12	Top Five Epistemic Lexical Verbs in JPICLE.....	82
13	Top Five Epistemic Lexical Verbs in TRICLE	83
14	Overall Frequency of Epistemic Strengths of Lexical Verbs in Four Corpora.....	89

LIST OF ABBREVIATIONS

CIA	: Contrastive Interlanguage Analysis
EFL	: English as Foreign Language
ELV	: Epistemic Lexical Verb
EMA	: Epistemic Modal Auxiliaries
ESL	: English as Second Language
JPICLE	: Japanese Subcorpus
L1	: First / Native language
L2	: Second / Foreign
LL	: Log-Likelihood
LOCNESS	: Louvain Corpus of Native English
N / n	: Number / Frequency
<i>nf</i>	: Normalized Frequency
NICLE	: Norwegian Subcorpus
TRICLE	: Turkish Subcorpus

INTRODUCTION

Argumentative essays can be described as a prominent genre of writing involving making arguments against or for other propositions by expressing opinions varying from certainty to uncertainty. As one of the ways of making argumentations, epistemic modality is used to convey the uncertainty of the writer to the proposition that the writer makes, which indicates the commitment of the writer to the claims the writer makes (Vold, 2006; Halliday et al., 2004). Epistemic modality can be expressed through epistemic modality markers, such as modal verbs and lexical verbs (Greenbaum, 1996). Concerning the various types of expressing epistemic modality in English, epistemic modality markers have been subjected to various studies regarding applied linguistics (Biber, 2006; Hyland & Milton, 1997).

In the context of EFL and ESL, previous studies investigated the use of epistemic modality markers in the writing of non-native speakers of English (Yang & Zeng, 2021; Bouhlal et al., 2018; Vandenhoeck, 2018; Yang, 2018; Karahan, 2017; Khoshshima, 2016; Hu & Li, 2015; Ngula, 2015; Zhang, 2015; Chen, 2010; McEnery & Kifle, 2002; Milton & Hyland, 1999). The teaching of epistemic modality markers acts as a role in the relationship between the reader and the writer in an intercultural pragmatic framework, because the writers are supposed to both draw a line between the facts and their opinions and convey their positions and confidence in their propositions (Hu & Li, 2015; Hyland & Milton, 1997). Furthermore, having knowledge of a wide range of usages of epistemic modality markers can allow the L2 learners “to make claims with the exact degree of certainty or doubt that they intend” (McEnery & Kifle, 2002: 183). Thus, the efficient use of epistemic modality markers can enable both native and non-native writers to position themselves in their claims in argumentative essays.

Concerning this, teaching and learning these epistemic devices can be a challenge for L2 learners. This challenge may stem from the nature of epistemic devices because epistemic devices are polypragmatic, which means they can convey different meanings, and the meanings of epistemic devices tend to overlap (Hu & Li, 2015). Moreover, involving in the target language’s discourse community can be distorted by the complexity of epistemic devices and lack of having enough intercultural pragmatic competence (Chen, 2010; Hyland & Milton, 1997). To overcome these challenges, the use of corpus can be regarded as a way to reach the authentic use of the target language (Granger, 2002). In this case, learner corpora can be used through corpus-based contrastive interlanguage analysis to see the similarities and differences regarding the target language’s discourse community (Granger, 1998). This type of investigation enables the researcher

to compare the use of language by those who have different L1 backgrounds for the improvement of L2 learners, teachers, and pedagogical materials (Granger, 1998). To this end, this study aimed to examine the use of epistemic devices involving epistemic modal auxiliaries and lexical verbs in the argumentative essays of undergraduate Turkish, Japanese, and Norwegian learners of English along with the undergraduate native speakers of English.



CHAPTER ONE

1. FRAMEWORK OF THE STUDY

1.1. Background of the study

There are several features of argumentative essays to be called a proficient writer. Making arguments can be regarded as one of the key elements of argumentative essays (Wentzel, 2018). In this regard, the meaning of argument plays a significant role in comprehending the ways of making argumentations in argumentative essays. Therefore, argument is “the act of making claims” suggesting different perspectives, beliefs, and thoughts, not a fact (Wentzel, 2018: 3). To this end, arguments can be created by using words that can help to change the degree of certainty, which involves epistemic modality (Wentzel, 2018). They can help to reduce the possibility that the claim is true, to create a distance between the claim and the person making the claim, and to make the claim closer to the personal level (Wentzel, 2018).

In relation to the expression of uncertainty in argumentative essays to make an argument, modality can be regarded as a way to express the opinion of the person making claims involving indeterminacy and uncertainty (Halliday et al., 2004). Modality can be divided into two: epistemic and deontic (root) modality, and it can be carried through modal auxiliaries, nouns, lexical verbs, adjectives, and adverbs (Palmer, 2001; Greenbaum, 1996). Among them, epistemic modality conveys the varying commitment degrees of the person making claims and arguments (Palmer, 2001). Epistemic modality markers indicate the varying degree of truth value of the content that is proposed (Vold, 2006). Thus, in argumentative essays, using epistemic modality markers can be a way to express the claim and argument by reducing the degree of certainty. These epistemic modality markers can be investigated through corpus-based studies regarding how they are used by both natives and non-natives.

In this regard, teaching and learning of English through corpus has been a topic that has been discussed by many scholars (Hunston, 2022; Flowerdew, 2011; O'Keeffe et al., 2007; Granger, 2002; Johns, 1991). There are two ways to teach English: direct and indirect ways (Breyer, 2011). Indirect ways of teaching English involves designing syllabus and textbooks by using corpus and creating learner corpora. Learner corpora can be used for contrastive analysis and error analysis studies to improve teaching materials for English learners concerning the understanding of the process of learning English and meeting the needs of the L2 learner (Breyer, 2011; Granger, 2002).

For the direct ways of teaching English, computer-assisted language learning and data-driven learning enable English instructors to teach English explicitly (Breyer, 2011). Data-driven learning is based on the discoveries made by the learners in the role of language detectives, which suggests an autonomous learning process (Johns, 1991).

To use corpus for teaching epistemic modality markers requires several steps. These can be the difficulties of the L2 learners, the differences or similarities between non-natives' and natives' use of epistemic markers, which are epistemic modal verbs and lexical verbs. With the help of contrastive interlanguage analysis, the challenges, differences, and similarities can be investigated. Moreover, in the related literature, using these modality markers appropriately to express uncertainty can be a challenging task for the language learners (Hyland & Milton, 1997). For this reason, the starting point of this study comes from the desire to learn these challenges if there is and to learn if there are any differences between natives and non-natives in using the epistemic lexical verbs and modal verbs. To this end, this study was inspired by studies of Ngula (2015) and Karahan (2017), who investigated the use of epistemic devices in native and non-native argumentative essays.

1.2. Statement of the Problem

For the novice and non-native writers, one of the difficulties in argumentative essays can be stating a position as a writer and making arguments (Wentzel, 2018). The writer can state his/her position by making arguments and expressing uncertainty in a proper way. The expression of uncertainty involves expressing the opinion of the person making claims and arguments (Halliday et al., 2004). Positioning oneself in the argumentative essay can be achieved by the epistemic modality, which serves to express uncertainty and reveal the positions. Epistemic modality indicates "the extent of the writer's conviction in their truth, which may range from uncertain possibility to confident prediction, but also convey a suitable degree of deference and modesty to the audience" (Hyland & Milton, 1997: 183). In this regard, the use of epistemic modality markers to express uncertainty as rhetorical devices can be a complex and challenging task for language learners (Hyland & Milton, 1997). Moreover, this challenge can be because of the multiple ways to express the writer's commitment to his/her statements and the multiple meanings conveyed with epistemic modality markers (Hyland & Milton, 1997). Regarding this, since even not all native writers can employ the epistemic markers to make arguments, there can be a question concerning the novice and non-native writers' situations in how learners from different L1 backgrounds differ in the expression of uncertainty in English (Hyland & Milton, 1997). The fact that these learners come from different L1 backgrounds and having different cultural backgrounds may reveal several differences in the use of epistemic modal auxiliaries and lexical verbs in the argumentative essays. Furthermore, modal auxiliaries are used with the main verb to express possibility while lexical verbs are regarded as the main verbs of a sentence (Halliday et al., 2004). Since these two can be

used together in a sentence and they are the essentials parts of a sentence syntactically, epistemic modal auxiliaries and lexical verbs were analyzed together in this study.

1.3. Purpose of the Study

This study aimed at investigating epistemic devices, especially both epistemic modal auxiliaries (EMAs) and epistemic lexical verbs (ELVs), used in the argumentative essays of Norwegian, Japanese, Turkish, and English undergraduate students. Norwegian, Japanese, and Turkish corpora were collected from The International Corpus of Learner English (henceforth ICLE). For English corpora, LOCNESS (The Louvain Corpus of Native English Essays) was used. After gathering the data for epistemic devices, an inter-rater reliability test was applied to ensure the reliability. Also, this study used a three-way categorization adopted by Ngula (2015) for the categorization of the epistemic devices according to their epistemic strengths: weak, medium, and strong. Both quantitative data and qualitative data were analyzed. To this end, the frequencies of epistemic devices and epistemic strengths of epistemic devices were analyzed. Moreover, a log-likelihood test was applied to the epistemic strengths of four corpora. For qualitative data, the concordance lines from four corpora were tentatively analyzed to see the grammatical structures in which epistemic devices were used.

1.4. Significance of the Study

Although there have been studies that focused on the use of epistemic modality in the argumentative essays of natives and non-natives of English, there is still room for the investigation of how Turkish, Japanese, and Norwegian non-native speakers of English use the epistemic modal auxiliaries and lexical verbs when compared to native speakers of English. Even though there are studies examining these languages separately, there is not any study investigating them through a contrastive interlanguage analysis of argumentative essays of Turkish, Japanese, and Norwegian non-native speakers of English and native speakers and English. In addition, Turkish and Japanese are agglutinating languages coming from the same language family: Ural Altaic. On the other hand, Norwegian and English come also from the same language family: Germanic. Since there have not been any previous studies conducted with these language backgrounds, this study tried to fill this research gap with the hope that the differences and similarities among these language background related to the use of epistemic modal auxiliaries and lexical verbs can be traced through a corpus-based contrastive interlanguage analysis. Thus, the argumentative essays of those who have these language backgrounds as their L1 were examined in terms of the use of epistemic modal auxiliaries and lexical verbs.

Furthermore, writing argumentative essays can pose a challenge for both native and non-native speakers of English, especially for novice writers in the sense that making an argument can

be a challenging task (Wentzel, 2018). In the process of making an argument in argumentative essays, epistemic modality can be regarded as an element to suggest varying degrees of commitment of the person making the argument or the claim (Palmer, 2001). In this regard, the choice of epistemic modality markers appropriately to convey the meaning of uncertainty for making an argumentation can create a challenge even for the native speakers of English (Hyland & Milton, 1997). Therefore, through a corpus-based approach, a corpus-based contrastive analysis of epistemic modal auxiliaries and lexical verbs in different languages answers these questions to some extent by revealing the process of learning English for non-native speakers of English and native speakers' use of epistemic modality. In addition, whereas epistemic modal auxiliaries were investigated in non-native writings, the use of epistemic lexical verbs was not given enough attention in previous studies (Vass, 2017). All in all, to give answers to these research niches, this study aimed to investigate both epistemic modal auxiliaries and lexical verbs in the argumentative essays of Turkish, Japanese, and Norwegian non-native speakers and native speakers of English through corpus-based contrastive language analysis.

1.5. Research Questions

This study aims at finding answers to the following research questions:

1. What is the frequency and diversity of epistemic modal auxiliaries used in four corpora?
2. What is the frequency and diversity of epistemic lexical verbs used in four corpora?
3. What is the frequency and diversity of epistemic strengths (weak, medium, and strong) in four corpora?
 - 3.1. Is there a statistically significant difference between the four corpora in terms of weak, medium, and strong epistemic strengths?
4. How do grammatical structures in which epistemic devices were used change in four corpora?

1.6. Definition of Terms

Epistemic Modality: “The degree of commitment by the speaker to what he says” (Palmer, 1986: 51).

Contrastive Interlanguage Analysis: A method that makes comparisons between the between interlanguages, having different native language backgrounds, of the same target language to detect the similarities and differences between them (Granger, 1998a).

Corpus Linguistics: An approach to language studies that includes a large amount of authentic language sample collection to investigate frequencies, concordances, and meanings by using a corpus software program (Hunston, 2022).

ICLE: International corpus of learner English (Granger et al., 2020).

1.7. Organization of Thesis

This study investigates the contrastive interlanguage analysis of epistemic modal auxiliaries and epistemic lexical verbs in argumentative essays of Norwegian, Japanese, Turkish, and English undergraduate students. Thus, the outline of this thesis was presented in the five following chapters.

Chapter 1, Framework of the study: This section presents the background of the study to set the rationale and aim of the study along with the significance of the study.

Chapter 2, Literature Review: This chapter is dedicated to the review of literature related to epistemic modality and its relationship with corpus-based studies by introducing seminal works.

Chapter 3, Methodology: This chapter presents the method this study used and the procedure during the application of the method.

Chapter 4, Results and Discussion: This section presents the results of both qualitative and quantitative data along with the discussions. It shows the results of the frequencies of epistemic modal auxiliaries and epistemic lexical verbs, the frequencies of epistemic strengths of both modal auxiliaries and epistemic lexical verbs, the log-likelihood tests, and grammatical structures in which epistemic devices were used.

Chapter 5, Conclusion and Recommendations: This chapter outlines the concluding remarks of the study and the recommendations for future studies.

CHAPTER TWO

2. LITERATURE REVIEW

2.1. Modality

Modality, regarding a semantic category, can refer to two kinds of judgements, which are about the factuality degree of what has been stated involving possibility and about the speaker's control over what has been said involving necessity (Auwera & Plungian, 1998; Greenbaum, 1996). In this regard, these meanings can be called epistemic and deontic sequentially conveyed (Greenbaum, 1996). Modality can be conveyed through modal auxiliaries, nouns, lexical verbs, adjectives, and adverbs (Facchinetti et al., 2003; Greenbaum, 1996). A more recent definition of modality is stated by Halliday et al. (2004) describing it as a way to express the speaker's opinion involving indeterminacy and uncertainty.

There are several types of modality along with epistemic and deontic modality. In this regard, Auwera and Plungian (1998) maintain four domains: participant-internal, participant-external, deontic, and epistemic modalities. First, participant-internal modality is related to both possibility (ability and capacity) and necessity of the speaker's internal situation that has been stated. On the other hand, participant-external modality refers to the situations external to the speaker related to a possibility or necessity. As a subordinate of participant-external modality, deontic modality expresses the situations external to the speaker related to permission or obligation. Lastly, epistemic modality expresses the speaker's judgment related to uncertainty and probability (Auwera & Plungian, 1998) (Table 1).

Table 1: Types of Modality

Possibility			
Non-epistemic Possibility			Epistemic possibility (Uncertainty)
Participant-internal possibility (Dynamic possibility, Ability, Capacity)	Participant-external possibility		
	(Non-deontic possibility)	Deontic possibility (Permission)	
Participant-internal necessity (Need)	(Non-deontic necessity)	Deontic necessity (obligation)	Epistemic necessity (Probability)
	Participant-external necessity		
Non-epistemic necessity			
Necessity			

Source: Auwera & Plungian, 1998: 32

A simple categorization of different kinds of modality has been made by Palmer (2003) stating them as epistemic, dynamic, and deontic modality. While deontic modality expresses the situation stated in the clause is external to the speaker's control, dynamic expresses the external situations controlled by the speaker. On the other hand, epistemic modality indicates the attitude of the speaker to the proposition's status.

Another classification of modality maintained by Halliday et al. (2004) includes modalization consisting of probability and usuality and modulation involving obligation and inclination. In this regard, probability in modalization means epistemic modality whereas obligation in modulation means deontic modality, as in (Halliday et al., 2004: 618):

- There can't be many candlestick-makers left.
- The roads should pay for themselves, like the railways.

As it can be seen, the first example expresses a probability referring to an epistemic meaning. On the other hand, the other example conveys a deontic meaning expressing an obligation. In addition, these types can be categorized in terms of subjectivity or objectivity, explicit or implicit, and degrees (Table 2).

Table 2: Categorization of Modality Types

	Subjective	Subjective	Objective	Objective
	explicit	implicit	implicit	explicit
Modalization: probability	I think [in my opinion] Mary knows.	Mary'll know	Mary probably knows [in all probability]	It's likely that Mary knows [Mary is likely to]
Modulation: obligation	I want John to go	John should go	John's supposed to go	It's expected that John goes

Source: Halliday et al., 2004: 620

Among these categorizations, probability can be categorized into two different ways: subjective and objective. Subjective probability refers to construing a subjectivity like the phrase *I think* whereas the objective probability expresses objectivity like the phrase *it is likely*. However, these examples are explicitly stated. When it comes to implicit way to say create these sentences, the modal auxiliaries *will* and *must* can be used to express implicit subjective probability while the adverbs *probably* and *certainly* can convey explicit objective probability (Halliday et al., 2004).

2.2. Epistemic Modality

Epistemic meaning can refer to a modal system not only including possibility but also indicating "the degree of commitment by the speaker to what he says. In particular, it should

include evidentials such as ‘heresay’ and ‘report’ (the quotative)” (Palmer, 1986: 51). Epistemic modality can mark the confidence of the writer in the knowledge and the source (Vold, 2006; Hyland, 1998). It includes an evaluation of the situation which is from absolute certainty about the situation that has been stated as real to absolute uncertainty about the situation that has been stated as not real. These two sides of certainty indicate a continuum involving probability to possibility (Cornillie, 2012). Thus, adverbs that convey probability explicitly can be regarded as expressing epistemic modality (Guziurová, 2016). Furthermore, hedges relate to epistemic modality since both convey the degree of probability (Guziurová, 2016).

These varying degrees can be conveyed in different ways: possibility, probability, and certainty, as in (Vold, 2006: 65):

- It is possible that smoking causes lung cancer.
- Smoking probably causes lung cancer.
- We know that smoking causes lung cancer.

The first example indicates a possibility expressed by an adjective in a noun-clause. The second one indicates probability by using an adverb in a clause. On the other hand, the last one does not express epistemic modality since it indicates certainty.

Epistemic modality can be related to interpersonal metafunction feature. Suggesting a modal called Systemic Functional Linguistics (SLF), Halliday (1994) notes situational context is a determiner of both the functions and meanings conveyed by language. As for the basic functions, there can be two basic ones: “making sense of our experience, and acting out our social relationships” (Halliday et al., 2004: 29). In this regard, Halliday et al. (2004) suggested three metafunctions as ideational, interpersonal, and lastly textual metafunctions (Table 3).

Table 3: Interpersonal Function and Its Dimensions

Category	Function	Examples
Interactive	Help to guide the reader through the text	Resources
Transitions	express relations between main clauses	in addition; but; thus; and finally
Frame markers	refer to discourse acts, sequences or stages	to conclude; my purpose is noted above;
Endophoric markers	refer to information in other parts of the text	see Fig; in section 2
Evidentials	refer to information from other texts	according to X; Z states
Code glosses	elaborate prepositional meanings	namely; e.g.; such as; in other words
Interactional	Involve the reader in the text	Resources
Hedges	withhold commitment and open dialogue	might; perhaps; possible; about
Boosters	emphasize certainty or close dialogue	in fact; definitely; it is clear that
Attitude markers	express writer's attitude to proposition	unfortunately; I agree; surprisingly
Self mentions	explicit reference to author(s)	I; we; my; me; our
Engagement markers	explicitly build relationship with reader	consider; note; you can see that

Source: Hyland, 2005a: 49

The ideational metafunction refers to construing the ideas and experiences of the subjects. Interpersonal metafunction refers to engaging with social relationships involving acting. Lastly, textual metafunction is to organize the textual elements coherently. Among these metafunctions, since interpersonal metafunction is dealing with the expression of “our appraisal of and attitude towards whoever we are addressing and what we are talking about” by using modal verbs or adverbs, epistemic modality can be related to the interpersonal function (Ngula, 2015; Haliday et al., 2004; Flowerdew, 1996).

Similarly, relating interpersonal feature with metadiscourse, Hyland (2005a) signifies metadiscourse includes the attitudes and assumptions of people who are communicating by indicating the attitudes towards the statements and the person whom the speaker is communicating with. In this regard, interpersonal function is related to metadiscourse indicating a social engagement. Interpersonal function can be divided into two dimensions: interactive and interactional. Even though epistemic modality can be used as a hedge or boosters, the use of other epistemic devices in this study such as lexical verbs may require wider categorization including the degree of probability.

When it comes to the writer’s attitude towards both the text and its readers, another aspect of epistemic modality can be the term *stance*. Analyzing the university registers regarding the markers of stance, Biber (2006: 99) reports on stance and its markers as a way to convey “attitudes that a speaker has about certain information, how certain they are about its veracity, how they obtained access to the information, and what perspective they are taking”. Related to the present study, lexico-grammatical features to express the stance have been defined by Biber (2006) (Table 4).

Table 4: Lexico-Grammatical Features to Express the Stance

Modal and semi-modal verbs		Stance complement clauses controlled by verbs			
Possibility/ permission/ ability	Can, could, may , might	Stance verb + that-clause		Stance verb + to-clause	
Necessity/ obligation	Must, should, had (better), have to, got to, ought to	Epistemic verbs:		Probability (likelihood) verbs	Appear, happen, seem, tend
Prediction/ volition	Will, would, shall, be going to	1. Certainty	Conclude, determine, know	Mental (cognition/perception) verbs (likelihood)	Believe, consider
		1. Likelihood	Believe, doubt, think	Desire/ intention/ decision verbs	Intend, need, want
		Attitude verbs	Expect, hope, worry	verbs of effort/facilitation	Attempt, help, try
		Speech act and other communication verbs (non-factual)	Argue, claim, report, say	Speech act and communication verbs	Advise, remind, request

Source: Biber, 2006: 101

According to Biber's (2006) categorization, only the epistemic aspect of stance is the concern of this study. Although this categorization is wide in range, the epistemic aspect of stance, such as probability, possibility, and likelihood can be regarded as a way to give information regarding the certainty of the subject on the statement.

2.2.1. Epistemic Modal Auxiliary

Modal auxiliary verbs can be regarded as a straightforward way used widely to express modality in English (Salager-Meyer, 1997). Modal auxiliaries can have two different meanings: deontic and epistemic (Greenbaum, 1996). Whereas deontic meanings have the speaker's control over the stated situation, epistemic meanings have the judgement of the speaker regarding the statement's truth value. Thus, a clause created by using a modal "always conveys a judgment by its source about the motivations of the modal clause's semantic subject" (Roberts et al., 2010: 243). However, there can be a situation in which epistemic and deontic meanings are used together, as in (Greenbaum, 1996: 260):

- Can you tell us how you first got involved in this project?

In this example, deontic meaning can be conveyed as an ability and epistemic meaning can be used as a possibility. Regarding this, modal auxiliaries can be used both with their deontic meaning and epistemic meaning along with the possibility of merging them together (Table 5).

Table 5: The Meanings of Deontic and Epistemic Modal Auxiliaries

Deontic and Epistemic Meanings of Modal Auxiliaries		
	Deontic Meanings	Epistemic Meanings
Can/could	Ability: Could you be a bit more specific than that Permission: You can take these	Possibility: These forces can be known and expressed only indirectly, through the condensations and displacements of dream imagery
May/might	Permission: If I'm that interested I'll ask you if I may have a piece	Possibility: You might be interested in it
Will/would	Volition: Right I'll ask her	Prediction: That would be what John Stuart Mill called the tyranny of the majority
Shall/should	Volition: Shall I go first Regulative: The committee shall consider the case after hearing any representations which the teacher may make. Obligation: I can remember when common sense said that for instance women were weaker than men women shouldn't wear trousers women should earn less than men	Probability: Lesions such as these should be capable of a good recovery in the long term. Prediction: I shall regret this for the rest of my life!
Must	Obligation: You must keep them moist	Certainty: They must have been his daughters, mustn't they

Source: Grreenbaum, 1996: 260-264

2.2.2. Epistemic Lexical Verbs

In addition to modal auxiliaries, several lexical verbs can also convey epistemic meaning. Epistemic lexical verbs enable the writer to make the discourse less severe before indicating the research gap in academic writing (Pérez-Llantada, 2010a). The epistemic use of lexical verbs can be seen as rhetorical mechanisms carrying varying degrees of confidence of the writer on the claims (Pérez-Llantada, 2010b). It can be seen as “the most transparent means of coding the subjectivity of the epistemic source and are generally used to hedge either commitment or assertiveness” (Hyland, 1998: 119). The lexical verbs indicating doubt or evaluation can carry epistemic meaning, such as the lexical verbs *seem* and *appear* (Salager-Meyer, 1997). However, Salager-Meyer (1997) does not give any other example than the lexical verbs *seem* and *appear* for epistemic lexical verbs. Regarding this, epistemic verbs can be used along with the modal lexical verbs to express hedge (Salager-Meyer, 1997: 129):

- Our results *seem to suggest* that in Third World countries the extensive use of land to grow exportation products tends to impoverish these countries' populations even more.

In this example, the epistemic lexical verb *seem* and modal lexical verb *suggest* maintained by Salager-Meyer (1997) are used to avoid the expression of certainty by giving it a flexibility to discuss the claim.

For the lexical verbs *think* and *believe*, they can express epistemic modality in medium strength whereas the lexical verb *know* can convey an epistemic meaning in strong strength (Huddleston & Pullum, 2017). Along with this, the epistemic meanings of the lexical verbs *suggest*, *indicate*, and *predict* can be stated as the most frequently used epistemic lexical verbs to convey mitigation (Hyland, 1998). Among the categorization of evidentiality made by Palmer (2001), speculative, deductive, and assumptive ways to indicate evidentiality also convey the epistemic meaning including uncertainty (Hyland, 1998). In addition, the quotative and sensory ways also may include epistemic meaning because they indicate doubt for the justification of evidence (Hyland, 1998). In this regard, according to Hyland (1998) epistemic lexical verbs can be divided into two categories: epistemic judgement verbs and epistemic evidential verbs. Epistemic judgement verbs convey an uncertainty based on human evaluation. Among these verbs, speculative verbs involve a hypothesis about the truth of the claim (Hyland, 1998: 120):

- Thus, we *propose* that this insert is the major site of interaction with membrane.

In this example, speculative verbs can convey epistemic meanings and performing an action rather than the description of the acts. In addition, instead of a verb conveying performance of an action, the speculative verbs conveying epistemic meaning can also carry a meaning involving a cognitive process or inferential reasoning (Hyland, 1998: 121-122):

- I *believe* that the major organizational principle of thylakoids is that of continuous unstacking and restacking of sections of the membranes.
- We *conclude* that the fluorescence quenching seen here is not due...

In these examples, the first one shows that the lexical verb *believe* is used to carry an epistemic meaning including a degree of uncertainty along with the meaning involving a cognitive process. On the other hand, the second one implies that the lexical verb *conclude* conveys epistemic reasoning along with inferential reasoning. In addition to these verbs, the lexical verbs *indicate* and *imply* can make the discourse impersonalized by indicating the source of epistemic judgements (Hyland, 1998).

When it comes to epistemic evidential verbs, it involves the justification of evidence “either based on the reports of others, the evidence of the writer’s sense, or feasibility of matching evidence to goals” (Hyland, 1998: 124). Through reporting verbs, hearsay evidence and epistemic meaning can be conveyed by indicating the degree of commitment to the report (Hyland, 1998: 124):

- Jofuku and Goldberg (1989) *showed* the expression of several Kunitz protease inhibitors in soybean tissue.

In this example, it can be suggested the verb *show* indicates both an epistemic meaning and evidential meaning through reporting by hedging the proposition. The most common lexical verbs can be the verbs *suggest* and *propose* (Hyland, 1998).

The second categorization of evidential lexical verbs involves sensory evidence and epistemic meaning. In this regard, perception and comprehension can be conveyed through these epistemic lexical verbs (Hyland, 1998: 125):

- This hypothesis *seems* plausible because UV-B-absorbing flavonoids accumulate in leaf epidermal cells, where they may protect the...

According to the example, the lexical verb *seem* suggests both epistemic meaning and evidential meaning regarding sensory evidence. Thus, it indicates both a visual evidence and a degree of uncertainty.

2.3. Corpus Linguistics

Corpus can be defined as “a collection of naturally-occurring language text, chosen to characterize a state or variety of a language” (Sinclair, 1991: 171). A similar definition also is that it is the collection of texts for a specific purpose of compiling a corpus (Flowerdew, 2012). Corpus is a large collection of authentic texts that consist of naturally occurring language samples (Hunston, 2022; Flowerdew, 2012; Tognini-Bonelli, 2001; Biber et al., 1998). In addition, Flowerdew (2012) mentions that corpus has clear design criteria, representing selected language, and is compiled for a particular purpose, mainly for linguistic.

Corpus linguistics can be defined as an approach to language studies that includes a large amount of authentic language sample collection to investigate frequencies, concordances, and meanings by using a corpus software program (Hunston, 2022). Regarding this, corpus linguistics includes “manipulation and observation of ‘what has been said (or written or signed)’... more about moving from observation to generalization and categorization of language features” (Hunston, 2022: 1-2). As for the corpus-based analysis, it can be regarded empirical since the patterns used in the naturally occurring language are analyzed both quantitatively and qualitatively by using a compilation of authentic texts collected largely and principally via computers (Conrad, 2000; Biber et al., 1998). Furthermore, it aims to describe the language that is used whereas

generative grammar developed by Chomsky tries to prescribe the language with a specific set of rules (Meyer, 2002).

Corpus can help to uncover many of the undetectable aspects of language (Reppen, 2021). Similarly, Hunston (2022) maintains the difference between corpus and other types of texts collections is that corpus software program is to discover the patterns that can be difficult to uncover with the conventional types of reading. Corpus-based studies can answer questions about the representation of language and the purpose of using language Hunston (2022). The strength of corpus linguistics is to be able to answer questions concerning from word level to discourse level (Reppen, 2021). The relationship between form and the meaning, the units of meanings and the patterns used overtly can be identified through corpus analysis (Sinclair, 1991).

Applied linguistics has been influenced by the developments in corpus linguistics concerning the way it investigates authentic language samples with specific contexts (Hunston, 2022; Anthony, 2013). It includes authentic language samples from different contexts, settings, and times, which enables the researchers to investigate and compare the language varieties from the specific samples gathered from specific contexts (Hunston, 2022). Thus, a corpus should be designed purposefully to get information for specific contexts (Hunston, 2022). To this end, texts should be chosen regarding whether they meet the criteria determined by the corpus' purpose (Hunston, 2022). In this regard, it is about whether the corpus is suitable for the designed criteria and purpose (Hunston, 2022).

The criteria for corpus design may be set with similar concerns. One of the researchers who set the criteria is Sinclair (1991). He (1991) remarks on the choice of texts regarding whether they will consist of written texts, spoken language, or both of them. Moreover, before compiling a corpus, whether the texts will be formal or informal and literary or ordinary should be considered (Sinclair, 1991). Among them, formal and literary language have been suggested as the easiest ways to reach texts (Sinclair, 1991). Another concern is regarding practicality to hinder prejudice (Sinclair, 1991). Typicality is mainly about being realistic while compiling texts in a way that the patterns will occur enough to be visible and typical (Sinclair, 1991). Moreover, what periods the corpus includes should be defined in the clearest way (Sinclair, 1991). For the general size of the corpus, Sinclair (1991: 18) suggests "a corpus should be as large as possible, and should keep on growing" because a large corpus can enable more samples than a small corpus does. It can be measured by the number of both words and texts (Hunston, 2022). Nevertheless, it can be maintained that calling a corpus large or small is relative and changeable compared to other corpora (Hunston, 2022).

Moreover, Hunston (2022) suggests three criteria for compiling a corpus, which are in regard to its size, its balance, and its representativeness. As for its balance, it does not mean that sub-

corpora should be used in equal sizes; it means although one sub-corpus consists of more tokens than the other one, the small sub-corpus should be still comparable (Hunston, 2022). Lastly, the representativeness of a corpus means the consistency between specific language use and the texts gathered to represent this language use (Hunston, 2022). A corpus is compiled to represent the language that is aimed to be compiled to make proper generalizations about the language (Clancy, 2021; Biber, et al., 1998; Biber, 1993). To provide more representativeness, the corpus should include more texts and examples to represent the text types and linguistic patterns (Flowerdew, 1996; Biber, 1993). In addition, the size of the corpus, text diversity and length, and number of the texts should be considered to make a corpus representative of a language or its variety (Clancy, 2021). For the criteria of compiling a corpus, a detailed record should be kept concerning the materials used in compiling a corpus (Reppen, 2021; Sinclair, 1991). In this regard, a corpus designed with a purposeful criterion can help to match the corpus with the aim of compiling it (Flowerdew, 1996).

In addition, there are several types of corpora that can be created according to different purposes. One type of corpus is general corpus, which is compiled to be a representative of a language or regional variety (Hunston, 2022). This type of corpus consists of a wide range of texts to be a representative of a language and to present all usages specific to the contexts (Hunston, 2022). On the other hand, a specialized corpus includes the texts specific to one situation to investigate only that situation (Hunston, 2022). Regarding this, comparing a specialized corpus with a general corpus enables to see the differences of a language feature (Hunston, 2022). Furthermore, a corpus can be compiled as synchronic that presents a snapshot of a particular period; however, it can also be diachronic which means compiling a corpus to see the changes over years (Hunston, 2022). Diachronic corpora include another sub-type called as monitor corpus which is compiled continuously over time (Hunston, 2022). There is also another type called parallel corpus consisting of translated version of texts (Hunston, 2022). On the other hand, comparable corpus consists of different languages to investigate the similarities (Hunston, 2022). Multimodal corpus enables the researcher to add visual information along written and spoken language (Hunston, 2022).

Moreover, the authentic language samples can include the language produced by both experts and learners (Hunston, 2022). Learner corpora enable the researchers to discover the aspects in which a learner group can have difficulty and to shed light on the language acquisition process and the variations of learner groups (Hunston, 2022). To this end, learner corpora are compiled to compare particular language learners with different native language backgrounds (Hunston, 2022).

2.4. Contrastive Interlanguage Analysis

In the field of second language acquisition, Contrastive Analysis Hypothesis was formulated by Lado (1957, as cited in Gass et al., 2013) suggesting that comparing languages can reveal the differences and similarities between the languages. These differences and similarities between the target language and the native language of the learner may be a hint for the second language learners' easiness and difficulties in learning the second language. In the 1960s, behaviorist theory suggested learning a language includes habit formation; thus, there can be native language interference (Rustipa, 2011). In the same way, according to this hypothesis, the reason behind making errors when learning a second language might be the interference of the native language of the learner. The Contrastive Analysis Hypothesis maintained that if there are similarities between the mother tongue and the target language, it can reinforce the positive transfer; on the other hand, if there are differences between them, it may bring about the negative transfer (Gass et al., 2013).

The contrastive analysis includes two views: a strong view and a weak view. On the one hand, the strong view maintains the idea that contrastive analysis can help to predict the errors of second language learners; thus, the success of the learners. On the other hand, the weak view suggests that contrastive analysis can explain the errors made by the learners (Gass & Selinker, 2008).

Nevertheless, there were criticisms arising from these views. One of them is the predictions made by contrastive analysis about the difficulties that second language learners can encounter could not be evidenced because some predictions did not occur as the difficulties for the learners. In the same way, the difficulties that second language learners encountered were not predicted by contrastive analysis (Ruspita, 2011). Therefore, it lacked empirical evidence.

As an alternative way, error analysis emerged after the criticisms on Contrastive Analysis Hypothesis and Error analysis. Error analysis is the first methodology dealing with the errors made by language learners by examining the learner language and trying to explain them. In that sense, the focus shifted from the errors due to the mother tongue to the errors made by the second language learners; therefore, the second language learners were the focus of error analysis (Gass et al., 2013). While contrastive analysis considered the interference due to the mother tongue as the reason for making errors, error analysis held the idea that errors may come from the subsystem of the learners. Furthermore, in contrastive analysis studies, the learners were regarded as passive learners for learning a language including making errors. On the other hand, in error analysis studies, the learners were thought as active while learning a second language and making errors (Lennon, 2008).

The shift of focus on errors was based on consideration of errors as significant in development for second language learners. The seminal paper by Corder (1967) named “The significance of learners’ errors” discussed the language learners’ errors as the indicators of developmental stage in learning a language by likening it to a child’s errors when learning his/her mother tongue by drawing on the perspectives of Chomsky on the acquisition of first language. Moreover, the error was explained by Corder (1967) as the systematic repeated event made by the learners as a sign of progress in language learning. On the other hand, he differentiated error and mistake by suggesting that mistakes may not be taken as signs of development because it is not repeated and systematic.

Although the error analysis was considered a solution to the criticism made on contrastive analysis, there was also criticism on error analysis especially made by Schachter (1974). For Schachter (1974), error analysis may ignore the structures avoided by the learners; the learners can avoid the structures when they are not familiar with them from their mother tongue. Specific structures can be avoided by the L2 learners thinking that there can be errors when they use the structures that seem difficult to use; therefore, it leads to the lower frequency of some structures. On the other hand, those who do not think these structures as problematic will not avoid using them; thus, these structures can be used with higher frequencies. Therefore, it was maintained that error analysis examines learners’ errors, which can highlight a limited perspective of the nature of language learning. Also, the sources of errors may not be easy to find since they can be ambiguous. (Larsen-Freeman & Long, 2014).

For the system special to language learners, the term *interlanguage* was introduced by Selinker (1972) after the criticism made on error analysis. Interlanguage is a term explaining the learners’ language system while learning a language. Interlanguage includes the elements from both the mother tongue, and the target language, and none of them, which means that the learner can formulate a unique “consistent and dynamic” interlanguage system that is different from their mother tongue and target language (Gass et al., 2013: 11; Selinker, 1972). Although it is dynamic, interlanguage can fossilize or stabilize (Gass et al., 2013; Ruspita, 2011).

Based on this, Contrastive Interlanguage Analysis (henceforth CIA) was termed by Granger (1998a) enabling comparisons between the target language and interlanguage and comparisons between interlanguages, having different native language backgrounds, of the same target language. For learners of a target language with different L1 backgrounds, this analysis tries to find out if there are any common ways to learn the target language by comparing the interlanguages. Therefore, in her seminal study, Granger (1998a) emphasizes the overuse and underuse of the items by checking their frequencies to investigate the features of the interlanguage of a specific target language.

On the other hand, there are several criticisms on the CIA: comparative fallacy and the issue of seeing natives as the norm. First, suggesting the comparative fallacy, Bley-Vroman (1983, as cited in Granger 2015) maintains that comparing the target language with interlanguage can be misleading regarding understanding the nature of interlanguage since they are different in nature. The other criticism is on the issue of taking natives of a target language into consideration as a norm. Since English is a Lingua Franca and there are World Englishes, the question of what native is can come to the surface. Thus, considering natives as a norm to learn a second language may be misleading if it is compared with interlanguages (Granger, 2015).

To address these criticisms, Granger (2015) revised the method. While the old method of CIA compares native language with interlanguage, the newly created method includes “reference language varieties” and interlanguage varieties (Granger, 2015: 17). This emphasis on the varieties of reference language may address the issue of seeing natives as the norm because it also includes the varieties of the native language. As for the interlanguage varieties, it is clear that it tries to highlight the variations in interlanguages. Furthermore, since most of the criticism comes from the terms ‘overuse’ and ‘underuse’ due to their negative connotations, Granger (2015) suggested new terms: overrepresentation and underrepresentation. In this way, these new terms can enable one to highlight the frequencies more statistically without suggesting being prescriptive.

2.5. Corpus Research on Epistemic Modality

In the previous studies, the use of epistemic modality was investigated through corpus-based studies; thus, there have been several studies conducted to examine the use of epistemic modality regarding different aspects via corpus-based studies. There are previous studies focusing on the epistemic modality used by non-natives of English in academic writings. One of them was conducted by Hyland and Milton (1997) investigating the use of epistemic modality in academic texts. It was found that the texts written in English by Cantonese speakers, when compared to natives of English, have limited varying of epistemic devices, express stronger commitments, and have problems in conveying the degree of certainty. Extending their study, Milton and Hyland (1999) examined the use of lexical phrases written by native speakers of English and Chinese students resulting that hedges and boosters were used as multi-words with a limited number by the non-natives and that English proficiency is a factor in using these expressions. As an answer to this discussion, a study conducted by Oh and Kang (2013) investigated the use of epistemic modality by Korean students regarding English proficiency. It was concluded the higher proficient the students are, the more approximate they are to native-like use of epistemic modality. To be more precise, the students with higher proficiency were found to use modal verbs, nouns, and adjectives more while they used lexical verbs less. Moreover, it was revealed that the students with higher proficiency used epistemic devices from the probability category more in a more diverse

syntactical context. In this regard, the studies about the use of epistemic modality were conducted through a comparison between natives and non-natives via corpus linguistics.

When it comes to the sources that were investigated, there are studies that investigate research articles, academic writings, argumentative essays, textbooks, and theses (Bouhlal et al., 2018; Yang, 2018; Khoshsima, 2016; Zhang, 2015). Gathering the research articles written in English by American and Spanish scholars and in Spanish by Spanish scholars, Pérez-Llantada (2010a) compared two contexts by focusing on the use of epistemic lexical verbs. As a result, it was found that there is a hybrid discourse in the English articles written by Spanish scholars changing their discourse. Because, it was revealed there is a homogeneity between the articles in that the epistemic lexical verbs and textual features used by Spanish scholars were similar to that of English scholars in the parts of articles. Differently, Khamkhien (2021) examined adjectives indicating epistemic modality in research articles in applied linguistics. It was found that the articles were interaction-oriented, and epistemic adjectives were used when making propositions and discussing their claims. In addition, the use of epistemic devices was investigated through academic writings of European non-natives and natives of English resulting that they both similar in using epistemic devices although there were several features of non-natives found in their academic writings (Vandenhoeck, 2018). A similar study conducted by Btoosh (2019) examined the epistemic modal verbs written in academic writings of Arabian EFL students and natives of English. As a result, it was found that non-natives were underused and overused several epistemic modal verbs. Consequently, the use of epistemic modality with different aspects was examined through research articles and academic writings.

Along with the research articles, there are studies that have focused specifically on argumentative essays (Chen, 2010). One of these studies was conducted by McEnery and Kifle (2002) comparing a learner corpus and a corpus consisting of argumentative essays of English speakers. They investigated the use of epistemic modality resulting that non-natives were found to use weaker epistemic devices and more tentative ones. The same results were also found in the study of Hu and Li (2015) investigating the use of epistemic modality by comparing Asian learners of English and natives of English. Moreover, they have found that while lower-band students used limited number of epistemic devices indicating stronger claims, the higher-band students used more tentative expressions in a way that the natives use. Similarly, a study conducted by Yang and Zeng (2021) investigated the argumentative essays of Chinese non-native speakers of English regarding their use of epistemic modality. It was found the non-natives use a limited range of epistemic devices. Similarly, a study conducted by Karahan (2017) examined the use of epistemic modality in the argumentative essays of Turkish non-native speakers of English and native speakers of English. As a result, Turkish students were found to use a limited number of epistemic devices, which are mostly strong level of epistemic degree. Other than argumentative essays, a study conducted by Gabrielatos and McEnery (2005) investigated MA dissertations regarding the use of epistemic

modality resulting that non-native speakers were found to use a limited number of epistemic modals. As can be seen, the studies focusing on epistemic modality investigated different sources, such as academic writing and argumentative essays.

Moreover, several studies focusing on the use of epistemic modality in academic writing, research articles, or argumentative essays also included different genres, disciplines, and registers (Taşpınar, 2017). The use of epistemic modality in different genres and registers was investigated by Carretero (2002), who found that the expression of epistemic modality can depend on the purpose of the text and individuals, planning, accuracy, topic, and the social relationships between the tenors. Another research focused on the use of epistemic modality in medical research articles by non-natives of English resulting that the epistemic modality was used low and medium degrees along with the use of epistemic modality in objective and explicitly objective orientation (Yang et al., 2015). Furthermore, a study conducted by Ngula (2015) investigated the epistemic modality in the argumentative essays of Ghanaian research articles in English written in different disciplines resulting that the Ghanaian writers used epistemic modality markers with medium and strong levels and different underuses and overuses. Differently, Karahan (2022) examined the articles written in Humanities & Social Sciences concerning the use of epistemic modal verbs. Other than epistemic modal verbs, the use of epistemic adverbials was investigated across different disciplines by Rozumko (2017) resulting that some epistemic adverbials were more prominent in specific disciplines. In another study, the use of epistemic modal markers was examined in the academic writings of medicine and humanities disciplines. The results showed different co-occurrences of modal verbs in different disciplines (Panocová & Lukačín, 2019). In this regard, the studies investigated different disciplines or genres regarding the use of epistemic modality focusing on different markers of epistemic modality.

In conclusion, previous studies examined the use of epistemic modality in the several writing genres by comparing several languages with a reference corpus involving native writings. However, there have not been any study that investigated the argumentative essays written by Turkish, Japanese, and Norwegian students in terms of the use of epistemic modal auxiliaries and lexical verbs. In this regard, this study tried to fill this research gap in the literature through a corpus-based interlanguage analysis to find out the similarities and differences among four language backgrounds involving a reference corpus written by those who are the native speakers of English.

2.6. Using Corpus in the Teaching and Learning of English

Other than investigating epistemic modality through corpus linguistics, teaching and learning English through corpus can be suggested as another concern that has been discussed by many scholars (Hunston, 2022; Flowerdew, 2011; O'Keeffe et al., 2007; Granger, 2002; Johns, 1991). In

this regard, Breyer (2011) mentions two ways to use corpus linguistics in teaching English: indirect and direct ways. Indirect ways include designing syllabi and textbooks via corpora providing authentic language. Moreover, it involves learner corpora consisting of learner language that can be used for contrastive interlanguage analysis and error analysis to improve teaching materials. Learner corpora may enable learners to understand their process of learning English (Pérez-Paredes & Mark, 2021). As for the direct way, it involves computer-assisted language learning and data-driven learning (Breyer, 2011).

As a direct way to teach English via corpus use, data-driven learning involving an inductive approach to learning and teaching English can be defined as an approach based on the discoveries of the learners while the teacher is providing the context to help the learner to develop strategies (Johns, 1991). On the other hand, the deductive way to apply data-driven learning via corpus involves a verification procedure of the learners of the previously learned rules (Breyer, 2011). The data-driven learning approach enables the learners to be more autonomous while they are being the detectives of the language (Cheng & Lam, 2021; Breyer, 2011).

In addition to DDL approach, one of the ways to teach English via corpus involves specialized corpora that can be used to teach English for Academic Purposes and English for Specific Purposes (Flowerdew, 2012). Through specialized corpus, a register-specific description can be made to grammar along with the integration of lexicon and grammar (Conrad, 2000). Thus, corpus research can provide materials for grammar books and monolingual dictionaries (Conrad, 2000). In this regard, teaching materials based on corpus research can be created by enabling the teacher to choose consciousness-raising activities (Conrad, 2000). Along with the activities that can be found through corpus, corpus-based tasks can be used for pedagogical purposes (Cheng & Lam, 2021). Overall, corpus in both learning and teaching English can be utilized through indirect (creating teaching materials and learner corpora) and direct (data-driven learning approach) ways.

2.7. Evidentiality and Modality

The definition of evidentiality can be discussed from a linguistic aspect. Basically, evidentiality in linguistics refers to the source of information (Aikhenvald, 2004). Another definition suggests that evidentiality deals with “the indication of the source or kind of evidence speakers have for their statements” (Auwera & Plungian, 1998: 85). Evidentiality is a “functional category that refers to the perceptual and/or epistemological basis for making a speech act” divided into two types of evidentiality: direct evidentiality (hearsay) and indirect evidentiality (reasoning) (Cornillie, 2009: 45). Whereas direct evidential can be used when the speaker has seen the action from the first hand, indirect evidential can be used if the speaker is not witnessing the action by using a deduction or hearing from the others about the action (Cornillie, 2009). Other than these

definitions involving the notion of justification, Boye (2012) regards the justification as epistemic justification; thus, evidential meanings can carry varying types of epistemic justification.

In this regard, evidentiality has different semantic parameters and types. Similar to Willett's (1988) classification, Boye (2012) categorizes the types of epistemic justifications conveyed with evidential meanings as direct justification, indirect-reportive justification, and indirect-inferential justification. As one of the classical categorizations of evidences, Willett's (1988) categorization consists of direct (experienced), indirect (inferential), indirect (hearsay). According to this categorization, direct evidentiality refers to firsthand experience of the speaker with perceptions. Indirect evidential is divided into two: inferential and hearsay. Inferential evidential involves a deduction and inference from direct information or background knowledge by using logical reasoning. The other type of inferential evidential is hearsay involving the speaker's knowledge of the facts through someone else (Nuyts, 2017). On the other hand, Aikhenvald (2004) differentiates between inference and assumption and hearsay and reported (quotative). According to Aikhenvald (2004), evidentiality can cover visual, non-visual sensory, inference, assumption, hearsay, and quotative.

2.8. Evidentiality in English

To express evidentiality, either lexical elements or grammaticalized expressions can be utilized in English, as in (Cornillie, 2009: 46):

- The author is *allegedly* a member of a comedy troupe and *presumably* was trying to be witty.
- It *seems* to be a good movie.

In these examples, both hearsay readings with lexical elements (*allegedly* and *presumably*) and inferential reading with grammaticalized expressions (*seem to be*) can be utilized to convey evidentiality in English.

Depending on the classical categorization of evidentiality, differentiation has been made among experienced (direct), inferential (indirect), and hearsay (indirect) by Willett (1988). The statements can change regarding the degree of reliability of the speaker according to the quality of evidence, as in (Nuyts, 2017: 67):

- I've *seen* it happen myself!
- John *appears* to be home. [His car is in front of the door and the lights are on.]
- I *hear* you've got a tuition waiver at Stanford, congratulations.

According to these examples, the first example is a statement conveying an experience; therefore, it is an experienced evidential (direct). Since experience evidential includes the use of perceptions, the verbs indicating the use of perceptions can be used to convey this meaning in English. The second example is an example of inferential evidential indicating an inference from a fact by using reasoning; thus, the verbs indicating inference can be used to give this meaning. The last one implicates hearsay evidential suggesting an act of hearing from someone else.

To convey epistemic meaning in English, the perceptive verbs can be used. Since evidentiality carries a deictic meaning, perceptive verbs used for evidentiality will also convey deictic meaning along with their perceptive meanings (Whitt, 2010: 26):

- I **see** the house.
- I **see** the house burning.
- Karen **looks** sick.

The first example only indicates the visual experience of the speaker perceiving the house. On the other hand, in the second example, there is an extra deictic component regarding the speaker's visual evidence on the burning house in addition to the fact that house is burning. The last example shows that it has two layers: visual evidence and the inference of the speaker based on this evidence although the speaker is not stated in the syntactic structure, and it is object-oriented.

Moreover, the evidentiality can be conveyed through perception verbs in English. In this regard, Whitt (2010) suggested the complementation patterns used to convey evidential meaning by using perception verbs (Table 6).

Table 6: Complementation Patterns for Perception Verbs Indicating Evidentiality

Complementation patterns	Examples
Perception verb + complement clause	I saw he was in front of the restaurant last night.
Perception verb + WH- complement clause	I saw where he was last night.
Perception verb + Direct object + Non-finite clause	I saw him coming from there.
Perception verb + prepositional phrase	It smells like a bad situation.
Perception verb + adjectival complement	He was looking excited.
Perception verb + conjunction + clause	It sounds as if it was the truth.
Perception verb + (Infinitive copula) + adjective/noun/ adjective/noun	She appears to be a good woman.
Parenthetical construction	She is doing charity to which I hear the old people are given groceries.
Perception verb external to the 'Evidentialized' Clause	She was in a hurry to catch the bus. While she was running to the bus station, handle of her bag was torn. So, she missed both the bus and the class. All this I heard.

Source: Whitt, 2010: 39

2.9. Evidentiality in Argumentative Discourse

Evidentiality can be traced also in argumentative discourse in terms of several aspects. Differentiating between argument and argumentation, Barton (1993) describes argument as a support of a claim whereas argumentation is a genre of an essay. In the creation of arguments, evidential markers with all functions can be used to make claims or counterclaims (Barton, 1993). Furthermore, it can be suggested maintaining pieces of evidence may be the key element of argumentative essay to support the arguments (Kibler & Hardigree, 2017). In addition, the markers of inter(subjectivity) can be used in argumentative discourse as a strategic tool for negotiating the positions of the counterparts (Nuyts, 2017). Evidentiality and argumentative discourse can be regarded as related because Hyland (2005b: 176) maintained the “writers intrude to stamp their personal authority onto their arguments or step back and disguise their involvement” under the term stance as an attitudinal dimension involving evidentiality as its part. Thus, it can be maintained evidentiality and argumentative discourses can be strongly related to each other with different aspects.

2.10. The Relationship between Epistemic Modality and Evidentiality

Although evidentiality and epistemic modality can be differentiated (Nuyts, 2001), the inclusion and overlap between them can bring about combining them in the studies (Cornillie, 2009). The inclusion of evidentiality and epistemic modality can stem from both the possibility of conveying an epistemic meaning via an evidential marker and the possibility of the dependence of epistemic strength on the inclusion of evidential dimensions (Cornillie, 2009). As for overlapping, it stems from the possibility of the correlation between epistemic necessity and inferential evidentiality, both of which “are dealing with the certainty of a judgment relative to other judgments” (Auwera & Plungian, 1998: 86).

Moreover, evidentiality covers epistemic modality because it is an umbrella term involving epistemic modality (Nuyts, 2017). Inferential includes the degrees of confidence and reliability of the speaker’s inferencing related to the strength of the evidence and quality of the evidence. In this regard, both inferential evidentiality and epistemic modality are in attitudinal categories in that “they concern the question of the extent to which the speaker can commit him or herself to a state of affairs” (Nuyts, 2017: 70). On the other hand, these two have several differences in their natures. One of the differences between them is that while inferential evidentiality deals with the reliability of what is stated, epistemic modality involves the likelihood of what is stated. Regarding this, the first involves reasoning in view of facts; however, epistemic modality refers to hypothetical thinking stemming from reasoning saying nothing about the quality about the evidence (Nuyts, 2017).

2.11. Native and Non-native Speakers in Applied Linguistics

The notion of native speaker has been a discussion among scholars. One of them was Davies (2003) discussing the native speakers in terms of psycholinguistic, linguistic, sociolinguistic, and communicative competence aspects along with lingualism and knowledges regarding native speakers, intelligibility, and lastly speech community. In this regard, Davies (2003) asks the question of who the native speaker is. As a first consideration to this question, identity judgements were maintained as whether native and non-native speakers make the same judgments or not Davies (2003). Consequently, Davies (2003) characterizes native speakers with six characteristics:

- Acquiring L1, which becomes the native language of the person, in childhood,
- Having intuitions concerning Grammar 1,
- Having intuitions regarding the distinctive features of Grammar 2 from Grammar 1,
- Producing spontaneous discourse fluently with communicative competence,
- Writing creatively,
- And interpreting and translating into the L1.

Other than these characteristics, Davies (2003) maintained the characteristics of L2 learners concerning six criteria: not acquiring the L2 in early childhood, having a possibility to have intuitions about Grammar 1 and 2, discourse and pragmatic control, creative performance, and interpreting and translating. Although these characteristics can be acquired, it should require sufficient contact with the target language and practice. When it comes to differences, confidence and identity can be maintained as the differences between them (Davies, 2003). Moreover, for an adult non-native, “to gain the speed and the certainty of knowledge relevant to judgements of grammaticality” can be regarded as more difficult (Davies, 2003: 2013). In this regard, applied linguistics can focus on the non-native speakers’ need and the meaning of International English concerning whether it means a variety of English or the use of English in international settings (Davies, 2003).

Based on this, there are several studies comparing the use of epistemic modality of natives and non-natives of English in the literature (Karahan, 2017; Ngula, 2005; McEnery and Kifle, 2002). In the light of these studies, it can be suggested that there can be differences between the non-natives and natives concerning the use of epistemic modality in different frequencies, strengths, grammatical structures, and the overall usage. Thus, this study focused on the comparative interlanguage analysis of non-natives and native speakers of English through a corpus-based study investigating the use of epistemic modal auxiliary and lexical verbs so as to see the similarities and differences between L2 learners with different L1 background and native speakers of English.

CHAPTER THREE

3. METHODOLOGY

3.1. Introduction

This corpus-based study involves both quantitative and qualitative methods investigating the use of epistemic modal auxiliaries and lexical verbs in the argumentative essays of native undergraduate students and non-native undergraduate students involving Turkish, Norwegian, and Japanese. This section includes the rationale for the study, selection criteria for and recordings of epistemic devices, aspects of epistemic devices, adopted criteria for epistemic classification, participants, corpora, data collection and analysis, procedure, preparation of the epistemic devices, classification of epistemic devices concerning their epistemic strengths, the extraction and examination of epistemic devices, and normalized frequencies. Last but not least, the study of Karahan (2017) and Ngula (2015) investigating the use of epistemic devices by native speakers of English and non-native speakers of English inspired this study.

3.2. Rationale for the Study

This study investigated the use of epistemic modal auxiliaries and epistemic lexical verbs to take a deep look into the epistemic modality in the argumentative essays of Norwegian, Japanese, Turkish, and English undergraduate students. The epistemic use of modal auxiliaries and lexical verbs has been presented as a challenge to make use of it in argumentative essays by non-native English learners (Hyland & Milton, 1997). Epistemic modality has been a useful rhetorical aspect to make an argumentation and to express the position of the writer by indicating writer's certainty to his/her proposition (Ngula, 2015). Epistemic modality markers are commonly used in the argumentative essays of the students, especially when they are making an argumentation and positioning themselves regarding their propositions (Ngula, 2015). In this regard, epistemic modality markers in argumentative essays involve epistemic modal auxiliaries and lexical verbs. Nevertheless, modal auxiliaries and lexical verbs can be used both in their epistemic meanings and non-epistemic (deontic) meanings (Karahan, 2017). For this reason, using them with their epistemic meanings intentionally to make arguments requires rhetorical awareness in the target language; thus, it may create a challenge and show different use of epistemic modal auxiliaries and lexical verbs across the languages. In addition, modal auxiliaries can be used to help discursive use of language conveying the subjects' intention in such a way that it explains the rationale having

different cultural values; therefore, it may shed light on the cultural differences based on how people's expressions of their intentions discursively (Roberts et al., 2010).

Moreover, languages coming from different language family trees can utilize epistemic modality by using epistemic modal auxiliaries and epistemic lexical verbs differently. Regarding this, the use of epistemic modality by Norwegian, Japanese, Turkish, and English undergraduate students can show differences and challenges. To see the differences, similarities, and challenges between the languages regarding epistemic modality, this study examined the use of epistemic modal auxiliaries and epistemic lexical verbs in the argumentative essays of Norwegian, Japanese, Turkish, and English undergraduate students.

3.3. Selection Criteria for Epistemic Modality

The selection of both epistemic modal auxiliaries and epistemic lexical verbs was based on Ngula's (2015) criteria and Karahan (2017) criteria. In this regard, the list of epistemic modal auxiliaries and epistemic lexical verbs from Ngula (2015) and Karahan (2017) were taken as a basis to investigate epistemic modality in this study.

3.4. Recording of Epistemic Devices

Epistemic devices can also be polysemous, which means they can have two meanings. Some of them can have both epistemic and non-epistemic (deontic) meanings (Vold, 2006). In this regard, the epistemic modal auxiliaries and lexical verbs needed to have been distinguished from the non-epistemic uses. To this end, a set of criteria was created and followed by the researcher by drawing conclusions from similar studies (Coates, 2015; Karahan, 2017; Ngula, 2015; Quirk et al., 1985; Hyland & Milton, 1997).

To detect epistemic uses, a concordance tool called SketchEngine was used. The concordance function of SketchEngine was utilized to see the larger contexts of each item. By doing that, the researcher managed to see the larger contexts of each item to detect whether they were used in their epistemic meanings.

3.5. Development of the Criteria for Epistemic Classification

There are several concerns about the classification of epistemic modal auxiliaries and epistemic lexical verbs. The criteria for their classifications in this study were developed by looking at the literature review about epistemic classifications.

3.5.1. Some Aspects regarding Modal Verbs

As Palmer (2003) classifies, there are three main types of classification of modality: epistemic modality, deontic modality, and dynamic modality. Among them, modal auxiliaries can carry both epistemic meaning and root meanings. Epistemic modality is used to express the attitude of the speaker to the situation by making judgements, as in (Palmer, 2003: 7):

- They may be in the office. -They must be in the office.

In the example, it can be suggested the speaker is making a judgement about the possibility that they are in the office. On the other hand, one of the root meanings, deontic modality, is utilized when the situation is external, which means it is not related to the subjects' judgement, but to the external situations. As in the example given by Palmer (2003: 7):

- They may/can come in now. -They must come in now.

It is related to an external circumstance. These circumstances concern an obligation or permission expressed by deontic modality (Palmer, 2003). As for the dynamic modality, it is used to express ability or volition (Palmer, 2003). Unlike deontic modality, dynamic modality is related to internal situations as in (Palmer, 2003: 7):

- They can run very fast. -I will help you.

The modal auxiliary *can* shows an ability and the modal auxiliary *will* presents a willingness. Among these examples, the modal auxiliaries *must* and *may* are used both in epistemic and deontic meanings. The modal auxiliary *can* is used deontic and dynamic meanings. Regarding this, it can be concluded the meanings of modal auxiliaries can carry epistemic or non-epistemic modalities.

In the case of modal auxiliary *may*, the epistemic meaning indicating possibility and deontic meaning indicating permission can be both used. Guziurová (2016: 123) gives examples of epistemic and deontic meanings of modal auxiliary *may*, as in:

- Students **may** steal books from the library.
- Students **may** borrow books from the library if they have a student card.

In the first example, the modal auxiliary *may* is used in epistemic meaning making a judgement; on the other hand, the second example shows the modal auxiliary *may* is used in deontic meaning expressing a permission in the situation.

As for the negation of modal auxiliaries, there are modal auxiliaries whose meaning can change when they are in their negation forms. The problem is about expressing 'not possible' and 'possible not', both of which have different meanings. Since English language does not contain such sentence as 'possible not', different modal auxiliaries can be used to give this meaning, as in (Palmer, 2003: 9):

- They can't be in the office. ("not possible")
- They may not be in the office. ("possible not").

In these examples, the negation of *can* does not give the epistemic meaning. Instead, the negation of *may* gives the epistemic meaning 'possible not'. The same is also for the negation of modal auxiliary *must* concerning 'not necessary' and 'necessary not' (Palmer, 2003: 9):

- They must be in the office.
- They may not be in the office. ("possible not" for "not necessary")
- They can't be in the office. ("not possible" for "necessary not")

In this regard, since the epistemic meaning "possible not" cannot be given with the negation of *must*, which is *must not*, the modal auxiliary *may not* is used to give the negation of possibility. The reason why the modal auxiliary *cannot* is not used is because it does not give epistemic 'possible not', but gives the meaning 'not possible'. Therefore, the other modal auxiliaries can be used to give the negated meaning (Coates, 2015).

As for the modal auxiliary *should*, it can mean the deontic necessity along with the epistemic judgment, as in (Palmer, 2003: 11):

- You shouldn't/oughtn't to come in.

In the example, the speaker is giving a piece of advice to the person to whom he is speaking. This sentence gives the deontic meaning by giving a piece of advice (Palmer, 2003). To make the modal auxiliary *should* epistemic, it should express a judgement since the epistemic modality is about making a judgement (Palmer, 2003).

Moreover, the modal auxiliary *will* can give both epistemic and non-epistemic meanings. Hyland (1998) signifies the difficulty of differentiating between the epistemic meaning of *will* and the non-epistemic meaning of *will* involving a future action. Therefore, it can be suggested it has less potential to carry an epistemic meaning; however, using it with epistemic adverbs can be regarded as a hedge (Guziurová, 2016). For these polysemous modal auxiliaries, the strategies that

Vold (2006) used for the classification of epistemic and non-epistemic meanings were utilized. These strategies consist of a substitution test by either changing the polysemic marker with an epistemic marker or adding a phrase expressing uncertainty Vold (2006). Also, changing the syntax of the sentence is one of the strategies to understand whether the marker is epistemic or not Vold (2006).

3.5.2. Some Aspects regarding Lexical Verbs

For lexical verbs, they can also carry both epistemic and non-epistemic meanings (Ericsson, 2008). Since the epistemic markers modify propositional content, the lexical verbs modifying propositional content are epistemic whereas the ones that are not modifying any propositional content are not (Vold, 2006). Thus, a lexical verb can carry non-epistemic meaning, as in (Vold, 2006: 66):

- ...explicit content is ‘much more inferential and hence worthy of pragmatic investigation’ than is **assumed** by pragmatists in the Gricean tradition.

The lexical verb *assume* is not used in epistemic meaning, but deontic meaning because it does not modify any propositional content in this example. On the other hand, lexical verbs can also carry epistemic meanings, as in (Vold, 2006: 66):

- Our findings in non-smoking mothers **suggest** that the positive associations previously reported among first births **might** simply reflect inadequate adjustment for confounding variables.

In this example, both a modal auxiliary and a lexical verb are used. The lexical verb *suggest* carries an epistemic meaning because it modifies propositional content. In that sense, lexical verbs can change regarding their epistemic or non-epistemic meanings. For Hyland (1998), the epistemic lexical verbs are the ones that are used to mark the confidence of the writer in the knowledge and the source. Similarly, the lexical verb *appear* can carry both epistemic meaning and non-epistemic meaning (Ericsson, 2008: 8):

- It *appears* that all students passed the exam.
- Suddenly he *appeared* on the stage.

In this example, it can be suggested the first sentence carries an epistemic meaning while the other sentence carries a non-epistemic meaning. The first sentence is used with epistemic lexical verb to express a judgement about the students and whether they have passed the exam or not.

However, the other sentence uses lexical verb *appear* to express a situation about a person and whether he showed up on the stage or not.

Furthermore, another concern is that there are two viewpoints on the classification of epistemic lexical verbs. One of them was suggested by Crompton (1997: 283) maintaining that the subject should be responsible for the reported propositional content in order the lexical verb to carry epistemic meaning:

- Shmerling suggested that sensitization took place during the first hours after birth.

In this example, the subject is not reporting his/her proposition; thus, the lexical verb *suggest* does not carry any epistemic meaning. The lexical verb should be used to express the writer's own responsibility and interpretation (Crompton, 1997: 286):

- Previous estimates of the incidence of acute mountain sickness *suggest* that...

On the other hand, Hyland and Milton (1997) add also these sentences into the category of epistemic uses of lexical verbs suggesting that it is a way to hide the epistemic source by attributing the author's view to another subject. In this regard, this study adopted the view of Hyland and Milton (1997) to include also the sentences created by the people using the view of others to hide the epistemic source and expressing them with lexical verbs.

3.6. Adopted Criteria for Epistemic Classification

This study adopted the criteria applied in several previous studies. According to the criteria they have used, this study created midway for the criteria for epistemic classification. One of the studies whose criteria was adopted in the study conducted by Vold (2006). According to Vold (2006: 65), epistemic modality expresses the degree of relying on the information that is expressed; thus, the epistemic modal auxiliary and lexical verbs should be a lexical unit or a grammatical one explicitly expressing "the truth value of a certain propositional content". Furthermore, Palmer's (2013) definitions and examples were taken into consideration in this study. He (2013) differentiates between 'possible not' and 'not possible' and 'necessary not' and 'not necessary'. In addition, the negations of modal auxiliaries mentioned by Palmer (2003) were taken into consideration for classifications.

Along with these criteria, other criteria mentioned in Guziurová (2016) were adopted in this study. According to Guziurová (2016), epistemic modality expresses possibility indicating degrees

of uncertainty. Moreover, Guziurová (2016) mentions the classification for epistemic modal auxiliaries adopted in this study.

Apart from these criteria, a list for the epistemic classification was given by Ericsson (2008). This list is also adopted for the criteria adopted in this study. According to the list, the markers carrying a meaning other than epistemic were excluded while markers that can be read as with an epistemic meaning even if it is ambiguous were included. For lexical verbs, the sentences expressing an attitude of the writer both implicitly and explicitly were added. However, even though the list does not add the modality markers in the form of question to the list as the epistemic classification, Ericsson (2008) mentions the question forms can also convey epistemic meaning. Thus, in this study, the question forms carrying epistemic meaning were added. Lastly, the viewpoints of Hyland and Milton (1997), Quirk et al. (1985), Ngula (2015), Karahan (2017), and Coates (2015) were taken into consideration for the criteria.

3.7. Participants

The corpora used in this study were chosen from Turkish, Norwegian, Japanese, and English undergraduate students. These students' argumentative essays were analyzed. The participants contributed to the Norwegian subcorpus, called NICLE, from nine different universities in Norway. These universities are "(1) the American College of Norway, (2) Bergen College, (3) Hedmark College, (4) Oslo College, (5) Ostfold college, (6) University of Oslo, (7) Agder College, (8) Buskerud University and (9) University of Bergen" (Granger et al., 2020: 47). The average age of the Norwegian participants is 23.94 ranging from 18 to 55. The participants were 74.13% female while 25.87% of participants were male in Norwegian corpus. 97.48% of the participants in NICLE have been learning English 6 years or more whereas the 0.32 of the participants have been learning English for 2 years. Looking at the 20 essays from NICLE, the CEFR results showed 8 students are in B2 level, 7 students are in C1 level, and 5 students are in C2 level. The 98.42 of the essays in Norwegian corpus were argumentative.

In Japanese corpus, called JPICLE, the participants contributed from 21 different universities in Japan (Granger et al., 2020). The average age of the Japanese participants is 20.06 ranging from 18 to 36. The participants in JPICLE were 72.95% female while the 26.50% of participants were male. All participants in JPICLE have been learning English 6 years or more. For the CEFR level of participants in JPICLE, according to 20 essays in JPICLE, 18 of the students are in B2 level whereas 2 of the students are in C1 level. All the essays in JPICLE gathered from the participants were argumentative.

For Turkish participants, the students participated to the study from 3 different universities. The average age of the Turkish participants is 22.08 ranging from 19 to 38. The participants in

Turkish corpus, called TRICLE, were 81.43% female while the 18.57% of participants were male. The 97.50% of the participants in TRICLE have been learning English 6 years or more whereas the 0.36% of the participants have been learning English for 3 years and 2.14 of the participants have been learning English for 4 years. According to the CEFR levels of the students, 16 of the students are in B1 level out of 20 students while 4 of the students are in C1 level. All the essays in TRICLE gathered from the participants were argumentative (Granger et al., 2020).

As for LOCNESS, the students were British and American undergraduate students and participated to the study from the universities in Britain and America. The essays written by native speakers of English were gathered by Louvain University in Belgium (Granger, 1998b).

3.8. Corpora

The NICLE consisted of 317 essays of the undergraduate students. It has 213,701 words in total. However, in this study, only argumentative essays were included. Argumentative essays comprised 312 essays with 208,101 words. 97.79% of the essays were collected under untimed conditions. Regarding the exam condition, 99.37% of essays were written in a non-examination environment while 0.63% of the essays were not known. The average number of words written in these essays was 674.14 words. The essays written by using reference tools were 63.09%; on the other hand, 36.91% of the essays were written with no support. In addition, the minimum length of these essays was 321 words whereas the maximum length is 17.82 (Granger et al., 2020).

JPICLE consisted of 366 argumentative essays of undergraduate students with 200,958 words. In JPICLE, the average length of the essays was 549.07 words. Moreover, the minimum length of the essays was 399 words while the maximum length was 1009 words. 37.43% of the essays were collected under untimed conditions whereas the essays written under timed conditions were 56.01%. Moreover, 36.89% of the essays were collected via an examination whereas 59.02% of them were collected with non-examination. On the other hand, 6.56% of the essays were not known whether they were written under timed conditions or not. The essays written by using reference tools were 13.11% whereas 81.97% of the essays were written with no support (Granger et al., 2020).

TRICLE consisted of 280 argumentative essays with 200,601 words. In TRICLE, the average length of the essays was 716.43 words. Moreover, the minimum length of the essays was 500 words while the maximum length was 1423 words. All essays were collected under untimed and non-examination conditions. 52.14% of the essays were written by using reference tools while 47.86% of the essays were written with no support (Granger et al., 2020).

LOCNESS can be directly compared to ICLE; however, it has 324,304 words constituted by the essays of students, which can be a disadvantage regarding not to include professional writings (Granger, 1998b). These essays were constituted of argumentative and literary-mixed involving 322 essays. Overall, these four corpora were found to be comparable according to the types and number of essays, size of the corpus, average essay length, average age of students, study level, and topics of essays although the completion time of corpora were found as having low comparability (Table 7).

Table 7. Overall Comparability of Four Corpora

	Corpus				Comparability*
	LOCNESS	TRICLE	NICLE	JPICLE	
Type of essay	Argumentative and Literary-mixed	Argumentative	Argumentative	Argumentative	++++
Number of essays	322	280	317	366	++
Size of the corpus (words)	324,304	200,601	213,701	200,958	+++
Average essay length	=/+ 500 (tokens)	716.43 (in words)	674.14 (in words)	549.07 (in words)	++
The average age of students	17-22	22	23	20	++++
Study level	University level (A-level students)	University level (third to above years)	University level (first to the third year)	University level (first to fifth year)	++
Topics of essays	French intellectual tradition Loss of sovereignty for Britain Abortion Euthanasia Capital punishment Nuclear power Homosexuality Sex equality Crimes Feminists Divorce	Divorce Abortion Sex Equality Nuclear Energy Suicide Capital Punishment Money Inventions Euthanasia Freedom of Press Human Rights Animal Testing University Degrees	Natural equality Feminism Professional military Science and technology Military army Crimes University degree Equality of men Dreaming and imagination Prison system Abortion	Language acquisition English as a second language American people	+++
Completion time	1998	2016	2002	2005	+
* +: low; ++: between low and average; +++: between average and high; ++++: high					

Therefore, LOCNESS was found as the most suitable for comparison with the ICLE corpora among the other native corpora. Regarding this, the fact that LOCNESS was used as a reference corpus in this study does not mean that it has all the correct usages of epistemic modal auxiliaries and lexical verbs. However, it can be seen as a way to see the full picture of how TRICLE, JPICLE, and NICLE used epistemic modal auxiliaries and lexical verbs when compared to the reference corpus.

3.9. Data Collection

The data were collected from three sub-corpora: NICLE, JPICLE, and TRICLE. These corpora were reached via ICLE corpus. In addition, LOCNESS was used because the corpus consisted of native students' essays. While collecting the data, the website of ICLE was utilized. From the website, the data was downloaded and uploaded to the corpus software tool SketchEngine to analyze them. For the ICLE corpora, the criteria were arranged according to the nationality, native language, and task variables. Regarding this, the criteria were arranged as Norwegian as the nationality, Norwegian as native language, and argumentative essay as the task variable. The same criteria were applied also for Turkish and Japanese corpora. After the first collection of the epistemic uses, another rater did the same procedure. Then, an inter-rater reliability test was applied to decide the final version of the epistemic uses. As for the LOCNESS, the data from LOCNESS corpus were uploaded to the SketchEngine to analyze it as a reference corpus.

3.10. Data Analysis

The data were analyzed via the corpus software tool called SketchEngine. By using SketchEngine, several functions can be reached to analyze the data (Hunston, 2022:10). In this study, the concordance function of SketchEngine was utilized most. This function enables the researcher to see the context in which the word is used and to find out why, how, and where a collocate occurred (Kilgarriff et al., 2014; Flowerdew, 1996; Sinclair, 1991). Along with this, the corpus software tool showed frequency, dispersion, collocation, and keywords concordances (Anthony, 2013). On the other hand, in this study, the concordance function was utilized to see the context and frequency of the word.

In this regard, the corpora that were uploaded to the corpus software program called SketchEngine were analyzed by using the concordance function. The concordance function enabled the researcher to see the linguistic features and whether the modal auxiliaries and lexical verbs were used as epistemic or not. It is decided by analyzing the word's context. In this way, the researcher was able to see the frequency of epistemic uses of the keyword. Thus, the corpora were uploaded to SketchEngine. After deciding the list of epistemic devices, the keywords were searched via the concordance function to analyze the context in which the keyword was used. Then, the epistemic uses of the keywords and their frequencies were recorded.

For the analysis of epistemic strength, a log-likelihood test was conducted via a calculator by Paul Rayson at Lancaster University (available at <http://ucrel.lancs.ac.uk/llwizard.html>) to see whether there is a difference between the corpora. The reason for the use of log-likelihood test is that it enables the researcher to compare the frequencies in two corpora by showing whether the difference in the results is based on a random or significant association (Ngula, 2015). Although

log-likelihood can seem similar to chi-square test, log-likelihood test is used when the frequencies are smaller, and it is easier to interpret the results of log-likelihood test than that of chi-square because of not needing to use critical values' table (Desagulier, 2017). Thus, a log-likelihood test was found more applicable for this study.

3.11. Corpus Tool Used in This Study

There are several corpus software programs that can be used for corpus-based studies. Hunston (2022) lists the software programs that are mostly used in the studies as Antconc, English-Corpora.org, #Lancsbox, SketchEngine, Wmatrix, and Wordsmith Tools. In this study, among these software programs, SketchEngine was used. SketchEngine is an online software program developed by Adam Kilgarriff (Hunston, 2022). Kilgarriff (2005) describes language as non-random because we produce language with purposes and may not produce them without computational aid. It has several functions including concordance, n-grams, Word Sketch, and Wordlist (Hunston, 2022). Among them, n-grams can be described as “strings of words”, which can include two words, three words, or more (Hunston, 2022: 7). In addition, the concordance function that is used in this study to see the context of a word (Flowerdew, 1996; Sinclair, 1991). In SketchEngine, the concordance function presents the raw data to provide information regarding why, how, and where a collocate occurred (Kilgarriff et al., 2014). These corpus software programs help to analyze large amounts of texts by giving information about frequency, dispersion, collocations, keywords, and concordances (Anthony, 2013). Moreover, there are ready-made corpora that can be used in SketchEngine; or, the new corpora can be uploaded by the users to SketchEngine (Kilgarriff et al., 2014).

3.12. Procedure

After deciding the corpora used in this study, the epistemic modal auxiliaries and lexical verbs were chosen, and a list was prepared. Then, the epistemic strength of the EMAs and ELVs were decided. Each EMAs and ELVs were extracted by using concordance function of SketchEngine. The concordance lines were examined to see whether the modal auxiliaries and lexical verbs were used as epistemic or not. Moreover, an inter-rater reliability test was applied to find out to what extent the first rater's examination matched with the second rater's examination. Lastly, raw and normalized frequencies of EMAs and ELVs were found out.

3.12.1. Preparation of the List of Epistemic Devices

The list of epistemic modal auxiliaries was adapted from the study of Ngula (2015). According to the list of epistemic modal auxiliaries, 7 epistemic modal auxiliaries were used in this study: could (not), may (not), might (not), must (not), should (not), would (not), and will (not).

Although Ngula (2015) included the negation of modal auxiliaries as separate items, the modal auxiliaries were investigated along with their negations in this study. When epistemic devices were searched in SketchEngine, the negation of the epistemic devices were also included in the frequency result. Along with the negations of epistemic devices, the reductions of both epistemic devices and negation of these epistemic devices, such as epistemic modal auxiliary ‘*ll* and negation *n’t*, was included in the frequency results.

Moreover, the modal auxiliary *can* was not included to the list because the previous seminal studies did not identify and include the modal auxiliary *can* as an epistemic marker in their studies (Ngula, 2015; Hyland & Milton, 1997; Coates, 1983). On the other hand, since the list of epistemic lexical verbs created by Ngula (2015) may not be inclusive, the list with 49 epistemic lexical verbs adapted from the study of Karahan (2017) was utilized. Although the original list of epistemic lexical verbs created by Karahan (2017) includes 52 epistemic lexical verbs, 49 of them were used in this study. In Karahan’s (2017) list, the epistemic lexical verbs *know*, *seem*, and *guess* were written twice; therefore, it made the list as 49 epistemic lexical verbs in this study (Table 8).

Table 8: The List of Epistemic Modal Auxiliaries and Lexical Verbs

Epistemic Modal Auxiliaries and Lexical Verbs				
Appear	Confirm	Guess	Judge	Presume
Argue	Could (not)	Highlight	Point out	Propose
Attest	Demonstrate	Hope	Prove	Reckon
Assume	Deduce	Know(n)	Realize	Regard
Assure	Describe (as)	Look as if	Recognize	Suggest
Believe	Doubt	Look (like)	Reveal	Suspect
Claim	Emphasize	May (not)	Seem	Suppose
Conclude	Establish	Might (not)	Show	Tend
Consider	Estimate	Must (not)	Should (not)	Think
Convince	Expect	Identify	Speculate	Will (not)
	Feel	Indicate	Stress	Would (not)
		Infer		
		Interpret		

Source: Karahan, 2017: 52

3.12.2. Classification of Epistemic Devices Regarding Their Epistemic Strengths

There are several classifications of epistemic devices according to the epistemic commitment they express. One of the well-known classifications made by Hyland (1998) is hedges and boosters. According to Hyland (1998), hedges are used to avoid expressing a commitment to what is said by changing the speaker’s confidence. Along with the boosters, hedges are used to categorize the markers along a continuum “ranging from uncertain possibility to confident necessity” (Hyland, 1998: 2). In addition, Biber et al. (1999) suggested a continuum ranging from doubt to certainty

along with the intensifiers and downtoners. One of the three-way classifications is suggested by Halliday et al. (2004) maintaining the modality's three values: high, median, and low. Furthermore, Huddleston and Pullum (2017) suggested a classification with four categories by adding a fourth category called quasi-strong to strong, medium, and weak. In addition, Rubin (2010) maintained a classification with five categories by adding absolute certainty and uncertainty to high certainty, moderate certainty, and low certainty. Lastly, mentioning epistemic support, Boye (2012) suggests full support involving certainty or epistemic impossibility, partial support involving probability and doubt, and neutral support involving uncertainty.

Other than these classifications, Ngula's (2015) three-way classification was utilized in this study. This list was created by looking at the epistemic markers' probability degrees. As for the other classifications, Ngula's (2015) classification can be more applicable to this study because the epistemic modal auxiliaries and lexical verbs were investigated. The other classifications may remain less neat, complicated, and unnecessary. Two-way classification might not allow the researcher to classify epistemic modal auxiliaries and lexical verbs in detail. Using four and five categories to classify would be unnecessary because it may not be effective with epistemic modal auxiliaries and lexical verbs. Thus, three categories *strong*, *medium*, and *weak* were used to classify epistemic modal auxiliaries and lexical verbs. Although all epistemic modal auxiliaries used in this study were classified in this list, some of the epistemic lexical verbs used in this study were indicated above (Table 9).

Table 9: Epistemic Strengths Expressed by Epistemic Modal Auxiliaries and Lexical Verbs

The Epistemic Strengths Expressed by Epistemic Modal Auxiliaries and Lexical Verbs		
Strong	Medium	Weak
Assure Attest Claim Convince Know Must Show Will (not)	Argue Assume Believe Consider Estimate Expect Feel Indicate Infer Presume Reckon Should (not) Suppose Tend Think Would (not)	Appear Could (not) Doubt Guess Hope Look (as if) Look (like) May Might Propose Seem Speculate Suggest

Source: Ngula, 2015: 119

3.13. The Extraction and Examination of Epistemic Devices in Four Corpora

For the extraction of epistemic lexical verbs and modal auxiliaries from LOCNESS, NICLE, JICLE, and TRICLE, the concordance function of SketchEngine was utilized. The lexical verbs and modal auxiliaries were identified regarding whether they carry epistemic meanings or not in their contexts via the concordance function. By scrutinizing each concordance line, the researcher discriminated the ones with the epistemic meanings from the non-epistemic ones according to the criteria created from the seminal works. However, to make this extraction more reliable, another rater specializing in linguistics analyzed the same concordance lines as the first rater, i.e. the researcher. A statistical method called Cohen's Kappa was utilized by using IBM SPSS 27 to measure inter-rater reliability. Cohen's kappa coefficient (κ), which is frequently used in determining reliability between raters, was proposed by Cohen (1960) and developed to determine the degree of agreement between two raters scoring at the classification level. κ is calculated with a formula based on subtracting the agreement based on chance/chance from the observed agreement between raters. (Bikmaz Bilgen & Doğan, 2017).

Cohen's kappa coefficient of each lexical verb and modal verb from each corpus was measured separately to ensure reliability. In this regard, the inter-rater reliability test of lexical verbs in LOCNESS shows that there were low agreements between the first and the second rater regarding five ELVs (Table 10). From none to slight agreement to fair agreement, the lexical verbs *judge*, *argue*, *point out*, *think*, and *highlight* can be ordered sequentially. The lexical verbs *judge* and *argue* are the verbs that can be interpreted as none to slight agreement in LOCNESS. Although the rest can be categorized as fair agreement, the agreements between the first and the second rater may be still low.

Table 10: Inter-rater Reliability Test for ELVs Resulting Low Agreement in LOCNESS

LOCNESS	Cohen's Kappa (κ) value	p
Argue	,187	<,001
Highlight	,385	,168
Judge	,083	,115
Point out	,245	<,001
Think	,273	<,001

As for the rest of the lexical verbs, they can be interpreted as moderate, substantial, and almost perfect agreement (Appendix 1). The only verb that resulted in moderate agreement was found as *assume*. The verbs found as substantial were the verbs *believe*, *doubt*, *expect*, *interpret*, *know*, *propose*, *prove*, *realize*, *suggest*, and *suspect*. Lastly, the lexical verbs resulted in having almost perfect agreement are the verbs *appear*, *assure*, *claim*, *conclude*, *confirm*, *consider*, *convince*, *demonstrate*, *describe*, *emphasize*, *establish*, *estimate*, *feel*, *guess*, *hope*, *identify*,

indicate, infer, look (like), look as if, presume, reckon, recognize, regard, reveal, seem, show, speculate. In conclusion, it can be maintained the agreement between the first and the second rater was found as moderate and high for these lexical verbs.

Similarly, the results of the inter-rater reliability test for the lexical verbs in NICLE were obtained (Table 11). It was concluded that there is only one lexical verb found to have a low agreement between the first and the second rater. Although there is not any verb interpreted as none to slight agreement, the verb *seem* was concluded as fair agreement, which can be interpreted as rather low agreement.

Table 11: Inter-rater Reliability Test for ELVs Resulting Low Agreement in NICLE

NICLE	Cohen's Kappa (κ) value	p
Seem	,396	<,001

Since there is only one verb having a low agreement, the rest of the verbs were found as either moderate or high agreement (Appendix 2). The verb *argue* was concluded as having moderate agreement. For substantial agreement, the verbs *conclude, emphasize, and know* were identified. The rest of the verbs *appear, assume, assure, believe, claim, confirm, consider, convince, demonstrate, describe, doubt, establish, expect, feel, guess, highlight, hope, identify, indicate, interpret, judge, look (like), point out, presume, propose, reckon, recognize, prove, realize, think, regard, reveal, show, speculate, stress, suggest, suppose, and tend* were found as almost perfect agreement. Regarding these results, only one verb *seem* was found to be in fair agreement, and the rest of the verbs were concluded as having moderate, substantial, and almost perfect agreement.

Moreover, an inter-rater reliability test was applied to the lexical verbs in JPICLE (Table 12). It was revealed the only lexical verb having low agreement was identified as the verb *think*. It can be categorized as having none to slight agreement. However, no verb was found as fair agreement in JPICLE.

Table 12: Inter-rater Reliability Test for ELVs Resulting Low Agreement in JPICLE

JPICLE	Cohen's Kappa (κ) value	p
Think	,402	<,001

Although one verb was concluded as none to slight agreement, the rest of the verbs were identified as having moderate and high agreement (Appendix 3). For the verbs having moderate agreement, it was found as the verb *confirm*. The lexical verbs *appear, argue, believe, conclude, expect, feel, guess, and judge* were found substantial agreement. Lastly, the verbs *assume, claim, consider, convince, describe, doubt, emphasize, establish, estimate, hope, identify, interpret, know,*

look (like), point out, presume, prove, realize, reckon, regard, reveal, seem, stress, suggest, suppose, suspect, and tend were concluded as having almost perfect agreement.

As for the inter-rater reliability test applied to the lexical verbs in TRICLE, it was revealed that the verb *think* was found as having none to slight agreement. Furthermore, the verbs *argue*, and *judge* resulted in having a fair agreement (Table 13). In this regard, it may be suggested that these ELVs had lower inter-rater reliability results when compared to other epistemic lexical verbs used in TRICLE.

Table 13: Inter-rater Reliability Test for ELVs Resulting Low Agreement in TRICLE

TRICLE	Cohen's Kappa (κ) value	p
Argue	,395	,002
Judge	,406	,028
Think	,190	<,001

Nevertheless, the inter-rater reliability test for other verbs was in high agreement (Appendix 4). For the substantial agreement, the verbs *guess* and *indicate* were identified. Lastly, the lexical verbs *appear, assume, believe, claim, conclude, consider, convince, demonstrate, describe, doubt, establish, estimate, expect, feel, highlight, hope, infer, know, look (like), point out, propose, prove, realize, recognize, regard, reveal, seem, speculate, stress, suggest, suppose, and tend* were found as almost perfect agreement.

Along with the inter-rater reliability tests for ELVs, the inter-rater reliability tests were also applied to the epistemic modal auxiliaries to test the degree of agreement between two raters. Regarding the results, there was not found low agreement in any of the corpora regarding epistemic modal auxiliaries. In LOCNESS, it was found the epistemic modal auxiliaries used in this study were agreed upon to a high degree (Table 14).

Table 14: Inter-rater Reliability Test for Epistemic Modal Auxiliaries in LOCNESS

LOCNESS	Cohen's Kappa (κ) value	p
Could	1	<,001
May	1	<,001
Might	1	<,001
Must	,840	<,001
Should	,806	<,001
Will	,791	<,001
Would	,981	<,001

According to the results, the agreement on epistemic *will* was found as substantial ($\kappa = ,791$). The degree of agreement between the two raters on the other epistemic modal auxiliaries *should* (κ

=,806), *must* (κ =,840), *would* (κ =,981), *could* (κ =1), *may* (κ =1), and *might* (κ =1) was found as almost perfect agreement. Therefore, it means both raters agreed on the EMAs with a high degree in LOCNESS.

As for the inter-rater reliability test for epistemic modal auxiliaries in NICLE, it was found that all EMAs used in this study were agreed on with a high degree. The inter-rater reliability tests of epistemic modal auxiliaries *should* (κ =,867), *will* (κ =,896), *must* (κ =,937), *would* (κ =,983), *could* (κ =1), *may* (κ =1), and *might* (κ =1) were found as almost perfect agreement between the two raters (Table 15).

Table 15: Inter-rater Reliability Test for Epistemic Modal Auxiliaries in NICLE

NICLE	Cohen's Kappa (κ) value	p
Could	1	<,001
May	1	<,001
Might	1	<,001
Must	,937	<,001
Should	,867	<,001
Will	,896	<,001
Would	,983	<,001

When it comes to the results of inter-rater reliability test for the epistemic modal auxiliaries used in JPICLE, it was revealed the degree of agreement between the two raters on EMAs was found as high. The inter-rater reliability results of epistemic modal auxiliaries *must* (κ =,877), *may* (κ =,887), *will* (κ =,915), *should* (κ =,936), *might* (κ =1), *would* (κ =1), and *could* (κ =1) were found as the almost perfect agreement between the two raters (Table 16). In this regard, it means the agreement degree on all EMAs in JPICLE was found as high agreement between the two raters.

Table 16: Inter-rater Reliability Test for Epistemic Modal Auxiliaries in JPICLE

JPICLE	Cohen's Kappa (κ) value	p
Could	1	<,001
May	,887	<,001
Might	1	<,001
Must	,877	<,001
Should	,936	<,001
Will	,915	<,001
Would	1	<,001

Lastly, the inter-rater reliability test for epistemic modal auxiliaries in TRICLE was applied. According to the results, it was revealed the degree of agreement between the two raters on the EMAs in TRICLE was found to be high. The inter-rater reliability test on the epistemic modal

auxiliaries *will* ($\kappa = .871$), *must* ($\kappa = .887$), *should* ($\kappa = .932$), *would* ($\kappa = .954$), *could* ($\kappa = 1$), *may* ($\kappa = 1$), and *might* ($\kappa = 1$) was found as the almost perfect agreement between the two raters (Table 17).

Table 17: Inter-rater Reliability Test for Epistemic Modal Auxiliaries in TRICLE

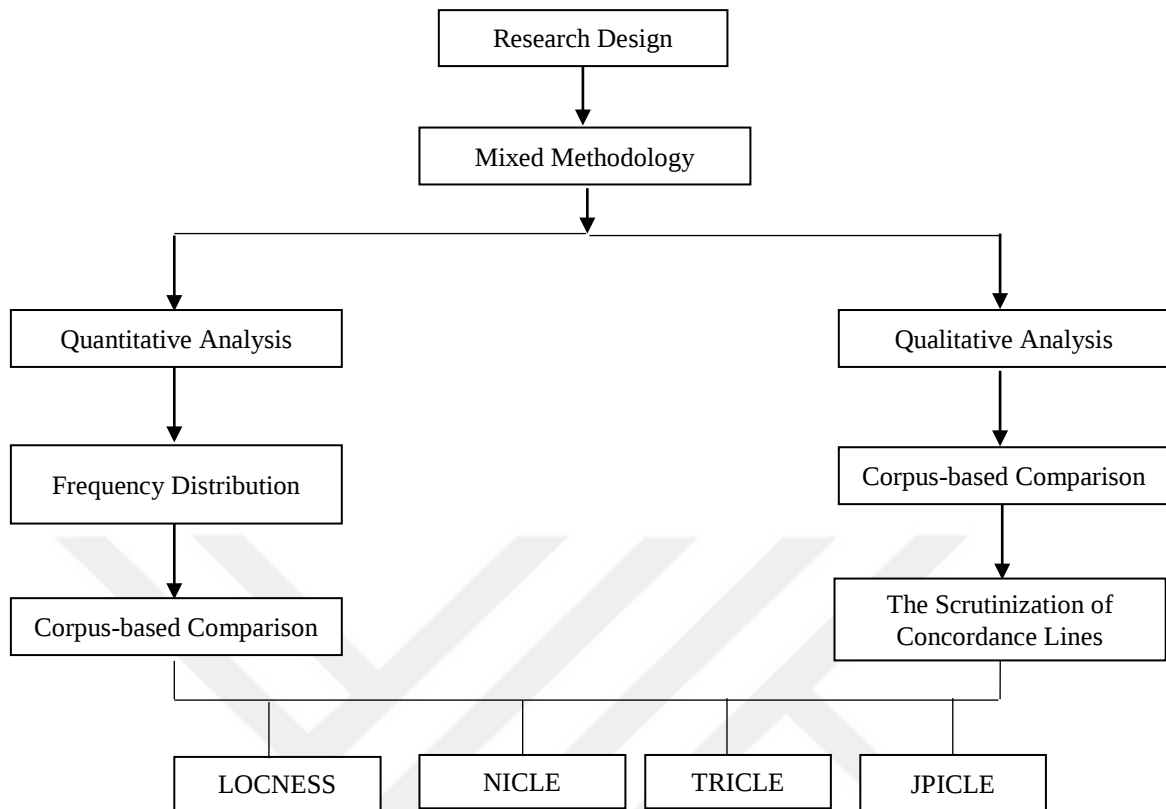
TRICLE	Cohen's Kappa (κ) value	p
Could	1	<,001
May	1	<,001
Might	1	<,001
Must	,887	<,001
Should	,932	<,001
Will	,871	<,001
Would	,954	<,001

The epistemic modal auxiliaries and lexical verbs whose inter-rater reliability tests were found to be low were identified. A final decision was made regarding whether these lexical verbs and modal auxiliaries are epistemic or not by the researcher. According to the final decision of the researcher, the data was analyzed to obtain the results. In this regard, the concordance lines of modal auxiliaries and lexical verbs that had low inter-rater reliability were examined to make a final decision on how many epistemic uses there were in each item again by taking both raters' decisions into consideration.

3.14. Raw and Normalized Frequencies

After extracting the concordance lines and deciding whether they have been used with their epistemic meanings, an inter-rater reliability test was applied. According to the results, the raw frequencies were found out. To make comparisons between the results from different corpora, normalized frequencies were used (Hunston, 2022). The results reached from different corpora with varying lengths, normalizing the results enables not to encounter misleading results (Meyer, 2002; Reppen et al., 1998). In this regard, normalization of raw frequency can be defined as “a way to adjust raw frequency counts from texts of different lengths so that they can be compared accurately” (Reppen et al., 1998: 263). To normalize the raw frequency, raw frequency was divided into the total number of words in the corpus. The result from this calculation was multiplied with the number that was chosen as a basis for norming. In this study, the raw frequencies were normalized per 10,000 because choosing a higher basis was advised by (Meyer, 2002). Also, the normalized results seemed more clearly by using 10,00 as a basis. In addition, this study both quantitative and qualitative methods investigating the use of epistemic modal auxiliaries and lexical verbs in the argumentative essays of native undergraduate students and non-native undergraduate students involving Turkish, Norwegian, and Japanese (Figure 1).

Figure 1: Research Design



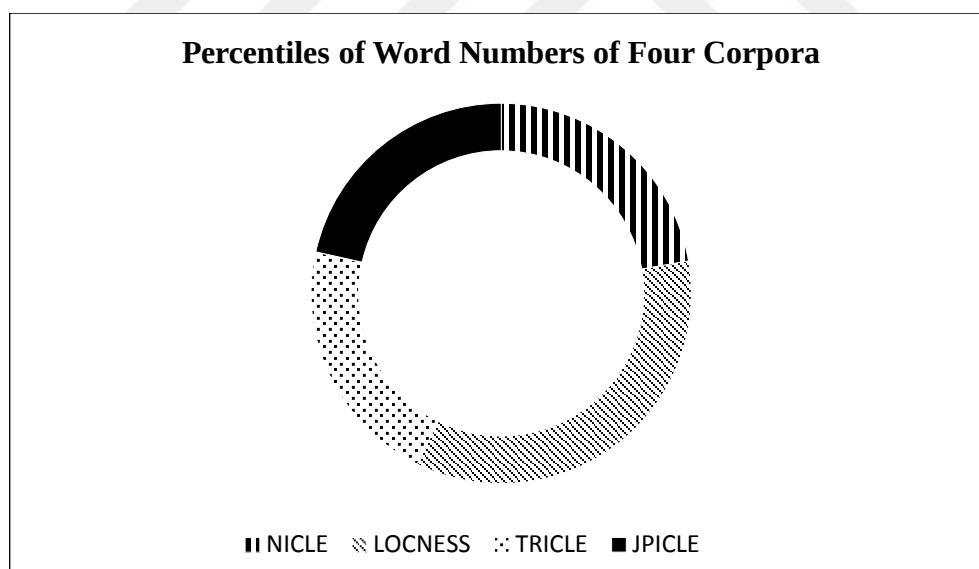
CHAPTER FOUR

4. RESULTS AND DISCUSSION

4.1. Introduction

This chapter presents the results obtained from the corpus analyses of four corpora regarding the use of epistemic lexical verbs and epistemic modal auxiliaries. Since the number of words was different in each corpus, the researcher did not compare the raw frequencies. Instead, the frequencies were normalized per 10.000 words. The number of words used in LOCNESS was 324,304 (35%). In TRICLE, it was 196,644 (21,22%). The number of words in JPICL was 199,137 (21,49%). Lastly, in NICLE, the number of words was 206,408 (22,27%) (Figure 2). Therefore, the results of corpus analysis were obtained in terms of normalized frequencies.

Figure 2: Comparison of the Number of Words in Four Corpora

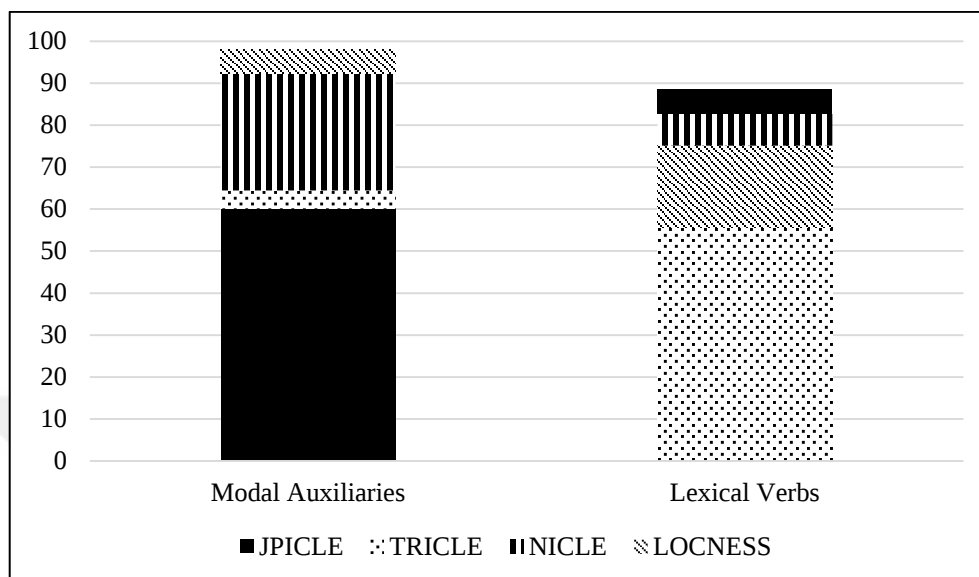


4.2. Frequency Distribution of Epistemic Lexical Verbs and Modal Auxiliaries

The study involved 56 epistemic devices: 49 epistemic lexical verbs and 7 modal auxiliaries. The frequency distribution of both epistemic modal auxiliaries and lexical verbs were found in four corpora. The frequencies were normalized into per 10.000 words because the sizes of four corpora were different. According to the overall frequencies of epistemic modal auxiliaries and lexical

verbs in four corpora, JPICLE (n=60,13) was the one that used the EMAs the least among the four corpora (Figure 3). Since the table is a stacked table, the frequencies were added up bottom to top.

Figure 3: Overall Frequencies of EMAs and ELVs in Four Corpora



Therefore, TRICLE (n=64,47) comes next after JPICLE regarding the use of EMAs. On the other hand, NICLE (n=92,21) and LOCNESS (n=98,09) used the EMAs the most while there was a gap between the frequencies of these two corpora and that of TRICLE and JPICLE. Regarding this, while NICLE and LOCNESS used the EMAs the most, TRICLE and JPICLE used it the least with a huge gap between the frequency of EMAs of JPICLE and TRICLE and that of NICLE and LOCNESS.

The frequencies of ELVs in four corpora were also investigated (Figure 3). According to the results, similar to the frequency of epistemic modal auxiliaries, TRICLE (n=55,58) was the one that used the ELVs the least among four corpora. It was followed by LOCNESS (n=75,09) and NICLE (n=82,7) with a huge gap. Differently, JPICLE (n=88,43) used the ELVs the most among the four corpora. In this regard, while JPICLE did not prefer to use EMAs that much, it mostly preferred to use lexical verbs to convey the epistemic meaning. Moreover, TRICLE used both with nearly the same frequencies; however, it fell behind the other three corpora in using ELVs and behind the other two corpora in using epistemic modal auxiliaries. On the other hand, LOCNESS preferred to use EMAs (n=98,09) more than lexical verbs (n=75,09) to convey the epistemic meaning. While LOCNESS used the EMAs the most among four corpora, it used ELVs in the middle of three corpora. For NICLE, the frequencies of EMAs (n=92,21) and lexical verbs used by NICLE (n=82,7) were near to that of LOCNESS. Similarly, NICLE, too, preferred to use EMAs more than the lexical verbs to express modality.

4.3. Epistemic Devices: Modal Auxiliaries

4.3.1. Frequency of Epistemic Modal Auxiliaries

The top ten epistemic modal auxiliaries used in four corpora were found (Table 18). According to the results, the epistemic *would (not)* was the most used modal auxiliary in LOCNESS (40,7) and NICLE (30,78). Differently, the mostly used epistemic modal auxiliary in JPICLE (24,47) and TRICLE (33,87) was the modal auxiliary *will (not)*. On the other hand, in LOCNESS (28,61) and NICLE (28,53), the modal auxiliary *will (not)* was used in the second rank. Moreover, EMAs used in LOCNESS and NICLE were used nearly in the same rank. Except for the modal auxiliaries *should (not)*, *must (not)*, *could (not)*, and *might (not)*, both had the epistemic modal auxiliary *would (not)* in the first rank and the epistemic *will (not)* in the second rank. In addition, they had *may (not)* in the third rank. In NICLE, the modal auxiliaries *must (not)* (1,81) was used in the sixth rank and *should (not)* (0,43) in the seventh rank. However, in LOCNESS, the modal auxiliaries *should (not)* (0,47) was used in the sixth rank and *must (not)* (0,33) was used in the seventh rank. Moreover, NICLE had the epistemic *might (not)* (11,71) in the fourth rank and *could (not)* (8,69) in the fifth rank whereas LOCNESS had the epistemic *might (not)* (2,35) in the fifth rank and *could (not)* (12,47) in the fourth rank. Although the epistemic modal auxiliary *might (not)* was used in the fifth rank in LOCNESS the frequency of it was found to be lower than both the other EMAs and the use of it in NICLE.

Table 18: Frequency Distribution of Epistemic Modal Auxiliaries in Four Corpora

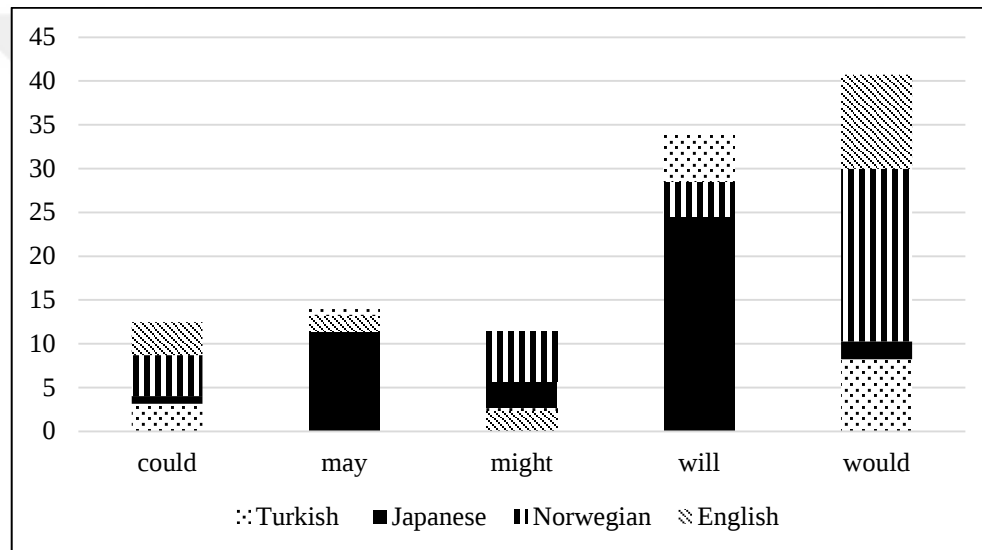
LOCNESS		NICLE		JPICLE		TRICLE	
Modals	<i>nf</i>	Modals	<i>nf</i>	Modals	<i>nf</i>	Modals	<i>nf</i>
Would (not)	40,7	Would (not)	30,78	Will (not)	24,47	Will (not)	33,87
Will (not)	28,61	Will (not)	28,53	May (not)	13,27	May (not)	13,95
May (not)	13,16	May (not)	11,75	Would (not)	10,27	Would (not)	8,22
Could (not)	12,47	Might (not)	11,71	Might (not)	5,7	Could (not)	3,13
Might (not)	2,35	Could (not)	8,69	Could (not)	4,01	Might (not)	2,63
Should (not)	0,47	Must (not)	1,81	Must (not)	1,8	Must (not)	1,72
Must (not)	0,33	Should (not)	0,43	Should (not)	0,66	Should (not)	0,95

As for JPICLE and TRICLE, the epistemic *may (not)* was used in the second rank in both corpora (13,27 for JPICLE and 13,95 for TRICLE. In JPICLE, the epistemic *would (not)* (10,27) and *might (not)* (5,7) were in the third and the fourth rank; however, in TRICLE, epistemic *would (not)* (8,22) was in the third rank and *could (not)* (3,13) was in the fourth rank. Furthermore, the epistemic modal auxiliaries *could (not)* (4,01), *must (not)* (1,8), and *should (not)* (0,66) were used in the fifth, sixth, and seventh ranks sequentially in JPICLE. Differently, in TRICLE, the modal auxiliary *might (not)* (2,63) was used in the fifth rank, *must (not)* (1,72) in the sixth rank, and *should (not)* (0,95) in the seventh rank. Overall, the epistemic *would (not)* was mostly preferred in

LOCNESS and NICLE whereas it was used with lower frequency in JPICLE and TRICLE. Moreover, while the epistemic *might (not)* was used with higher frequency in NICLE, it was less preferred in JPICLE, LOCNESS, and TRICLE. Lastly, apart from these, the rest of the results were found near regarding frequencies and ranks in four corpora.

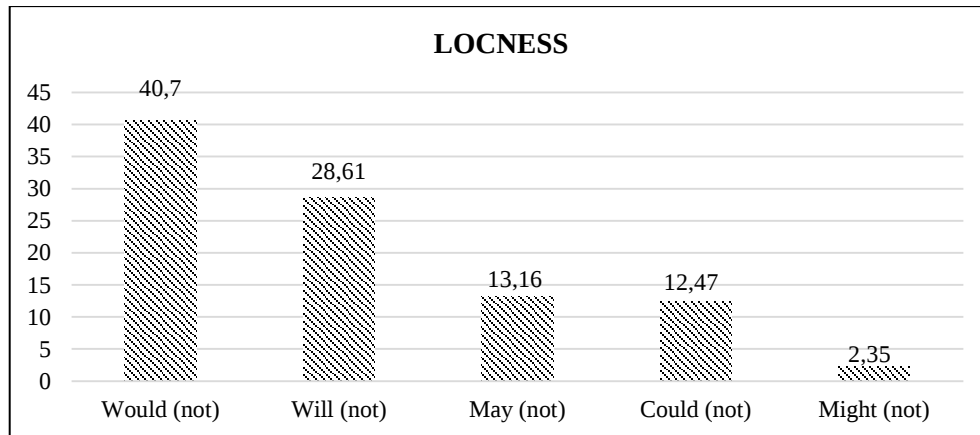
Overall, when comparing the epistemic modal auxiliaries used in four corpora, the top five EMAs were used with different frequencies in four corpora (Figure 4). While epistemic *would* and *could* was used by LOCNESS the most, the epistemic *will* and *may* was utilized by TRICLE the most. Moreover, the epistemic *might* was used by NICLE. In this regard, the use EMAs showed differences in four corpora.

Figure 4: Top Five Epistemic Modal Auxiliaries in Four Corpora



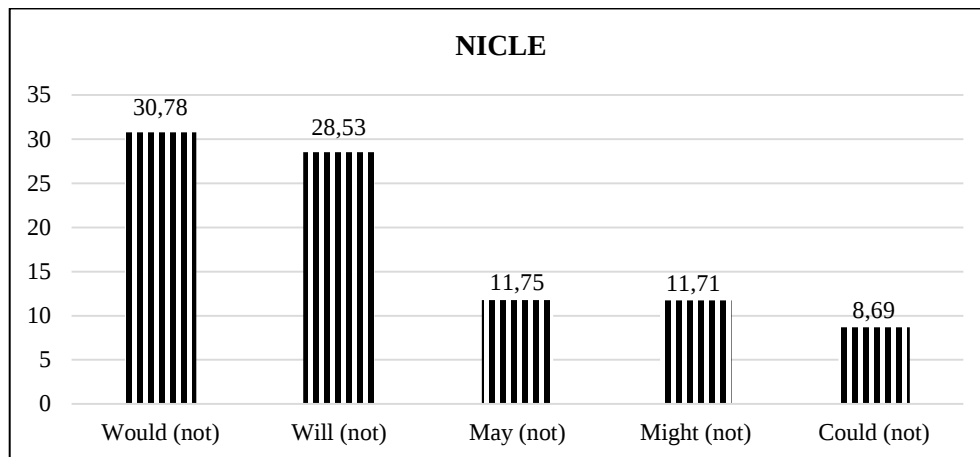
The most used top five epistemic modal auxiliaries were identified in LOCNESS (Figure 5). The most used one was found as *would (not)* (40,7) with a higher frequency than the others. The second one was the modal auxiliary *will (not)* (28,61). It was followed by the epistemic modal auxiliary *may (not)* (13,16). The subsequent one was found as *could (not)* (112,47), whose frequency is near the frequency of *may (not)*. Lastly, the epistemic *might (not)* (2,35) is the fifth one with a rather low frequency. In this regard, while the epistemic modal auxiliaries *would (not)* and *will (not)* were preferred with a higher frequency, the epistemic modal auxiliaries *could (not)* and *may (not)* were used with moderate frequencies, and the epistemic *might (not)* was hardly preferred.

Figure 5: Top Five Epistemic Modal Auxiliaries in LOCNESS



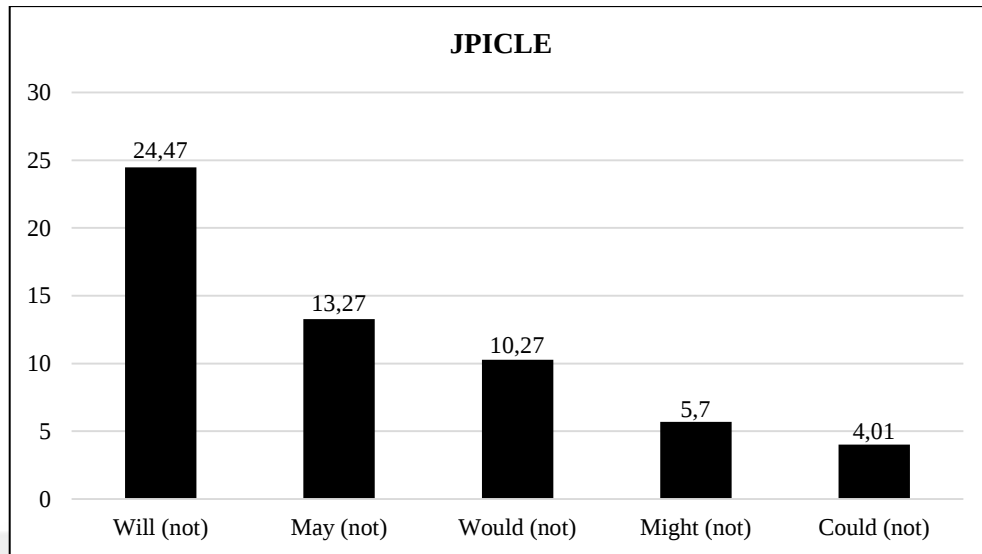
For NICLE, the top five epistemic modal auxiliaries were found (Figure 6). The most preferred EMAs were *would (not)* (30,78) and *will (not)* (28,53). Then, they were followed by the epistemic modal auxiliaries *may (not)* (11,75), *might (not)* (11,71), and *could (not)* (8,69). According to the results, it can be maintained the epistemic modal auxiliaries *would (not)* and *will (not)* were preferred with rather higher frequencies whereas the epistemic modal auxiliaries *could (not)*, *may (not)*, and *might (not)* were used with moderate frequencies and lower than the first two. Although the gap between the frequencies was not so much between the first two and the last three EMAs, it can be suggested *would (not)* and *will (not)* were the more preferred EMAs than the last three ones.

Figure 6: Top Five Epistemic Modal Auxiliaries in NICLE



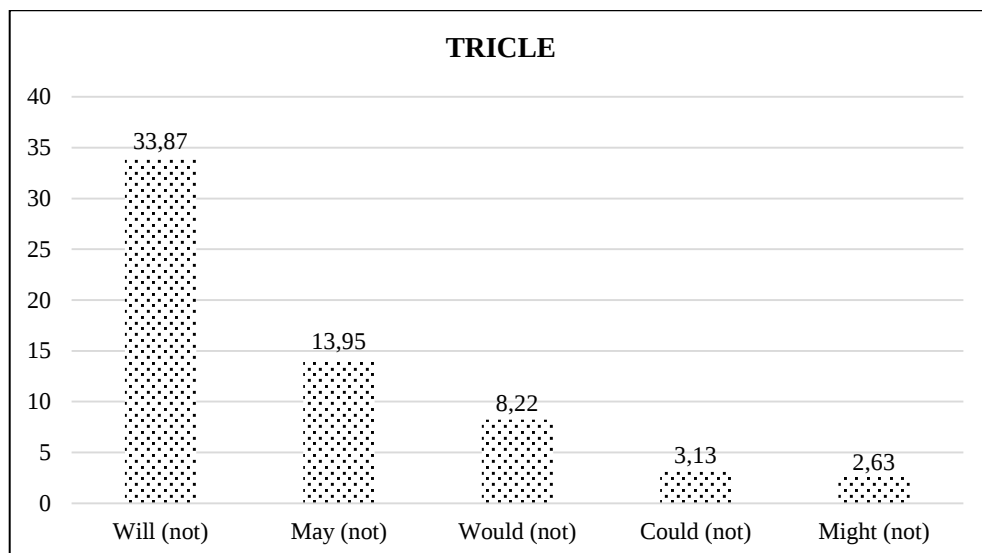
The top five epistemic modal auxiliaries used in JPICLE were found (Figure 7). According to the results, the epistemic *will (not)* (24,47) was mostly used with higher frequency than the other four EMAs. Subsequently, the epistemic modal auxiliaries *may (not)* (13,27) and *would (not)* (10,27) were used moderately in the second and third ranks. On the other hand, the epistemic modal auxiliaries *might (not)* (5,7) and *could (not)* (4,01) were used with a rather low frequency.

Figure 7: Top Five Epistemic Modal Auxiliaries in JPICLE



As for T-ICLE, the top five epistemic modal auxiliaries in TRICLE were found (Figure 8). The most used EMAs was *will (not)* (33,87). It was followed by the epistemic modal auxiliaries *may (not)* (13,95), *would (not)* (8,22), *could (not)* (3,13), and *might (not)* (2,63). In this regard, the epistemic *will (not)* was preferred with a higher frequency than the other four EMAs. Moreover, epistemic *may (not)* were used nearly half as much as the epistemic *will (not)*. The epistemic modal auxiliaries *would (not)* and *could (not)* were preferred with rather low frequencies. Lastly, the epistemic *might (not)* was used with a pretty low frequency.

Figure 8: Top Five Epistemic Modal Auxiliaries in TRICLE



4.3.2. Epistemic Strengths of Epistemic Modal Auxiliaries

The epistemic modal auxiliaries were categorized according to Ngula's (2015) three-way classification in terms of epistemic strength. According to their commitment and probability degrees, their epistemic strengths were identified. Thus, the EMAs that indicate the weak epistemic strength are *could*, *may*, and *might*. The EMAs having medium epistemic strength are *should* and *would*. Lastly, the EMAs that indicate strong epistemic strength are *must* and *will* (Table 19).

Table 19: Epistemic Strengths of Epistemic Modal Auxiliaries

Weak	Medium	Strong
Could	Should	Must
May	Would	Will
Might		

The frequency distribution of EMAs indicating weak epistemic strength was identified regarding four corpora (Table 20). The EMAs indicating weak epistemic strength were used more in NICLE (32,15). It is followed by LOCNESS (27,98), JPICLE (22,98), and TRICLE (12,94) sequentially. According to the results, NICLE, JPICLE, and LOCNESS used EMAs with nearly close frequencies. On the other hand, TRICLE had the lowest frequency utilizing the weak epistemic strength with a rather low frequency.

Table 20: Frequency Distribution of Weak Epistemic Modal Auxiliaries in Four Corpora

	LOCNESS	JPICLE	NICLE	TRICLE
	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>
Weak				
Could	12,47	4,01	8,69	3,13
May	13,16	13,27	11,75	7,18
Might	2,35	5,7	11,71	2,63
Total	27,98	22,98	27,98	12,94

For the frequency distribution of EMAs indicating medium epistemic strength, LOCNESS (41,17) used the EMAs utilizing medium epistemic strength with the highest frequency among the four corpora (Table 21). It was followed by NICLE (31,21) with a moderate frequency. However, JPICLE (10,93) and TRICLE (9,17) used EMAs with a rather low frequency. Moreover, the epistemic *should* was used less than the epistemic *would* in four corpora. Regarding this, to utilize medium epistemic strength, four corpora preferred to use the epistemic *would* than *should*.

Table 21: Frequency Distribution of Medium Epistemic Modal Auxiliaries in Four Corpora

	LOCNESS	JPICLE	NICLE	TRICLE
	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>
Medium				
Should	0,47	0,66	0,43	0,95
Would	40,7	10,27	30,78	8,22
Total	41,17	10,93	31,21	9,17

Lastly, TRICLE (35,59) used more EMAs indicating strong epistemic strength than the other three corpora with a higher frequency (Table 22). It was followed by LOCNESS (28,94), NICLE (30,34), and JPICLE (26,27) in sequence. Although the frequency of NICLE, JPICLE, and LOCNESS was found near, the frequency of TRICLE was rather high regarding strong epistemic strength. In addition, to utilize the strong epistemic strength, four corpora preferred to use the epistemic *will* more than the epistemic *must*.

Table 22: Frequency Distribution of Strong Epistemic Modal Auxiliaries in Four Corpora

	LOCNESS	JPICLE	NICLE	TRICLE
	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>
Strong				
Must	0,33	1,8	1,81	1,72
Will	28,61	24,47	28,59	33,87
Total	28,94	26,27	30,34	35,59

The overall frequency distribution of epistemic strength in four corpora was identified to see the similarities and differences among the four corpora (Table 23). According to the results, TRICLE was found to be the corpus using the epistemic modal auxiliaries the least to indicate weak epistemic strength. On the other hand, while TRICLE used weak epistemic strength moderately, it was found it used medium epistemic strength with a pretty low frequency and used strong epistemic strength with a pretty high frequency in terms of both its own corpus and among four corpora.

Table 23: Frequency Distribution of Epistemic Strength in Four Corpora

Strength of Epistemic Modal Auxiliaries	LOCNESS	JPICLE	NICLE	TRICLE
Weak	27,98	22,98	32,15	12,94
Medium	41,17	10,93	31,21	9,17
Strong	28,94	26,27	30,34	35,59
Total	98,09	60,18	93,7	57,7

Regarding the weak epistemic strength, the frequencies of NICLE, JPICLE, and LOCNESS were close. Moreover, the frequencies of JPICLE and TRICLE were rather low in terms of medium epistemic strength while the frequencies of NICLE and LOCNESS were found pretty high. In

LOCNESS, the EMAs indicating medium epistemic strength were used more; however, the ones that indicate strong epistemic modality were utilized rather low. Differently, in JPICLE, the EMAs indicating medium epistemic strength were used the lowest although the weak epistemic strength was utilized the most, and medium epistemic strength was used moderately. Similarly, in NICLE, the EMAs to indicate weak epistemic strength were used the most. However, the strong epistemic strength was utilized the least.

Overall, TRICLE utilized medium epistemic strength the least whereas it used strong epistemic strength the most. For NICLE, it used strong epistemic strength the least whereas it utilized weak epistemic strength the most. Similarly, in JPICLE, weak epistemic strength was utilized the most; however, medium epistemic strength was used the least. In LOCNESS, medium epistemic strength was used the most while strong epistemic strength was utilized the least. Furthermore, LOCNESS (98,09) was found to be the one indicating epistemic strength in total the most. Similarly, the total frequency indicating epistemic strength in NICLE (93,7) was near to that of LOCNESS. On the other hand, the total frequency of utilizing epistemic strength in JPICLE (60,18) and TRICLE (57,7) was rather low. Furthermore, from this result, it can be understood the use of EMAs was also the highest in LOCNESS among the four corpora.

The log-likelihood tests for epistemic strength in NICLE and JPICLE were found (Table 24). Regarding the results, there is not any statistically significant difference between NICLE and JPICLE in terms of utilizing weak (LL=1.53, p-value=0.17) and strong (LL=0.29, p-value=0.59) epistemic strength. However, NICLE used more epistemic modal auxiliary indicating weak and strong epistemic strength than JPICLE. On the other hand, there is a statistically significant difference (LL=10.18, p-value=0.00) between NICLE and JPICLE in terms of utilizing medium epistemic strength. Moreover, it was revealed NICLE used more medium epistemic strength than JPICLE with a statistically significant difference.

Table 24: Log-Likelihood Tests for Epistemic Strength in NICLE and JPICLE

Epistemic Strength	NICLE	JPICLE	LL		p-value
Weak	32,15	22,98	1.53	+	0.17
Medium	31,21	10,93	10.18	+	0.00
Strong	30,34	26,27	0.29	+	0.59

As for the log-likelihood test of epistemic strengths in NICLE and LOCNESS, it was found there is not any statistically significant difference between NICLE and LOCNESS regarding weak (LL=0.81, p-value=0.30), medium (LL=1.37, p-value=0.23), and strong (LL=0.03, p-value=0.79) epistemic strength. According to the results, N-ICLE used more weak epistemic strength than LOCNESS. Also, LOCNESS utilized more medium and strong epistemic strength than NICLE. Nevertheless, there is not any statistically significant difference between them (Table 25).

Table 25: Log-Likelihood Tests for Epistemic Strength in NICLE and LOCNESS

Epistemic Strength	NICLE	LOCNESS	LL		p-value
Weak	32,15	27,98	0.81	+	0.30
Medium	31,21	41,17	1.37	-	0.23
Strong	30,34	28,94	0.03	-	0.79

When it comes to NICLE and TRICLE, the log-likelihood tests were found and analyzed (Table 26). According to the results, there was not any statistical difference between NICLE and TRICLE regarding strong (LL=0.42, p-value=0.53) epistemic strength. However, TRICLE utilized more strong epistemic strength than NICLE. On the other hand, it was revealed there is a statistically significant difference between NICLE and TRICLE in terms of medium epistemic strength (LL=12,71, p-value=0.00) and weak epistemic strength (LL=8,45, p-value=0.00). Furthermore, NICLE was found to have used more EMAs indicating medium and weak epistemic strength than TRICLE.

Table 26: Log-Likelihood Tests for Epistemic Strength in NICLE and TRICLE

Epistemic Strength	NICLE	TRICLE	LL		p-value
Weak	32,15	12,94	8,45	+	0.00
Medium	31,21	9,17	12,71	+	0.00
Strong	30,34	35,59	0.42	-	0.53

Moreover, the log-likelihood tests for epistemic strength in JPICLE and LOCNESS were found and analyzed (Table 27). In this regard, there was not found any statistically significant difference between JPICLE and LOCNESS concerning weak (LL=0.49, p-value=0.47) and strong (LL=0.13, p-value=0.78) epistemic strength although LOCNESS used more EMAs indicating weak and strong epistemic strength. On the other hand, for medium epistemic strength, there is a statistically significant difference (LL=18.70, p-value=0.00) between JPICLE and LOCNESS. It was found LOCNESS utilized more medium epistemic strength than JPICLE.

Table 27: Log-Likelihood Tests for Epistemic Strength in JPICLE and LOCNESS

Epistemic Strength	JPICLE	LOCNESS	LL		p-value
Weak	22,98	27,98	0.49	-	0.47
Medium	10,93	41,17	18.70	-	0.00
Strong	26,27	28,94	0.13	-	0.78

Regarding JPICLE and TRICLE, the log-likelihood tests were found (Table 28). It was revealed there is not any statistically significant difference between JPICLE and TRICLE in terms of weak (LL=2.84, p-value=0.08), medium (LL=0.15, p-value=0.81), and strong (LL=1.41, p-value=0.24) epistemic strength. Nevertheless, JPICLE used more EMAs indicating weak and

medium epistemic strength than TRICLE. Also, TRICLE utilized more strong epistemic strength than JPICLE.

Table 28: Log-Likelihood Tests for Epistemic Strength in JPICLE and TRICLE

Epistemic Strength	JPICLE	TRICLE	LL		p-value
Weak	22,98	12,94	2.84	+	0.08
Medium	10,93	9,17	0.15	+	0.81
Strong	26,27	35,59	1.41	-	0.24

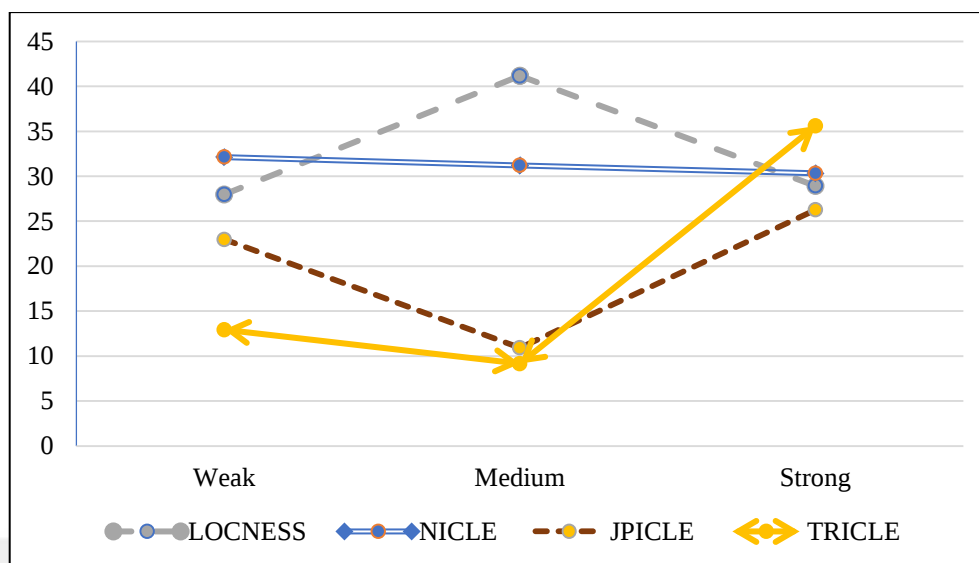
Lastly, for LOCNESS and TRICLE, the log-likelihood tests were found and analyzed (Table 29). According to the results, it was found there is not any statistically significant difference between LOCNESS and TRICLE regarding strong (LL=0.69, p-value=0.37) epistemic strength. Moreover, TRICLE utilized more strong epistemic strength than LOCNESS. On the other hand, there was found a statistically significant difference between LOCNESS and TRICLE in terms of medium (LL=22, p-value=0.00) and weak (LL=5.66, p-value=0.01) epistemic strength. LOCNESS used more EMAs indicating medium and weak epistemic strength than TRICLE.

Table 29: Log-Likelihood Tests for Epistemic Strength in LOCNESS and TRICLE

Epistemic Strength	LOCNESS	TRICLE	LL		p-value
Weak	27,98	12,94	5.66	+	0.01
Medium	41,17	9,17	22	+	0.00
Strong	28,94	35,59	0.69	-	0.37

All in all, varying degrees of epistemic strengths utilized by the four corpora showed significant differences (Figure 9). Concerning this, LOCNESS preferred to use medium epistemic strength more than strong and weak ones. NICLE used weak epistemic strength more than medium and strong epistemic strengths; however, it used the other epistemic strengths with near frequencies. On the other hand, TRICLE and JPICLE utilized the epistemic strengths as the exact opposite of LOCNESS' use of epistemic strengths. In this regard, they preferred to use medium epistemic strength the least while they preferred the strong epistemic strength the most.

Figure 9: Overall Frequencies of Epistemic Strength of Modal Auxiliaries in Four Corpora



4.3.3. Qualitative Analysis of Modal Auxiliaries' Elimination of Epistemic Uses from the Non-Epistemic Uses

To eliminate non-epistemic modal auxiliaries, concordance lines were examined by deciding which lines involving modal auxiliaries are epistemic. Regarding this, qualitative analysis was conducted to scrutinize the concordance outputs. In the case of modal auxiliaries in four corpora, the modal auxiliaries *could (not)*, *may (not)*, *will (not)*, *would (not)*, *should (not)*, and *must (not)* were used with both their epistemic meanings and non-epistemic meanings except the modal auxiliary *might*. As in the case of modal auxiliary *could (not)*, it can be used with both an epistemic meaning (1-2) and non-epistemic meaning (3-4):

1. So, a boxer who has been boxing for a number of years, ten for example, and retires is very likely to have brain damage and it **could** be very severe or hardly noticable but the damage is still there. (LOCNESS)
2. We can't allow ourselves to give up just because we think it **couldn't** possibly be any better. (NICLE)
3. When it has first started to be sold, it was too expensive to buy that only rich people **could** afford it. (TRICLE)
4. Moreover, women were not independent since they **could not** earn money by themselves. (JPICLE)

In this regard, whereas the concordance lines (1-2) from LOCNESS and NICLE show the modal auxiliary *could (not)* was used with its epistemic meaning conveying a possibility, the examples from the concordance lines (3-4) reached from TRICLE and JPICLE indicate it is used

with a non-epistemic meaning conveying an ability. In other cases, it was used with its non-epistemic meaning conveying permission in four corpora. As for the modal auxiliary *may*, it was used with epistemic (1-2) and non-epistemic (3) meanings in four corpora:

1. You **may** not use it tomorrow but there will be a time to use it. (TRICLE)
2. The way people dream today **may not** be the same as the dreams of older generations. (NICLE)
3. He was executed on **May** 20, 1982. (LOCNESS)

The examples (1-2) indicate that the modal auxiliary *may* can be used with its epistemic meaning conveying a possibility. On the other hand, the example (3) shows the word *may* was used to indicate a date in the calendar. In addition, it can be seen the word *may* was utilized as a name of a person in the concordance lines.

When it comes to the modal auxiliary *must*, it can convey epistemic and non-epistemic meanings. While the epistemic meaning of *must* conveys possibility, the non-epistemic meaning of it indicates necessity. The four corpora used in this study include both meanings:

1. This **must** be very difficult for them. (NICLE)
2. We **mustn't** destroy the culture and land of original people. (JPICLE)
3. Before any essay can be started, the key words in the essay title **must** be defined and understood. (LOCNESS)

The concordance line (1) shows the epistemic use of modal auxiliary *must* by conveying possibility. In example (1), it signifies the situation's possibility of being difficult. However, the examples (2-3) show the use of modal auxiliary *must* with its non-epistemic use conveying the meaning of necessity.

For the modal auxiliary *should*, there were two meanings conveyed by it: epistemic and non-epistemic meanings. The epistemic meaning involves prediction and probability whereas non-epistemic meaning carries a sense of obligation. In this regard, it was found that four corpora used in this study involves both meanings; therefore, it should be eliminated:

1. If somebody is to ready to do such a terrible thing, there **should** be incredibly huge reasons to force him to this. (TRICLE)
2. If the monitors are of better quality, this **should not** be a problem. (LOCNESS)
3. The original inhabitant and the emigrant **should** be equal. (JPICLE)
4. There is nothing wrong with watching television, but we **should not** overdo it. (NICLE)

In these examples (1-2), the modal auxiliary *should* was used with its epistemic meaning involving probability of a situation. On the other hand, the examples (3-4) show that the epistemic modal auxiliary *should* was used as non-epistemic modal auxiliary carrying a meaning involving an obligation.

For the modal auxiliary *will*, it has the same situation with other modal auxiliaries in four corpora. It was used with both epistemic and non-epistemic meanings. The epistemic meaning of the modal auxiliary *will* involves a guess on a proposition and prediction. However, non-epistemic meaning of it conveys the situations for future events:

1. In order to do this program we **will** probably lose attendance, because people will not be able to afford tuition. (LOCNESS)
2. People **will not** be interested in working if society is falling apart. (NICLE)
3. We **will** examine this under different categories. (TRICLE)
4. Now, we live in the society which we can select our **will** freely. (JPICLE)

The examples (1-2) present the epistemic use of modal auxiliary *will*. In these examples, the epistemic *will* involves a prediction about the future indicating a guess. On the other hand, the other example (3) shows the modal auxiliary *will* was used with its non-epistemic meaning indicating volition. In addition, the last example (4) presents the word *will* was used as noun in the concordance lines. In this regard, there was the word *will* used as a noun in four corpora other than the non-epistemic modal auxiliary *will*.

Similarly, the modal auxiliary *would* was used with both epistemic and non-epistemic meanings in four corpora. While the epistemic *would* indicates prediction, the non-epistemic *would* conveys volition as a past simple of *will*. Thus, the concordance lines was examined to eliminate them:

1. More deaths **would** occur through boxing if it were banned. (LOCNESS)
2. If the most lucrative profession there was, was to dig wells in Africa, there **wouldn't** be any problem getting someone to do that. (NICLE)
3. I'd like to study how to use English in real society than how make correct sentences. (JPICLE)

As it can be seen in the examples (1-2), the modal auxiliary *would* was used with its epistemic meaning involving prediction. On the other hand, the modal auxiliary *would* was used with the phrase like to and it created the phrase *would like to* indicating a volition. Thus, the example (3) shows the modal auxiliary *would* was used with its non-epistemic meaning.

Lastly, the modal auxiliary *might* was used only with its epistemic meaning involving possibility. However, non-epistemic modal auxiliary *might* indicate permission. In this regard, four corpora had epistemic use of modal auxiliary *might*:

1. If cell phones had not been invented, our daily lives **might** be more uncomfortable, troublesome, and monotonous. (LOCNESS)

The example (1) presents the modal auxiliary *might* was used with its epistemic meaning in the concordance line. It conveys the meaning of possibility; thus, the proposition in the concordance line suggests that a possibility in case of a hypothetical situation.

4.3.4. Cases of Epistemic Modal Auxiliaries in Four Corpora

Typical cases of epistemic modal auxiliaries were investigated in terms of salient uses, the lexico-grammatical patterns, and collocations used with epistemic modal auxiliaries. Therefore, the concordance lines of four corpora were subjected to a qualitative analysis to find out the salient uses of these lexico-grammatical patterns in four corpora. Regarding this, the similarities and differences among four corpora were discussed. The modal auxiliaries *may* and *might*, *will* and *would*, and *must* and *should* were discussed in couples whereas the modal auxiliary *could* was discussed separately because the former ones have mutual correspondences.

4.3.4.1. The Case of Epistemic ‘May’ and ‘Might’

The cases of epistemic modal auxiliaries *may* and *might* were examined through the concordance lines in four corpora to compare the use of these modal auxiliaries in terms of lexico-grammatical patterns. To this end, objective or subjective use of epistemic *may* and *might*, collocations, and the salient patterns were examined across four corpora. As a basis, epistemic modal auxiliaries *may* and *might* were used with nearly the same frequencies in NICLE whereas the epistemic modal auxiliary *may* was used more than *might* in the other three corpora. Therefore, the qualitative analysis also included categorization of the patterns and counting the uses of the salient patterns.

First of all, it was found the epistemic modal auxiliary *may* was used in objective modal use, rather than subjective modal, in four corpora. In this regard, the use of subjective *I* and ‘we’ were found few in four corpora. Instead, the four corpora had the use of objective modal by mostly utilizing the noun clauses, there, it, some people, and they:

1. I **may** advise him to read "How to win friends and Influence people", but to take it with a grain of salt and not to overdue the "lavish praise". (LOCNESS)

2. We **may** not be as eager to invent new things or dream of some revolutionary invention, because the world has come so far already. (NICLE)
3. Some people **may** be angry then they hear this. (JPICLE)
4. It **may** cause deaths. (TRICLE)

Although the examples (1-2) indicate the subjective use of epistemic *may* by utilizing subjective *I* and ‘we’, the other examples (3-4) show the objective use of epistemic *may* by using the unknown person ‘some people’ and ‘it’. In addition, the frequency of subjective epistemic *may* in four corpora was found as TICLE (n=11), NICLE (n=18), LOCNESS (n=19), and JPICLE (n=34) sequentially.

As for the negation of epistemic *may*, it was found the modal auxiliary *may* was used in a positive clause mostly, rather than the use of negation, in four corpora. In this regard, the negation of epistemic *may* was used few in four corpora (1):

1. Your child **may not** know the meanings of these words: poor, being hungry, spending time in rows. (TRICLE)

Regarding this, four corpora preferred using the epistemic *may* in a positive clause mostly while they used the negation of it few in four corpora: LOCNESS (n=65), JPICLE (n=38), TRICLE (n=30), and NICLE (n=20) sequentially.

In addition, it was found that the four corpora used the concessive *may* by utilizing *but*, *while*, *whereas*, *whilst*, *although*, *yet*, and *or*. The use of concessive *may* indicates the possibility of a proposition to be truth by not saying the proposition is not fully impossible to be true:

1. They **may** be fair **but** they also reduce the democracy within Britain. (LOCNESS)
2. It gives you insight, **while** religion **may** hinder this. (NICLE)
3. **Although** I **may** be a meddler, teacher should take oral lesson and reading more. (JPICLE)
4. The problem **may** be a government pressure **or may** be a media group and so on. (TRICLE)

As in the examples, the four corpora utilized the concessive *may* so as not to reject the proposition fully but to give a possibility with *but*, *while*, *although*, and *or*. Among these connectors, *but* is the one that was used most in four corpora: LOCNESS (n=90), TRICLE (n=69), NICLE (n=60), and JPICLE (n=49). Moreover, LOCNESS was found as the corpus that used these connectors to use concessive *may* at the highest frequency (n=149). It was followed by NICLE (n=90), TICLE (n=83), and JPICLE (n=62) sequentially.

Other than the concessive uses, it was revealed the concordance lines in four corpora involved the combination of epistemic *may* with the use of the if clause. So, the concordance lines were examined:

1. Equality for men and women **may** change **if** there was an imbalance in the population. (LOCNESS)
2. **If** you do not have the same opinions as them, you **may** not be welcome. (NICLE)
3. **If** my memory is correct, provably it **may** cost you \$250 (2 hours). (JPICLE)
4. **If** there is 20 years between man and woman, this situation **may** cause divorce. (TRICLE)

The examples (1-2-3-4) from the concordance lines of four corpora indicate the use of if conditionals with the epistemic *may*. These clauses convey the meaning of possibility and probability with a condition. In this regard, LOCNESS (n=50) is the one that utilized if conditional with epistemic *may* most. It was followed by TRICLE (n=30), NICLE (n=23), and JPICLE (n=2) sequentially.

Another most salient collocation used with epistemic *may* is the use of *also*. It was discovered that the epistemic *may* is mostly used with the use of *also* in four corpora:

1. Cutting down on national speed limit **may also** encourage drivers to take more care. (LOCNESS)
2. Looking at other children and mothers **may also** be a huge problem. (NICLE)
3. Not only **may** have many Japanese but **also** many foreign people many distressing and painful experiences I have never experienced. (JPICLE)
4. We should get rid of this dilemma made of glass which **may also** considered as theoretical information. (TRICLE)

Regarding this, the epistemic *may* was used with the use of *also* as both a collocation and a salient pattern. While LOCNESS (n=37) utilized this pattern most and higher than the other three corpora, the frequency of this pattern was nearly the same in TRICLE (n=13), JPICLE (n=12), and NICLE (n=11) sequentially.

Another salient pattern used with epistemic *may* is the use of *even* and *well* as the collocations of epistemic *may*. However, the use of *even* in NICLE, JPICLE, and TRICLE is very low. Similarly, these three corpora did not use *even* as the collocation of epistemic *may* at all. On the other hand, these collocations were found salient in LOCNESS:

1. In fact the increase in Sporting facilities and training academies **may well** increase our sporting prowess in the long term, something that would be most welcomed and for which many would be happy to pay for. (LOCNESS)
2. In the future, there will be various crimes and the number of them **may even** increase. (JPICLE)
3. Also, unconscious ideology occurs, where there exists an underlying attitude, or prejudice, which **may not even** be consciously noticed. (LOCNESS)

These examples (1-2-3) from LOCNESS and JPICLE present the epistemic *may* was used with *even* and *well* as collocations. LOCNESS is the one that utilized epistemic *may* with the use of *even* most (n=13). It was followed by JPICLE (n=2), NICLE (n=1), and TRICLE (n=1). Moreover, while LOCNESS used the collocation of epistemic *may* with *well* with higher frequency (n=12), the other three corpora did not use this collocation.

Lastly, it was discovered there were other collocations used in four corpora, such as the use of *as well*, *of course*, *only*, *possibly*, and *in fact*. However, their frequency was found rather low when compared to other patterns and collocations. The epistemic *may* was used with *as well* as a collocation the most (n=4) while TRICLE comes next (n=1) and NICLE and JPICLE did not use this collocation at all. For the collocation of epistemic *may* and *only*, LOCNESS is the one that use it most; on the other hand, JPICLE comes next (n=1) and TRICLE and NICLE did not use it.

1. Europe **may as well** then take on the form of a "super"-country. (LOCNESS)
2. Just for a fight against another top ranking boxer that **may only** last a couple of round. (LOCNESS)

As it can be seen in the examples (1-2) from LOCNESS, the use of *as well* and *only* comes after the epistemic *may* as collocations. (Appendix 6).

When it comes to the salient patterns, lexico-grammatical patterns, and collocations of epistemic modal auxiliary *might*, there were similar patterns and collocations as those of epistemic *may*. First, the objective and subjective epistemic *might* was examined through the lines in four corpora. In four corpora, the objective epistemic *might* was dominated; the visibility of the writer was discovered as low. Whereas the objective modal *might* was conveyed with the use of *they*, *it*, *there*, *this*, and noun clauses, the subjective modal *might* was carried with the subject *I* and 'we' in four corpora:

1. **I might** rightly make the same statement about the comic strip in the Sunday edition of the Clinton Chronicle and be just as correct. (LOCNESS)

2. Therefore, **we might** as well have no pornographic pictures on the internet at all. (JPICLE)
3. In worst case **it might** kill them. (NICLE)
4. A hopeless cancer of a few years before **might** be an ordinary illness of today which can be cured very easily nowadays. (TRICLE)

Regarding the examples (1-2), the epistemic *might* was used as a subjective modal; on the other hand, the examples (3-4) from the concordance lines present the use of epistemic *might* as objective modal. The use of subject *I* and ‘we’ with epistemic *might* made the writer’s voice visible while the use of *it* or a noun clause made the clause objective with an objective modal. In this regard, the objective modal *might* was used most in four corpora while the subjective one was used low. JPICLE was the one using the subjective modal *might* with the highest frequency (n=12); it was followed by NICLE (n=11), LOCNESS (n=6), and TRICLE (n=4).

The negation of epistemic *might* was not preferred while conveying the epistemic meaning of *might*. In this regard, the positive clauses was preferred most in four corpora:

1. Or the person **might not** be in pain but he may be living without consciousness. (TRICLE)
2. In a while they **might not** be able to play on their own any more. (NICLE)

The examples (1-2) from the concordance lines of TRICLE and NICLE show the negation of epistemic *might*. Among these four corpora, NICLE is the one that used the negation of epistemic *might* most (n=29). It was followed by JPICLE (n=15), TRICLE (n=7), and LOCNESS (n=4). In this case, it was revealed that four corpora preferred to use positive clauses while using epistemic *might*.

Furthermore, the concessive *might* was preferred in four corpora by utilizing *but* mostly. It was followed by the use of *or* with the concessive *might*. Other than the use of *but* and *or* to convey the concessive *might*, there was the use of *while*, *whilst*, *although*, *on the other hand*, and *yet* with few frequencies:

1. I agree that five to ten years ago this **might** have been true, **but** now many employers are the victims. (LOCNESS)
2. The light **might** have gone out, **or** trains might have stopped. (JPICLE)
3. **Although** religion **might** not be a painkiller, the belief that it is can do much by itself. (NICLE)
4. **While** it is true that life is a valuable thing to possess, it **might** only be valuable in certain circumstances. (TRICLE)

These examples from the concordance lines of four corpora show salient patterns used with concessive *might*. Among these patterns and collocations, the use of *but* is the mostly used one in four corpora. NICLE used it most (n=59) and it was followed by JPICLE (n=17), LOCNESS (n=11), and TRICLE (n=5) sequentially. For the use of *or*, NICLE (n=13), LOCNESS (n=10), JPICLE (n=6), and TRICLE (n=5) used this pattern sequentially in terms of frequency. In this case, NICLE (n=85) is the one that used concessive *might* most with these patterns and collocations. It was followed by JPICLE (n=25), LOCNESS (n=2), and TRICLE (n=11) sequentially.

As another salient pattern used with the epistemic *might*, *if* clauses were preferred in four corpora with high frequencies. Moreover, the meaning of possibility was conveyed with a meaning of condition:

1. **If** a couple did not cohabit, they **might** not experience those consequences. (JPICLE)
2. This **might** have proved possible, **if** the determination was there. (NICLE)

As in the examples (1-2) extracted from JPICLE and LOCNESS, epistemic *might* was used with *if* conditionals. The use of *if* clause with epistemic *might* gives the meaning of possibility with a condition. JPICLE (n=26) preferred this pattern most; it was followed by NICLE (n=25), LOCNESS (n=21), and TRICLE (n=11).

As collocations of epistemic *might*, *also* and *as well* were used in four corpora. However, epistemic *might* with the use of *as well* was not utilized in TRICLE. They were used mostly after the epistemic *might*; however, there are clauses that have them before the epistemic *might*:

1. **Also** in some cases there **might** be something wrong with the foetus, it might be disabled or born with one arm. (NICLE)
2. ...the computer will charge for their duties and they **might also** control the world as it is happening in the films. (TRICLE)
3. To answer this question, we **might as well** begin by deciding who should not. (JPICLE)
4. With all of the hassle students on campus deal with they **might as well** live at home. (LOCNESS)

In these examples (1-2-3-4) extracted from four corpora, the salient collocations of *might* was presented. In this regard, while the use of *also* was found before the epistemic *might*, the use of *as well* was not found before the epistemic *might*. Moreover, NICLE was found as the one using *also* (n=8) and *as well* (n=5) most among the four corpora. On the other hand, LOCNESS (n=2), TRICLE (n=2), and JPICLE (n=1) come after NICLE in terms of frequency of *also*. For the frequency of *as well*, LOCNESS (n=3) and JPICLE (n=2) come sequentially.

In addition, while LOCNESS, JPICLE, and TRICLE did not utilize it, NICLE used the word *even* as a collocation of epistemic *might* with a very high frequency:

1. Some persons **might even** be addicted to television. (NICLE)

As in the example (1) reached from the concordance lines of NICLE, the word *even* was used mostly after the epistemic *might*. NICLE used this collocation with a very high frequency (n=15) whereas the other three corpora did not (Appendix 7).

4.3.4.2. The Case of Epistemic ‘Will’ and ‘Would’

The cases of epistemic modal auxiliaries *will* and *would* were examined through the concordance lines in four corpora in terms of lexico-grammatical patterns, the most salient patterns, and collocations. As a basis, while JPICLE and TRICLE preferred to use epistemic *will* more than epistemic *would*, NICLE and LOCNESS used epistemic *would* more than epistemic *will*. To see the patterns used with the epistemic *will* and *would*, the concordance lines were examined through a qualitative analysis.

The subjects of the epistemic *will* was examined to see if the epistemic *will* was used with the subject *I* and ‘we’ and with the other objective subjects. In this regard, whereas four corpora had the objective epistemic modal *will* most, the number of subjective ones were notable. Four corpora utilized the subject *I* and ‘we’ along with the noun clauses, there, this, it, you, and other subjects:

1. ...if I don't know that difference, **I will** make a big mistake. (JPICLE)
2. If we replace religion with television in the 20th century, **we will** witness the same thing. (NICLE)
3. Who knows, maybe one day this fight will end and both sexes **will** live happily till the end of the days, like in the tales. (TRICLE)
4. ... maybe there **will** be a cure before they die. (LOCNESS)

As in the examples (1-2) reached from JPICLE and NICLE, the visibility of the writer was conveyed via the use of subjects *I* and ‘we’. On the other hand, it was discovered there were mostly cases involving the use of other subjects, demonstratives, the word *there*, and noun clauses to signify the objective epistemic *will* rather than the subjective epistemic modal *will*. Regarding this, JPICLE was found to be the corpus that used subjective modal *will* with the use of *I* (n=18) and ‘we’ (n=52) more than the other three corpora. It was followed by NICLE (n=48), LOCNESS (n=33), and TRICLE (n=27) sequentially. Thus, JPICLE preferred to have more writer’s visibility than the others whereas TRICLE was the one that used the subjective modal *will* least.

When looked at the negation of the epistemic *will*, both positive clauses and the negation of epistemic *will* were traced through the concordance lines. It was discovered that the reduction of *will* as *'ll* was used in four corpora and that the reduction of negation of *will not* was utilized as *won't*. Therefore, both the reductions and full versions of modal *will* were utilized in four corpora:

1. They just get the unnecessary theoretical knowledge most probably **will not** be used after graduation... (TRICLE)
2. However, I think this way **won't** suffice for becoming international humans. (JPICLE)
3. ...the number of cars using it **will** double by 2030. (LOCNESS)
4. By doing a certain job for years you'**ll** be highly qualified. (NICLE)

For the examples (1-2), the negation of epistemic *will* was scrutinized through the concordance lines. Both the reduction and full typing of the epistemic *will* were utilized in four corpora. In this regard, LOCNESS (n=103) was discovered as the corpus used the negation of epistemic *will* most. It was followed by NICLE (n=52), JPICLE (n=46), and TRICLE (n=37). In addition, the reduction of *will* was used in four corpora as *'ll* (3-4). On the other hand, it was revealed that four corpora preferred to use positive clauses with modal *will* to convey the epistemic meaning of *will*.

As one of the salient patterns used in four corpora, if conditional was preferred with the epistemic *will* in four corpora most saliently. Thus, the prediction of the writer was conveyed through the condition of the proposition:

1. **If** they didn't read it then they'**ll**, no doubt, have heard it on the news or from friends. (LOCNESS)
2. **If** real English education come true, Japan **will** be able to develop at important point, I think. (JPICLE)

Regarding the examples (1-2) extracted from LOCNESS and JPICLE, it was discovered that the epistemic use of *will* was utilized with the use of if conditionals. Thus, the prediction in the proposition was supported with a condition in the propositions in four corpora. This pattern was found saliently in four corpora. LOCNESS (n=179) was the one using this pattern more than JPICLE (n=125), TRICLE (n=96), and NICLE (n=81) sequentially.

As for the collocations used with the epistemic *will*, the adverb *also* was used in four corpora. It was found the adverb *also* was used as making the epistemic meaning of *will* more salient in four corpora. In addition, it was used before or after the epistemic *will*:

1. It is a decision that **also will** hurt those close to him , and furthermore will scar the image people have towards adoption. (LOCNESS)
2. **Also**, it **will** be to difficult to decided who students are successful or not. (TRICLE)
3. The face of America is changing, and the sound of it **will also** change. (JPICLE)
4. Breakfast is a meal that is very important for the concentration, and being taught to eat this meal as a child makes it more likely that this **will** happen **also** later in life. (NICLE)

In this regard, the adverb *also* was used in different places in a clause to make the epistemic meaning of *will* more salient in four corpora. Thus, it is the corpus LOCNESS (n=68) that used the collocation of epistemic *will* with the adverb *also* the most among four corpora. It was followed by TRICLE (n=28), NICLE (n=20), and JPICLE (n=11) in sequence.

Another collocation of epistemic *will* is the adverb and adjective *then* used in four corpora. Both meanings of *then* was used in four corpora with the epistemic *will*:

1. If there is "true love," **then** all our needs **will** be satisfied. (LOCNESS)
2. If these trials are successful, the drug **will then** go on to human trial. (TRICLE)
3. **Then** they **will** probably fight for a system they no longer believe in. (NICLE)
4. Most people probably think that I should go straightly home, and **then**, I **will** feel easy. (JPICLE)

As in the examples (1-2-3-4) extracted from the four corpora, the collocation of epistemic *will* was used as *then* in different syntactical clauses. Thus, it can be situated before or after the epistemic *will*. It can be used as both an adverb and adjective as the collocation of epistemic *will*. It is LOCNESS (n=40) that used the adverb and adjective *then* the most among four corpora. It was followed by TRICLE (n=23), JPICLE (n=20), and NICLE (n=19) in sequence.

Furthermore, the other collocation of epistemic *will* was discovered as the adverb *always*. It was used in different places in a clause to make the use of epistemic *will* more salient:

1. There **will always** be some people more alike than others. (NICLE)
2. They **always will** be affected by other people. (TRICLE)
3. Clothes are an important part of our every day lives and they **always will** be. (LOCNESS)
4. Thus Japanese **will always** have to use English. (JPICLE)

As in the examples (1-4) extracted from NICLE and JPICLE, the adverb *always* was used after the epistemic *will* whereas the examples (2-3) present it was also used before the epistemic

will as a collocation. NICLE (n=62) had the adverb *always* as the collocation of epistemic *will* the most; it was followed by TRICLE (n=24), LOCNESS (n=23), and JPICLE (n=3).

Other adverbs used as a collocation of epistemic *will* were found as adverbs *probably* and *never* in four corpora. In this case, the phrase *will probably* was used same as the adverb *likely*. It was revealed both adverbs had different placements in a clause:

1. When a well-educated person can not find a good job for himself, he **will probably** try to go abroad. (TRICLE)
2. If you go to prison it **probably will** not be easy. (NICLE)
3. The supremacy of Parliament **will never** be challenged. (LOCNESS)
4. My play is very poor and I **will never** be professional. (JPICLE)

As in the examples (1-2), the adverb *probably* was used before and after the epistemic *will* as a collocation of the epistemic *will*. NICLE (n=36) is the one using the adverb *probably* as the collocation of epistemic *will* the most. It was followed by TRICLE (n=16), LOCNESS (n=13), and JPICLE (n=10) in sequence. On the other hand, the adverb *never* was usually preferred to be used after the epistemic *will* when used as a collocation of the epistemic *will*. It is LOCNESS (n=30) that used the adverb *never* the most as a collocation of epistemic *will*. It was followed by NICLE (n=19), TRICLE (n=16), and JPICLE (n=7).

Lastly, it was discovered there were other salient collocations of epistemic *will* in four corpora with different frequencies. There were the adverbs *perhaps*, *eventually*, *surely*, and *soon* as the other salient collocations:

1. In the future this number **will perhaps** expand due to the events in Eastern Europe, whereby more countries are turning towards a market economy. (LOCNESS)
2. If you become a part of the prison system it **will eventually** seem like a home to you. (NICLE)
3. **Surely**, it **will** be convenient for us to live in the city. (JPICLE)
4. Of course people **will** take the shelter in the country and shortage **will** appear **soon**. (TRICLE)

These adverbs were used as collocations of epistemic *will* in four corpora with different placements in a clause. There were clauses, such as the example (3), that have adverbs at the beginning of the sentence and before the epistemic *will*. On the other hand, there were clauses, such as the examples (1-2-4), having the adverbs after the epistemic *will*. In this regard, the most salient patterns and collocations used with the epistemic *will* were scrutinized through the concordance lines (Appendix 8).

Similarly, the most salient lexico-grammatical patterns and collocations of epistemic modal auxiliary *would* was examined in four corpora to find out the differences and similarities in terms of the use of epistemic *would*. In this regard, whereas the epistemic *would* indicates a prediction in the proposition, the non-epistemic use of it includes the past simple of *will* and volition along with the phrase *would like to*. In this regard, the epistemic uses of modal auxiliary *would* were examined through the concordance lines in four corpora to see the salient patterns and collocations.

For the subjective and objective epistemic use of modal auxiliary *would*, the four corpora were scrutinized. In this regard, the subjects *I* and ‘we’ were utilized in four corpora to show the writer’s visibility along with the subjective epistemic *would*. On the other hand, the demonstratives, there, other subjects, and noun clauses were used for the objective epistemic *would*:

1. **This would** mean more money or higher taxes. (LOCNESS)
2. To be honest, **the world order would** probably be really shaken. (NICLE)
3. **Student would** be away from studying English. (JPICLE)
4. **No one would** ask how do you get this mark, by studying hard or cheating. (TRICLE)

As in the examples (1-2) from LOCNESS and NICLE, the subjects *I* and ‘we’ were used to convey subjective epistemic *would* leading to the visibility of writer in the proposition. However, other subjects, noun clauses, and demonstratives were also used to indicate the objective epistemic *would* decreasing the visibility of the writer. Regarding this, NICLE (n=104) is the one that used the subject *I* and ‘we’ to convey subjective epistemic *would* the most. It was followed by LOCNESS (n=83), JPICLE (n=67), and TRICLE (n=13) in sequence.

Moreover, the concordance lines showed that there was also the negation of epistemic *would* in four corpora. In addition, the reduction of modal auxiliary *would* was used as ‘*d* in four corpora:

1. If it were not for paper, our present life **would not** function. (JPICLE)
2. Men may think that women **would not** manage to be successful in politics and business world. (TRICLE)
3. I'm sure most people **wouldn't** complain about paying an extra penny in the pound to subsidise a re-Nationalised Service to the country. (LOCNESS)
4. They maybe dreamt so much of it that they'**d** walk around with weapons. (NICLE)

As in the examples (1-2) gathered from JPICLE and TRICLE, the negation of epistemic *would* was used in four corpora without the reduction along with the reduction of it. The example (3) presents there were instances that the negation of epistemic *would* was used in the form of reduction of *not*. In addition to the reduction of negation, epistemic *would* was also used with its reduction as ‘*d* as in the example (4). In this regard, among four corpora, LOCNESS (n=136) is the

one using the negation of epistemic *would* the most. It was followed by NICLE (n=70), TRICLE (n=41), and JPICLE (n=24) sequentially.

Similar to the epistemic modal auxiliary *will*, the epistemic *would* also was used with if conditionals saliently. Thus, the prediction was conveyed via the epistemic *would* in a condition expressed by if clause:

1. **If** I had listened to my father, I **would** not have been at university now and could not have written on this paper. (TRICLE)
2. **If** you were not working, you **would** mind all the time, and you must do. (JPICLE)
3. ...most people **would** be relieved **if** fox hunting was banned. (LOCNESS)
4. Maybe the system **would** be changed **if** there were granted more money. (NICLE)

As in the examples (1-2-3-4) extracted from the four corpora, if clause was used before and after the main clause. It was discovered the epistemic *would* was used with second conditional and third conditional to mention a hypothetical and imaginary situation by giving a prediction. In this regard, LOCNESS (n=346) is the corpus that used if conditionals as a salient pattern with the epistemic *would* the most. Then, NICLE (n=207), JPICLE (n=104), and TRICLE (n=77) comes in sequence.

Another salient pattern used with the epistemic *would* is the conjunction *when* in four corpora. In this regard, the conjunction *would* was used to introduce a completed event and longer event that takes place at the same time along with the meaning of prediction:

1. **When** there was no TV, you'd spend your time different. (NICLE)
2. That **would** become clear **when** we think about juvenile crimes. (JPICLE)

In this case, the conjunction *when* was used as a salient pattern in four corpora with the epistemic *would*. However, if the modal *would* carried the epistemic meaning, the conjunction *when* was used as in the meaning of if conditionals. NICLE (n=60) is the one that used the conjunction *when* as a salient pattern. It was followed by LOCNESS (n=54), TRICLE (n=25), and JPICLE (n=24).

Lastly, for the collocations of epistemic *would*, it was discovered the concordance lines involve the adverbs *probably*, *also*, and *then* as the collocations of epistemic *would*. The adverb *probably* was used with the epistemic *would* similar to the adverb *likely*. The adverb *also* was used with the epistemic *would* to add a prediction to the clause. The adverb *then* was utilized with the epistemic *would* to indicate the order and the time of the clause along with the meaning of *in addition*:

1. This **probably would** require understanding of philosophy & psychology as well as mathematics & science. (LOCNESS)
2. I **would also** say that the modern world that we know today started to "bloom" after WW2. (NICLE)
3. If the abortion is really necessary, **then it would** be no more a sin or an unethical operation. (TRICLE)

In the examples (1-2-3), it was discovered these adverbs were used before or after the epistemic *would* in four corpora. In addition, while these adverbs as the collocations of epistemic *would* were more salient in LOCNESS and NICLE, they were used few in TRICLE, and JPICLE (Appendix 9).

4.3.4.3. The Case of Epistemic ‘Must’ and ‘Should’

When it comes to the salient lexico-grammatical patterns and collocations of the epistemic modal auxiliary *must*, it was discovered that the non-epistemic (deontic) meaning of the epistemic *must* was mostly preferred in four corpora instead of the epistemic meaning of it. Thus, the concordance lines were examined to find out the salient patterns and collocations that were used with the epistemic modal auxiliary *must*.

It was discovered that the epistemic *must* was mostly used with the verb *to be*. The use of the verb *to be* with the epistemic *must* conveys a conclusion based on an evidence:

1. There **must be** some differences because we all see, perceive the world in different views so our comments and reactions differ. (TRICLE)
2. I 'm sure there **must be** much more problems for such a person. (JPICLE)
3. As far as I can see, the only reason for saving time **must be** a desire to fill this time with something worth while. (NICLE)
4. These women were led to believe that if they were not happy fulfilling this role that something **must be** wrong with them. (LOCNESS)

In the examples (1-2-3-4) extracted from four corpora, the epistemic *must* was used with the verb *to be* saliently in four corpora. This pattern conveys a conclusion based on the evidence. Therefore, the verb *to be* was used as a way to express this epistemic meaning. JPICLE (n=35) is the one that used the verb *to be* as a salient pattern of epistemic *must* the most among the four corpora. Then, NICLE (n=30), TRICLE (n=24), and LOCNESS (n=7) come in sequence.

Another salient pattern that was used with the epistemic *must* was discovered as the pattern *must + have + past participle*. In this case, this pattern conveys the past meaning of epistemic *must* involving a conclusion about the past time based on evidence:

1. Before a new invention can be seen in our physical world, it **must have been** in someone's head as an idea. (NICLE)
2. All this paper **must have been** used for something. (LOCNESS)
3. ...according to majority of the students, people, and also some instructors, plagiarism and cheating are normal for the students because everyone **must have copied** in schooling years... (TRICLE)
4. ...I felt very ashamed because I knew that they **must have thought** it strange.

As in the examples (1-2-3-4) extracted from the four corpora, the epistemic *must* was used with the pattern *have + past participle* to mention about a conclusion in the past based on the evidence. In this regard, NICLE (n=10) used the pattern *must + have + past participle* and the epistemic modal *must* most saliently. It was followed by LOCNESS (n=5), JPICLE (n=4), and TRICLE (n=2).

The other salient pattern used in four corpora was discovered as the use of epistemic modal *must* and the use of the conjunction *but*. It was used to give a contrastive meaning in the clause with the deduced conclusion in the proposition:

1. Obviously the traits of such people are not wholly genetic, **but** there **must** be some genetic influence in the way their minds work. (LOCNESS)
2. ...there is a great argument in saying that being born **must** be in the best interest of the child, **but** I will have to disagree. (NICLE)
3. Therefore nowadays, most of Japanese **must** have studied English at least for 3 years, **but** most of them can't speak English at all. (JPICLE)
4. It can seem to be cruel, **but** it **must** be, as they can't condemn the people to die in pain. (TRICLE)

As in the examples (1-4) gathered from LOCNESS and TRICLE, it was discovered the conjunction *but* was used before the epistemic modal *must*. Also, there were examples, such as the examples (2-3), that have the conjunction *but* after the epistemic modal *must*. In this regard, an opinion based on the evidence was given within a contrastive clause. TRICLE (n=5) is the one that used the conjunction *but* with the epistemic modal *must* the most. After that, JPICLE (n=4), NICLE (n=3), and LOCNESS (n=2) come in sequence.

Lastly, the negation of epistemic modal *must* was used in TRICLE; however, it was not used in the other three corpora. Thus, the negation of *must* was utilized in its epistemic meaning only in TRICLE:

1. Therefore the universities claim that they have not got enough capital, it **must not** be a reason. (TRICLE)

As in the example (1) extracted from TRICLE, the negation of *must* was used with its epistemic meaning involving a conclusion with a negative form. TRICLE (n=7) is the only one that used the negation of *must* with its epistemic meaning (Appendix 10).

As for the salient lexico-grammatical patterns and collocations of the epistemic modal auxiliary *should*, it was mostly used with its non-epistemic meaning (deontic) in four corpora. The epistemic meaning of *should* involves the prediction and probability. Thus, the salient patterns and collocations of the epistemic modal *should* was examined through the concordance lines of four corpora:

It was revealed most of the clauses with the epistemic modal *should* were used with the verb *to be* for conveying a prediction and a probability through the proposition in four corpora:

1. They **should be** out enjoying themselves and gaining experiences for themselves instead of reading about them on a flat screen. (LOCNESS)
2. It **should be** interesting. (JPICLE)
3. It **should be** easier for them to function normally in society, and thus stay away from crime. (NICLE)
4. So, there **should be** another ways of making lessons more attractive and easily understandable. (TRICLE)

As in the examples (1-2-3-4) extracted from the four corpora, the epistemic *should* was used with the verb *to be* saliently. This pattern was used for conveying the prediction and probability in a clause. It is TRICLE (n=14) that used the verb *to be* with the epistemic modal *should* the most. It was followed by LOCNESS (n=13), JPICLE (n=11), and NICLE (n=7) in sequence.

Another salient pattern used with the epistemic modal *should* was discovered as the pattern *should + have + past participle*. It was used to mention a prediction or a probability in the past:

1. I still believe she **should have been forced** to go through with the pregnancy. (LOCNESS)
2. Maybe there **should have been** a requirement before going to a teaching college to have a few years in between high school and college? (NICLE)

3. It **should have been** a wonderful feelings. (TRICLE)

In this regard, the epistemic *should* was used as the pattern *should + have + past participle*. This pattern was used to mention the prediction about the past event. While JPICLE did not use any of this pattern, NICLE (n=2), TRICLE (n=2), and LOCNESS (n=1) used this pattern few because of the limited number of the use of epistemic modal *should*.

As one of the collocations used with the epistemic modal *should*, the conjunction *but* was used in four corpora. The conjunction *but* was used to mention a contrast in the clause along with the use of epistemic modal *should* involving a prediction:

1. I don't know who to educate them **but** there **should** be a way. (JPICLE)
2. These slight obstacles will always be shared by some patients, **but should** be easily tolerable compared to the past depression many have suffered through. (LOCNESS)
3. In such sensitive issues, involving people's personal freedom and value, it **should** be easy, **but** it's not. (NICLE)
4. I know that having a job is a part of the real life, **but** there **should** be some other things in the meaning of the university. (TRICLE)

As in the examples (1-2-3-4) extracted from the four corpora, the conjunction *but* was used with the epistemic modal *should* as a salient pattern. It was used before or after the epistemic modal auxiliary *should* in four corpora. Regarding this, TRICLE (n=4) and LOCNESS (n=4) used the salient pattern epistemic modal *should* and the conjunction *but* together the most. It was followed by JPICLE (n=2) and NICLE (n=1).

The other salient pattern that was used with the epistemic modal auxiliary *should* was found as the use of if conditionals. The prediction was conveyed via epistemic modal *should* with a condition in the proposition:

1. **If** somebody is to ready to do such a terrible thing, there **should** be incredibly huge reasons to force him to this. (TRICLE)
2. **If** we work for the time, then volunteering **should** be easy, but it's not. (JPICLE)
3. However **if** the monitors are of better quality, this **should** not be a problem. (LOCNESS)

As in the examples (1-2-3) reached from the three corpora, if conditionals were used with the epistemic modal auxiliary *should*. A prediction on a situation was given within a condition. In this regard, LOCNESS (n=4) and TRICLE (n=4) used this pattern the most. It was followed by JPICLE (n=3) and NICLE (n=2) sequentially.

Lastly, the negation of *should* was used with its epistemic meaning in four corpora. However, the positive clauses were mostly preferred in four corpora instead of the negation of epistemic modal *should*:

1. To experience a life sentence in prison **should not** be a life longing for. (NICLE)
2. The thought that her husband is not dismissed is equal illusion in 20 century, we **shouldn't** be able to be made peaceful. (JPICLE)
3. However if the monitors are of better quality, this **should not** be a problem. (LOCNESS)
4. There **should not** be time for them to stand lazily. (TRICLE)

As in these examples (1-2-3-4) extracted from the four corpora, it was discovered the negation of the epistemic modal *should* was used in four corpora. However, it was used less than the positive clause of the epistemic modal *should*. In this regard, TRICLE (n=4) used the negation of epistemic modal *should* the most. It was followed by NICLE (n=2), JPICLE (n=2), and LOCNESS (n=2) (Appendix 11).

4.3.4.4. The Case of Epistemic 'Could'

For the typical cases of epistemic modal *could*, the concordance lines of four corpora were examined to discover the salient lexico-grammatical patterns and collocations used with the epistemic modal *could*. Concerning this, one of the salient patterns used with the epistemic modal *could* was found as the verb *to be* conveying the possibility in the proposition:

1. There is the chance that this is done purposefully ie biological warfare in which the effects **could be** catastrophic. (LOCNESS)
2. Another issue is that using cell phones **could be** a problem of memorizing things for human. (JPICLE)
3. The tutors could may be better in asking what the students mean at the same time as the students also **could be** better in telling the tutors what they mean. (NICLE)
4. We know that in some past societies, tradition of family **could be** based on matriarchal family. (TRICLE)

Regarding the examples (1-2-3-4) extracted from the four corpora. The epistemic modal auxiliary *could* was used with the verb *to be*. This pattern involves a possibility in the proposition. LOCNESS (n=186) was the one that used this pattern the most. It was followed by NICLE (n=66), TRICLE (n=25), and JPICLE (n=21).

Another salient pattern used with the epistemic modal *could* was found as the pattern *could + have + past participle*. It was used to convey a possibility in the past:

1. When my uncle walked away after words were exchanged, the acquaintance **could have done** the same to solve the confusion between the two of them. (LOCNESS)
2. Her mother regretted that if she could know that she had a problem with her heart earlier then her daughter **could have recovered** earlier. (JPICLE)
3. Another solution **could have been** three years of theoretical studies in college and than a whole year of practise. (NICLE)
4. May be it **could have been used** in beneficial of humanity but unfortunately everything human being create is due to eliminate themselves. (TRICLE)

As in the examples (1-2-3-4) gathered from the four corpora, the epistemic modal auxiliary *could* was used with the pattern *have + past participle*. This pattern enabled the writer to convey the possibility about past events. This pattern was used by NICLE (n=30) the most among the four corpora. Then, LOCNESS (n=25), TRICLE (n=5), and JPICLE (n=2) comes in sequence.

For another pattern that was used with the epistemic modal *could*, it was discovered that the conjunction *but* was used for conveying the possibility through a contraction:

1. Sure, there may a few cases where that may happen, **but** that **could** happen anywhere. (LOCNESS)
2. It is nothing **but** a racism that **could** bring about hate crime. (JPICLE)
3. **But** now, in this century we just **could** n't have managed to live without it. (NICLE)
4. **But** educational establishment **could** use from hiring international students at the university. (TRICLE)

The examples (1-2-3-4) gathered from the four corpora present the use of the conjunction *but* with the epistemic modal auxiliary *could* as a salient pattern. Thus, LOCNESS (n=57) is the one that used this pattern the most; it was followed by NICLE (n=30), TRICLE (n=13), and JPICLE (n=9) in sequence.

Furthermore, one of the salient patterns used in for corpora regarding the epistemic modal *could* was discovered as the use of if conditional with the epistemic modal *could*. In this case, the possibility in the proposition was conveyed within a condition indicated in the clause:

1. Television **could** be a better informal medium **if** it was used the right way. (NICLE)
2. **If** he does not have a good relationship with journalist his career **could** end in the team. (TRICLE)
3. **If** they started learning English early, they **could** be good at English. (JPICLE)

4. The welfare problem **could** easily be solved if both parties would quit butting heads for power. (LOCNESS)

Regarding the examples (1-2-3-4) extracted from the four corpora, the possibility of a situation was given within a condition in the clause in four corpora by using if conditionals and the epistemic modal *could* together as a salient pattern. This pattern was used by LOCNESS (n=74) the most among the four corpora. It was followed by NICLE (n=42), JPICLE (n=29) and TRICLE (n=20) in sequence.

In addition, there were two collocations of the epistemic modal *could* that were mostly used by LOCNESS although the other three corpora used them, too. These collocations were found as adverbs *also* and *then*. The adverb *also* was used with the epistemic modal *could* to convey the possibility as an addition to the other proposition. For the adverb *then*, it was used to mention about the possibility in an order and time along with the meaning of the phrase *in addition*:

1. This book **could also** be labeled as a satire, farce or even a picaresque novel. (LOCNESS)
2. It **could also** make the inmates feel more secure of themselves. (NICLE)
3. Should the attitudes of the people involved in this persevere or get stronger however, **then** there **could** be some major problems ahead. (LOCNESS)

As in the examples (1-2) extracted from LOCNESS and NICLE, the adverb *also* was used with the epistemic modal *could* to express possibility in addition to another proposition. For the example (3) from LOCNESS, it was found out the adverb *then* was used with the epistemic modal *could* to convey the possibility within an order, time, or in the meaning of the phrase *in addition*. Therefore, LOCNESS (n=23 for also, n=24 for then) was used these salient patterns the most among the four corpora. On the other hand, these patterns were used few in the other three corpora.

Lastly, the negation of the epistemic modal *could* was used in four corpora; however, the positive clause for the use of epistemic modal *could* was preferred more than the negation of it in four corpora:

1. We can't allow ourselves to give up just because we think it **couldn't** possibly be any better. (NICLE)
2. In this viewpoint, a person **could not** chose to die, and that everything must be done to preserve life. (TRICLE)
3. A person **could** not be smart enough to pass the second grade, but as long as he can play ball he is going to be rich. (LOCNESS)

As in the examples (1-2-3), the negation of epistemic modal *could* was used for conveying the negation of a possibility. In this regard, it was used mostly by LOCNESS (n=21). It was followed by JPICLE (n=14), TRICLE (n=12), and NICLE (n=9) in sequence. However, the epistemic modal *could* was mostly used within a positive clause instead of a negative one (Appendix 12).

4.4. Epistemic Devices: Lexical Verbs

4.4.1. Frequency of Epistemic Lexical Verbs

The top ten ELVs used in four corpora were investigated (Table 30). According to results, the lexical verbs *think* is the most used verb in the three corpora except LOCNESS. Differently, JPICLE has the most frequency with the verb *think* (52,82) whereas LOCNESS has the lowest frequency with the verb *think* (7,76). The verb *think* is used as 25,55 for NICLE and 20,41 for TRICLE in the first ranks. On the other hand, the epistemic lexical verb *show* was used in the first rank in LOCNESS (9,2) while it is used lower in TRICLE (3,59), NICLE (3,45), and JPICLE (2,55). Moreover, the epistemic lexical verb *believe* was utilized in the second rank of LOCNESS (8,95) and JPICLE (4,62). Although the verb *believe* was used in the third rank in NICLE and TRICLE, it was used more than JPICLE in NICLE (7,86) and TRICLE (5,72). As for the epistemic lexical verb *seem*, NICLE (10,46) used it in the second rank while it is used in LOCNESS (8,12) and JPICLE (4,14) in the third rank and it is used with lesser frequencies in TRICLE (3,45) in the fifth rank. Moreover, in the second rank of TRICLE (6,41), there was the verb *know/n* although it was used in NICLE (4,79) and JPICLE (3,92) in the fourth rank and in LOCNESS in the sixth rank (4,07). On the other hand, the epistemic verb *feel* was used in the fifth rank in LOCNESS (5,93) and NICLE (3,97), it was used in the tenth rank of JPICLE (1,36) and not showed up in the top ten of TRICLE.

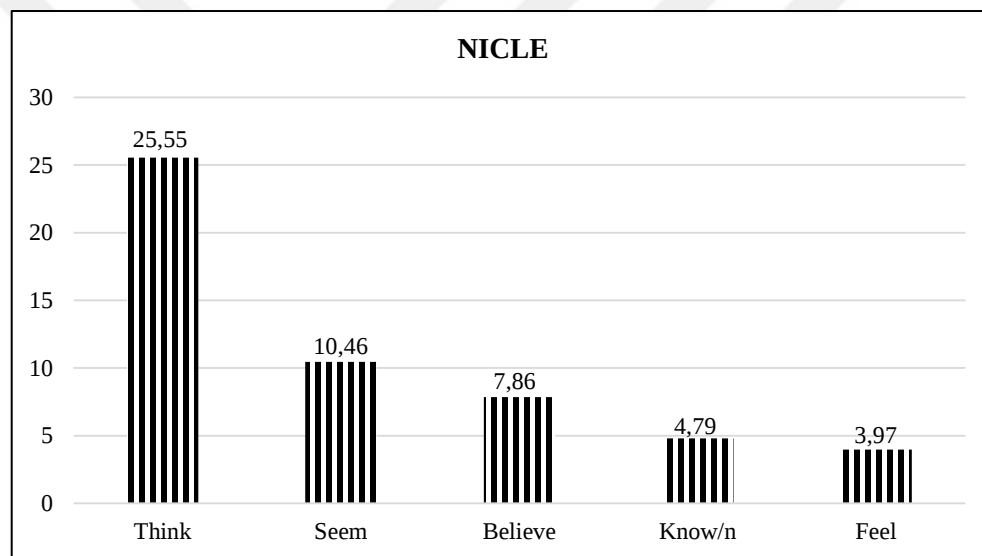
Table 30: Top Ten Epistemic Lexical Verbs in Four Corpora

LOCNESS		NICLE		JPICLE		TRICLE	
Verb	<i>nf</i>	Verb	<i>nf</i>	Verb	<i>nf</i>	Verb	<i>nf</i>
Show	9,2	Think	25,55	Think	52,82	Think	20,41
Believe	8,95	Seem	10,46	Believe	4,62	Know/n	6,41
Seem	8,12	Believe	7,86	Seem	4,14	Believe	5,72
Think	7,76	Know/n	4,79	Know/n	3,92	Show	3,59
Feel	5,93	Feel	3,97	Tend	2,86	Seem	3,45
Know/n	4,07	Suppose	3,76	Show	2,55	Consider	2,95
Argue	3,52	Claim	3,63	Hope	1,8	Claim	2,54
Prove	2,96	Show	3,45	Consider	1,54	Argue	1,45
Consider	2,66	Argue	3,02	Suppose	1,49	Suppose	1,4
Realize	2,46	Look (like)	2,59	Feel	1,36	Expect	0,9

Last but not least, although LOCNESS used ELVs in the top ten list with near frequencies without huge gaps between the frequencies of these verbs, NICLE, TRICLE, and JPICLE more favored using specific ELVs with higher frequencies as it can be seen in the use of epistemic verbs *think*. The other lexical verbs that were subjected to the investigation were mentioned in Appendix 5.

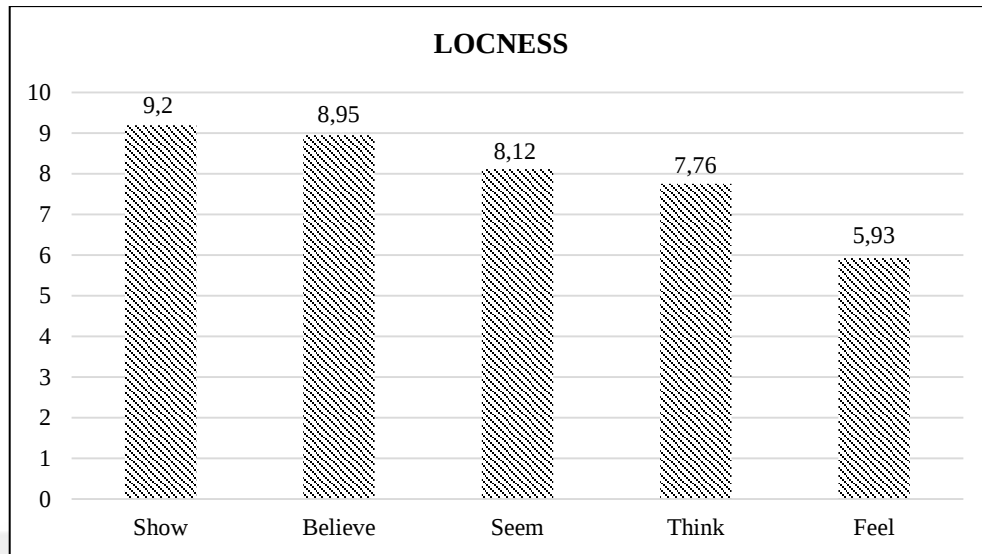
In LOCNESS, the most used epistemic lexical verb is *show* (Figure 10). It is followed by the epistemic lexical verbs *believe*, *seem*, *think*, and *feel*. As for the frequencies, it can be seen while the frequencies of epistemic verbs *show*, *believe*, *seem*, and *think* are high and near, the frequency of the epistemic verb *feel* was used lesser.

Figure 10: Top Five Epistemic Lexical Verbs in NICLE



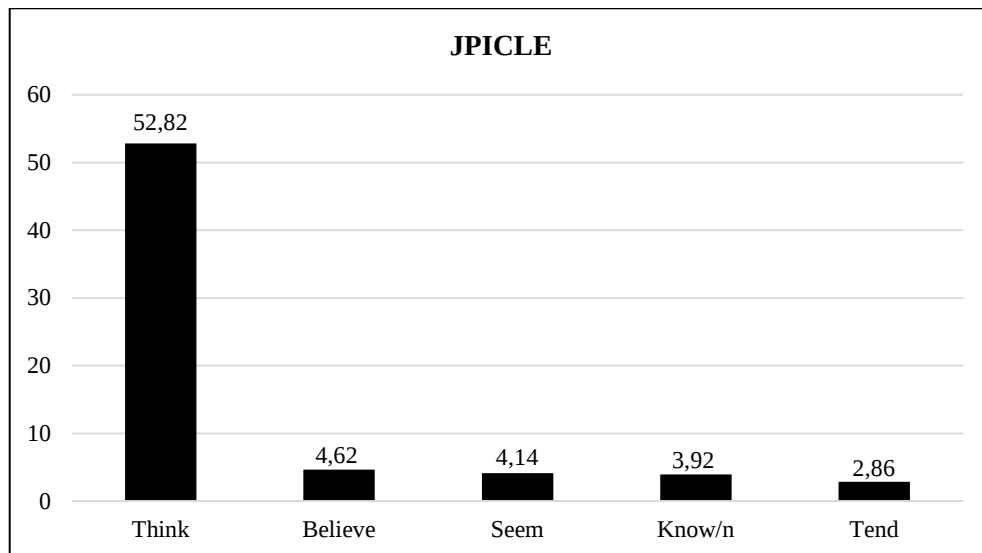
As for the top five epistemic verbs list of the NICLE, the epistemic verb *think* was used mostly with a huge gap from the other four epistemic verbs (Figure 11). Unlike LOCNESS, although the epistemic verb *think* is used more frequently, the other epistemic verbs were not used as much as the epistemic verb *think* in NICLE. Moreover, the epistemic verb *seem* was used more than the verbs *believe*, *know/n*, and *feel*; on the other hand, the frequencies of these three verbs *know/n* and *feel* are rather low and nearly similar.

Figure 11: Top Five Epistemic Lexical Verbs in LOCNESS



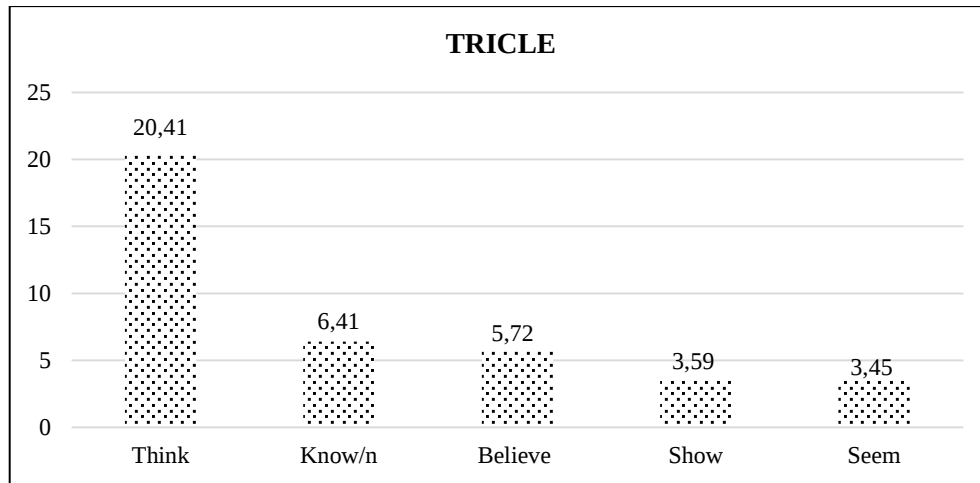
Similarly, in the JPICLE, the epistemic lexical verb *think* is the mostly used epistemic verb (Figure 12). It is followed by the epistemic verbs *believe*, *seem*, *know/n*, and *tend*. Nevertheless, the epistemic lexical verbs *believe*, *know/n*, *tend*, and *show* were used rather less. There is a huge gap between the frequency of *think* and the frequencies of these three verbs.

Figure 12: Top Five Epistemic Lexical Verbs in JPICLE



The same situation can be identified in the TRICLE (Figure 13). The mostly used epistemic verb is *think* in TRICLE. Subsequently, it is followed by *know/n*, *believe*, *show*, and *seem*. The epistemic verb *think* is used more than the other verbs; the frequencies of these other three epistemic verbs are lower than the verb *think*. In this regard, there was a huge gap between the use of epistemic *think* and the other three verbs.

Figure 13: Top Five Epistemic Lexical Verbs in TRICLE



4.4.2. Epistemic Strengths of Epistemic Lexical Verbs

There are categories for epistemic strength, such as weak, medium, and strong. It is adapted from Ngula's (2015) three-way classification. According to this classification, the ELVs were categorized into strong, medium, and weak according to their commitment and probability degrees to investigate the epistemic strength used in four corpora. For example, whereas the epistemic verb *guess* indicates weak epistemic strength, the epistemic verb *claim* shows a strong degree of epistemic strength related to their probability degrees (Table 31).

Table 31: Epistemic Strengths of Epistemic Lexical Verbs

Weak	Medium	Strong
Appear	Argue	Attest
Doubt	Assume	Assure
Guess	Believe	Claim
Hope	Consider	Convince
Look as if	Estimate	Know/n
Look (like)	Expect	Show
Seem	Feel	
Speculate	Indicate	
Propose	Presume	
	Reckon	
	Suppose	
	Tend	
	Think	
	Infer	

For the frequency distribution of ELVs indicating weak epistemic strength, the epistemic verbs indicating weak epistemic strength were used more in LOCNESS (Table 32). It is followed

by NICLE, JPICLE, and TRICLE. However, TRICLE has the lowest frequency (4,34) utilizing weak epistemic strength. When TRICLE is compared with the other three corpora, it can be suggested that while LOCNESS and NICLE and JPICLE are utilizing weak epistemic strength to some degree, TRICLE may refrain from using the ELVs indicating weak epistemic strength.

Table 32: Frequency Distribution of Weak Epistemic Lexical Verbs in Four Corpora

	LOCNESS	JPICLE	NICLE	TRICLE
	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>
Weak				
Appear	1,77	0,13	0,43	0,09
Doubt	0,13	0,61	0,25	0
Guess	0,19	0,66	0,99	0,27
Hope	0,69	1,8	0,77	0,18
Look as if	0,02	0	0	0
Look (like)	0,63	1,19	2,59	0,27
Seem	8,12	4,14	10,46	3,45
Speculate	0,02	0	0,08	0,04
Propose	0,47	0,39	0,08	0,04
Total	12,04	8,92	15,65	4,34

Furthermore, the most used epistemic verb indicating weak epistemic strength in LOCNESS, NICLE, JPICLE, and TRICLE is *seem*. It is followed by the epistemic verb *appear* in LOCNESS while it is the verb *look* in NICLE. On the other hand, in the JPICLE, the second most frequently used epistemic lexical verb indicating weak epistemic strength is *hope* (1,67) whereas it was the verb *look* in TRICLE.

When it comes to the frequency distribution of medium epistemic strength, the medium epistemic strength is used the most in the JPICLE (67,26) (Table 33). It is followed by the NICLE (49,92). Also, LOCNESS (34) and TRICLE (34,97) have utilized ELVs indicating medium epistemic strength with nearly similar frequencies. Furthermore, the mostly used epistemic verb indicating medium epistemic strength is *think* in NICLE, JPICLE, and TRICLE whereas it is the verb *believe* in LOCNESS. On the other hand, there is a huge gap between the frequencies of the first and the second verb indicating the medium degree of epistemic strength in NICLE, JPICLE, and TRICLE.

Table 33: Frequency Distribution of Medium Epistemic Lexical Verbs in Four Corpora

	LOCNESS	JPICLE	NICLE	TRICLE
	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>
Medium				
Argue	3,52	0,83	3,02	1,45
Assume	1,1	0,66	0,3	0,31
Believe	8,95	4,62	7,86	5,72
Consider	2,66	1,54	2,03	2,95
Estimate	0,05	0,04	0	0,09
Expect	1,13	0,79	0,77	0,9
Feel	5,93	1,36	3,97	0,81
Indicate	0,27	0,17	0,51	0,31
Presume	0,08	0,04	0,12	0
Reckon	0,02	0,04	0	0
Suppose	1,1	1,49	3,76	1,4
Tend	1,35	2,86	2,03	0,54
Think	7,76	52,82	25,55	20,41
Infer	0,08	0	0	0,08
Total	34	67,26	49,92	34,97

As for the frequency distribution of strong epistemic lexical verbs, LOCNESS (15,95) used more ELVs indicating strong epistemic strength (Table 34). It is followed by TRICLE (12,71), NICLE (12,29), and JPICLE (6,77) subsequently. In addition, the most used epistemic verb indicating a strong degree of epistemic strength in LOCNESS is *show* (9,2) followed by *know/n* (4,07). On the other hand, the most used epistemic verb is *know/n* in NICLE (4,79), JPICLE (3,92), and TRICLE (6,41). While the epistemic verb *show* is in the second rank indicating strong epistemic strength in JPICLE (2,55) and TRICLE (3,59) corpora, the second one is *claim* in the NICLE (3,63).

Table 34: Frequency Distribution of Strong Epistemic Lexical Verbs in Four Corpora

	LOCNESS	JPICLE	NICLE	TRICLE
	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>
Strong				
Attest	0	0	0	0
Assure	0,05	0	0,04	0,04
Claim	2,27	0,17	3,63	2,54
Convince	0,36	0,13	0,38	0,13
Know/n	4,07	3,92	4,79	6,41
Show	9,2	2,55	3,45	3,59
Total	15,95	6,77	12,29	12,71

There were differences and similarities in four corpora regarding the use of ELVs showing the degrees of epistemic strength (Table 35). The use of ELVs in terms of a strong degree of epistemic strength is much higher in LOCNESS (15,95) than in the other three corpora. Accordingly, the JPICLE (6,77) has rather fewer use of ELVs expressing a strong degree. In that sense, while LOCNESS are using strong ELVs that indicate more certainty of the writer, JPICLE has less strong ELVs, which means less certainty of the writers. On the other hand, JPICLE (67,26) has a medium degree of epistemic strength much more than its use of weak and strong and the other three corpora. In this corpus, that means Japanese writers use medium strength and avoid expressing both weak and strong certainty in argumentative essays.

Table 35: Frequency Distribution of Epistemic Strength in Four Corpora

Strength of Epistemic Lexical Verbs	NICLE	JPICLE	LOCNESS	TRICLE
Weak	15,65	8,92	12,04	4,34
Medium	49,92	67,26	34	34,97
Strong	12,29	6,77	15,95	12,71
Total	77,86	82,95	61,99	52,02

Similarly, in TRICLE and NICLE, there is less use of ELVs expressing a strong and weak degree of epistemic strength, but they have more medium strength of ELVs. In LOCNESS (34), the ELVs in medium strength are used more than weak and strong ones. However, unlike these three corpora, LOCNESS uses weak and strong ELVs more than JPICLE and TRICLE. It can be claimed that although these NICLE, JPICLE, and TRICLE avoid expressing both strong and weak certainty and they tend to stay in the middle by using medium epistemic strength, the English corpus uses the weak and strong expression as much as medium expression. So, instead of avoiding using weak and strong ELVs, the writers in LOCNESS seem to use them, when necessary, unlike NICLE, JPICLE, and TRICLE.

The Log-Likelihood tests for epistemic strength in NICLE and JPICLE indicate that there was not found any statistically significant difference between these two corpora in terms of their use of the weak (LL=1.87, p-value=0.14), medium (LL=2.58, p-value=0.09), and strong (LL=1.62, p-value=0.15) epistemic strengths (Table 36). Nevertheless, JPICLE had more ELVs indicating medium epistemic strength than NICLE. For the weak and strong epistemic strengths, NICLE preferred them more than JPICLE.

Table 36: Log-Likelihood Tests for Epistemic Strength in NICLE and JPICLE

Epistemic Strength	NICLE	JPICLE	LL		p-value
Weak	15.65	8.92	1.87	+	0.14
Medium	49.92	67.26	2.58	-	0.09
Strong	12.29	6.77	1.62	+	0.15

Differently, there was not found any statistically significant difference between the NICLE and LOCNESS related to weak (LL=0.47, p-value = 0.56), medium (LL=3.04, p-value = 0.09), and strong (LL=0.48, p-value = 0.56) epistemic strengths (Table 37). However, LOCNESS utilized more ELVs having strong and weak epistemic strengths than the NICLE; NICLE has more medium epistemic strength than LOCNESS.

Table 37: Log-Likelihood Tests for Epistemic Strength in NICLE and LOCNESS

Epistemic Strength	NICLE	LOCNESS	LL		p-value
Weak	15.65	12.04	0.47	-	0.56
Medium	49.92	34	3.04	+	0.09
Strong	12.29	15.95	0.48	-	0.56

In terms of NICLE and TRICLE, it was revealed there is a statistically significant difference between them in terms of weak (LL=6.79, p-value=0.00) epistemic strength (Table 38). On the other hand, no statistically significant difference was found between these two corpora regarding medium strengths (LL=2.65, p-value=0.09) and strong epistemic strengths (LL=0.01, p-value=1). On the other hand, strong epistemic strength was utilized nearly with the same frequencies in both corpora while NICLE used more weak and medium epistemic strengths than TRICLE.

Table 38: Log-Likelihood Tests for Epistemic Strength in NICLE and TRICLE

Epistemic Strength	NICLE	TRICLE	LL		p-value
Weak	15.65	4.34	6.79	+	0.00
Medium	49.92	34.97	2.65	+	0.09
Strong	12.29	12.71	0.01	-	1

According to the JPICLE and LOCNESS corpora's log-likelihood results, there is a statistically significant difference between these two corpora in terms of medium (LL=11.13, p-value=0.00) and strong (LL=3.82, p-value=0.04) epistemic strengths (Table 39). LOCNESS utilized more epistemic verbs indicating weak and strong epistemic strengths than JPICLE. Nonetheless, JPICLE used more medium epistemic strength than LOCNESS.

Table 39: Log-Likelihood Tests for Epistemic Strength in JPICLE and LOCNESS

Epistemic Strength	JPICLE	LOCNESS	LL		p-value
Weak	8.92	12.04	0.47	-	0.36
Medium	67.26	34	11.13	+	0.00
Strong	6.77	15.95	3.82	-	0.04

When it comes to the log-likelihood results of JPICLE and TRICLE, the results suggested that there is a statistically significant difference between these two corpora in terms of medium epistemic strength (LL=10.38, p-value=0.00) (Table 40). JPICLE used more weak and medium

epistemic strength than TRICLE. In addition, TRICLE utilized more strong epistemic strength than JPICLE. However, there was not found any statistically significant difference between these corpora in terms of weak (LL=1.61, p-value=0.24) and strong (LL=1.84, p-value=0.15) epistemic strengths.

Table 40: Log-Likelihood Tests for Epistemic Strength in JPICLE and TRICLE

Epistemic Strength	JPICLE	TRICLE	LL		p-value
Weak	8.92	4.34	1.61	+	0.24
Medium	67.26	34.97	10.38	+	0.00
Strong	6.77	12.71	1.84	-	0.15

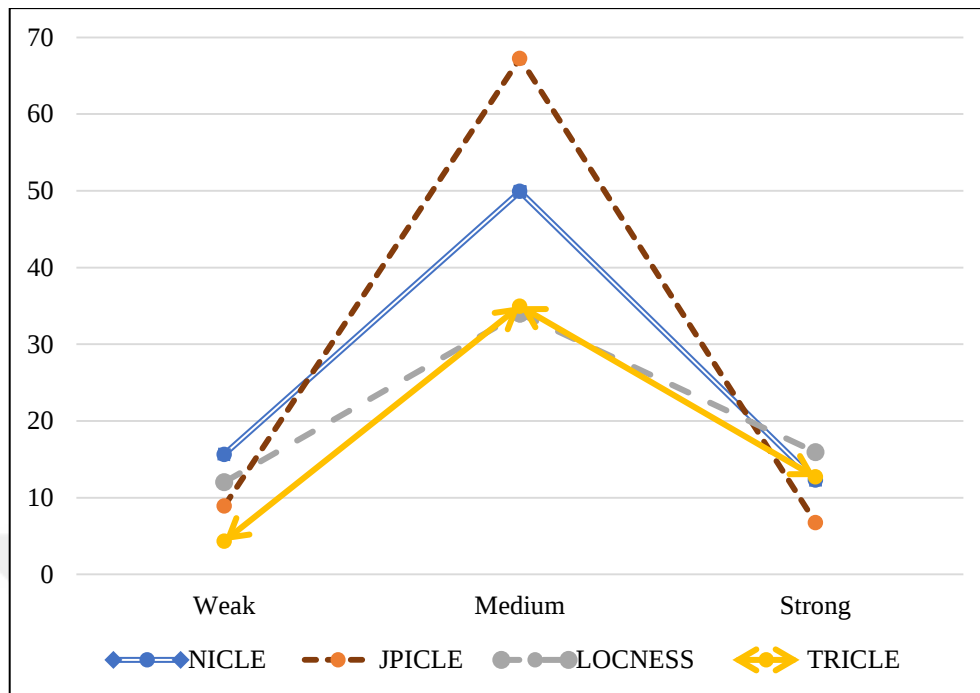
Lastly, it was found that there is a statistically significant difference between LOCNESS and TRICLE in terms of weak epistemic strength (Table 41). LOCNESS utilized more weak epistemic strength than TRICLE (LL=3.77, p-value=0.04). Although LOCNESS also used more strong (LL=0.37, p-value=0.56) epistemic strength than TRICLE and used nearly the same medium (LL=0.01, p-value=1) epistemic strength, there was not found any statistically significant difference between them.

Table 41: Log-Likelihood Tests for Epistemic Strength in LOCNESS and TRICLE

Epistemic Strength	LOCNESS	TRICLE	LL		p-value
Weak	12.04	4.34	3.77	+	0.04
Medium	34	34.97	0.01	+	1
Strong	15.95	12.71	0.37	+	0.56

The use of epistemic strengths indicated by lexical verbs showed similarities and differences among the four corpora (Figure 14). While LOCNESS and TRICLE showed similarities in terms of the use of epistemic strengths, JPICLE overrepresented the medium epistemic strength. JPICLE used weak and strong epistemic strengths less than LOCNESS. On the other hand, NICLE also used medium epistemic strengths more than LOCNESS and TRICLE.

Figure 14: Overall Frequency of Epistemic Strengths of Lexical Verbs in Four Corpora



As for the epistemic lexical verbs that are not used in four corpora, the results showed that there is not any ELVs that are not used in LOCNESS, which mean every epistemic lexical verb was used in LOCNESS. In NICLE, the epistemic lexical verbs *estimate*, *infer*, *look as if*, and *suspect* were not used. The verb *look as if* was not used also in the JPICL and TRICLE. Also, the verb *infer* was not used in NICLE and JPICL, but it was used in TRICLE. Whereas JPICL used the verb *suspect*, this verb was not used in both NICLE and TRICLE. For the verb *assure*, JPICL and TRICLE did not use the verb *assure* while NICLE used this verb. The other verbs that were not used in JPICL are *demonstrate*, *highlight*, *indicate*, and *speculate*. However, these verbs were used in both NICLE and TRICLE. Lastly, the verbs that were not used in TRICLE are identified as *confirm*, *identify*, *interpret*, *presume*, *reckon*, and *suspect*. While these verbs were not used in TRICLE, they were used in both JPICL and NICLE (Table 42).

Table 42: Epistemic Lexical Verbs that are not used in Four Corpora

NICLE	JPICL	TRICLE
Estimate	Assure	Look as if
Infer	Highlight	Presume
Look as if	Infer	Reckon
Suspect	Look as if	Suspect
	Speculate	

4.4.3. Qualitative Analysis of Lexical Verbs' Elimination of Epistemic Uses from the Non-Epistemic Uses

Lexical verbs had both epistemic and non-epistemic meanings in four corpora; therefore, a qualitative analysis was conducted through the concordance lines in four corpora to eliminate the epistemic uses of lexical verbs from the non-epistemic uses. The epistemic uses of lexical verbs were examined by taking the description of Hyland and Milton (1997: 190) into consideration for epistemic lexical verbs "signaling the writer's commitment to a proposition" other than the descriptions of Vold (2006), Coates (2015), Karahan (2017), Ngula (2015), and other scholars.

As one of lexical verbs that was examined through the concordance lines in four corpora, the lexical verb *appear* was subjected to a qualitative analysis to see whether it was used as epistemic or not. It was discovered that the epistemic meaning of lexical verb *appear* can be used to mean the verb *seem*. On the other hand, non-epistemic meaning of *appear* can be utilized to mean to be present and come into sight:

1. When a character is shot, he often **appears** in the next scene. (LOCNESS)
2. But even though on the surface racism **appears** to be much less prominent, I think unluckily this is naive. (LOCNESS)

The examples from the concordance lines of LOCNESS show both the epistemic and non-epistemic meaning of the lexical verb *appear*. In the first example (1), the lexical verb *appear* was used with its non-epistemic meaning involving the meaning of coming into sight. However, the latter example (2) presents the epistemic meaning of the lexical verb *appear* involving the meaning of the verb *seem*.

Moreover, the lexical verb *argue* was used with epistemic and non-epistemic meanings in four corpora. The epistemic lexical verb *argue* means to disagree with something and express an opinion while non-epistemic meaning of it can be to quarrel:

1. It's hard to **argue** with these statements. (NICLE)
2. However, the heatedly **argued** welfare issue can be solved by combining the ideas of both parties. (LOCNESS)
3. Many **argue** that the death capital punishment does not deter. (TRICLE)

As in the examples (1-2) from NICLE and LOCNESS, the word *argue* was used with its non-epistemic meaning. The first example presents the non-epistemic lexical verb does not involve the expression of an opinion using *argue* as a main epistemic lexical verb. The second example involves the word *argue* as an adjective not a lexical verb. On the other hand, the last example (3)

shows the lexical verb *argue* was used with its epistemic meaning involving a disagreement and expression of an opinion.

For the lexical verb *believe*, the concordance lines were examined to eliminate the epistemic uses of it from the non-epistemic uses. In this regard, the lexical verb *believe* can be used for both mentioning a belief towards an opinion and a certainty about the existence of something, usually intangible things, along with the meaning of expression of a trust for someone:

1. At the beginning of the play he is seen to **believe** in God and he says. (LOCNESS)
2. Having lots of money is one reason that people do not **believe** in him. (JPICLE)
3. I **believe** it is different. (JPICLE)

The examples (1-2) from LOCNESS and JPICLE indicate that the lexical verb *believe* was used with its non-epistemic meanings including the certainty about the existence of something and trust for someone. However, the last example (3) shows an expression of an opinion and the writer's confidence in the proposition.

The other lexical verb that was used as both epistemic and non-epistemic was found as the lexical verb *claim*. Whereas the non-epistemic meaning of it involves a demand, the epistemic meaning of the lexical verb *claim* function the same as the verb say; it involves saying a fact regardless of whether it is true or not and whether the others believe it is true or not:

1. Today it's difficult to be a family in balance where everybody **claim** their rights. (NICLE)
2. But this must not be seen as a **claim** but as a certainty. (TRICLE)
3. First, I **claim** that the Japan government shouldn't follow in the Chinese government 's pressure. (JPICLE)

The examples (1-2) extracted from NICLE and TRICLE present the non-epistemic uses of lexical verb *claim*. The first example (1) used the lexical verb *claim* in the meaning of demanding something. The second example (2) involved the word *claim* as a noun. The last example (3) used the lexical verb as an epistemic lexical verb involving stating something true regardless of the opinion of the others.

Furthermore, the lexical verb *feel* was examined through the concordance lines in four corpora to see if it was used in its epistemic meaning. The non-epistemic meaning of lexical verb *feel* involves feeling something emotional while the epistemic meaning of it includes an expression of an opinion:

1. The wrong clothes can often make a person **feel** uncomfortable and self-conscious. (LOCNESS)
2. But sometimes people **feel** that their lives are coming to an end and there's nothing else to do. (TRICLE)

The first example (1) shows the use of non-epistemic lexical verb *feel* in LOCNESS. It was used to express experience emotion. The last example (2) involves the use lexical verb *feel* in its epistemic meaning. It involves expression of an opinion. Other than that, there were examples that have the word *feel* used as a noun.

The other lexical verb that was qualitatively examined through the concordance lines in four corpora is the lexical verb *know/n*. It was used with both epistemic and non-epistemic meanings in four corpora. The non-epistemic meaning of it involves being familiar with someone or something whereas the epistemic meaning of it expresses a range of certainty about a piece of information:

1. The two persons get to **know** each other very well as both positive and negative trait of characters will emerge when they live together. (NICLE)
2. We **know** that in the US, it always happens that a company buys the other company. (JPICLE)

In the first example (1), the lexical verb *know* was used with its non-epistemic meaning to express a familiarity with somebody. The next example (2) presents the lexical verb *know* was utilized with its epistemic meaning involving a certainty about a piece of information.

In addition, the lexical verb *show* was utilized in both epistemic meaning and non-epistemic meaning. The epistemic meaning of it expresses proving something although the non-epistemic meaning of it involves several meanings involving making something visible and expressing the feelings by using the actions. Moreover, several propositions were used with the lexical verb *show* to create different non-epistemic phrasal verbs and meanings:

1. The consequences of these long battles, to the children, may not even **show** up until 10 years later. (LOCNESS)
2. But they will, sooner or later, have to deal with it again: Remorse might **show** her face any day. (NICLE)
3. All in all, all of these **show** that animals exist in order to make the human life easier. (TRICLE)

In these examples (1-2) extracted LOCNESS and NICLE, the lexical verb *show* was used with its non-epistemic meanings. The first example (1) was created by using a proposition to create

a phrasal verb *show up* in the meaning of come into sight. The second example (2) presents the non-epistemic use of the lexical verbs *show* in the meaning of making something visible. The last example (3) expresses a proposition that proves something.

As another lexical verb that was subjected to a qualitative analysis to find out the epistemic uses and eliminate them from the non-epistemic uses, the lexical verb *think* was examined through the concordance lines in four corpora:

1. Now, many of us say that they cannot **think** of living without it. (JPICLE)
2. Some people **think** that a university degree opens all doors of real world. (TRICLE)

Regarding the examples, the first example (1) shows the non-epistemic use of the lexical verb *think*. It was used to express an act to consider on something or somebody. On the other hand, the last example (2) involves the use of epistemic lexical verb *think* to express an opinion.

Lastly, all lexical verbs examined in this study were subjected to qualitative analysis; however, only the findings of the top epistemic lexical verbs used in four corpora were discussed in this section.

4.4.4. Cases of Epistemic Lexical Verbs in Four Corpora

The typical cases of epistemic lexical verbs were examined through a qualitative analysis of concordance lines from the four corpora in terms of salient uses, the lexico-grammatical patterns, and collocations used with epistemic lexical. Based on the qualitative analysis, the similarities and differences between four corpora were discussed. Only the top epistemic lexical verbs *think*, *believe*, *show*, *seem*, *know/n*, *claim*, *feel*, *argue*, and *consider* were discussed in terms of their typical cases involving salient lexico-grammatical patterns and collocations.

4.4.4.1. The Case of Epistemic ‘Think’

The cases of epistemic lexical verb *think* were examined through the concordance lines in terms of the lexico-grammatical patterns and collocations used saliently with the epistemic lexical verb *think*. In this regard, the subjects of the epistemic lexical verb *think* was scrutinized in four corpora. Thus, four corpora preferred to use the subjective *I* more than the other subjects or nearly the same with the other subjects. Other than the subject *I*, the subject *we* was used in four corpora to convey the subjective ELVs although its use was few:

1. I **think** the lottery is here to stay and next weekend it will bring joy to many and misery to just a few. (LOCNESS)

2. I **think** Gløshaugen is a lot different than Dragvoll. (NICLE)
3. We usually **think** it is a sin to think of mercy killing, but it is also a sin to keep people so painful if no hope is seen. (JPICLE)
4. We **think** that we know everything, how to communicate, behave, stand by our feet. (TRICLE)

As in the examples (1-2) from LOCNESS and NICLE, the epistemic lexical verb *think* was used with the subject *I* to express the subjectivity in an epistemic clause. Similarly, the examples (3-4) show the subject *we* was used with the epistemic lexical verb *think* to convey the subjectivity in the expression. In this case, four corpora used the subjects *I* and *we* more than the other subjects. JPICLE utilized both the epistemic lexical verb *think* and the subjective *I* and *we* (n=1062) the most among four corpora.

Another pattern that was utilized with the epistemic lexical verb *think* was found as *the subject + think (that) + clause* pattern. Nearly half of the frequency of epistemic lexical verb *think* in four corpora was used with this pattern. Also, this pattern was utilized by using both with and without *that*-clause:

1. He **thinks that** if he can achieve one impossible act, then this will change everything. (LOCNESS)
2. I **think** all the students at the university should get a chance to work out in the real. (NICLE)
3. So, I **think** governments shouldn't have the right to take land away without permission. (JPICLE)
4. I **think that** there are not any students who have not cheated in the exams or if there. (TRICLE)

The examples (1-4) from the concordance lines of LOCNESS and TRICLE present that the pattern *subject + think (that) + clause* was used with *that*. On the other hand, the examples (2-3) from NICLE and JPICLE utilized the epistemic lexical verb *think* without using *that*. For the frequency of the pattern with using *that*, JPICLE (n=460) and TRICLE (n=207) used this pattern with *that* nearly half of their total frequencies. NICLE (n=156) and LOCNESS (n=119) utilized the pattern with *that* less than the half of their total frequencies. When it comes to the pattern without using *that*, it was found in nearly the rest of concordance lines of four corpora. In this regard, the epistemic lexical verb *think* was used mostly with the pattern *subject + think (that) + clause* in four corpora.

The negation of epistemic lexical verb *think* was examined through the concordance lines of four corpora. It was used with reduction as *n't* and without reduction as *not* in four corpora. The negation of epistemic lexical verb *think* conveys a disagreement on subjective level:

1. I do **not think** that there is a certain. (TRICLE)
2. However, I don't **think** that it will be banned... (LOCNESS)

According to the concordance lines (1-2) from TRICLE and LOCNESS, the negation of epistemic lexical verb *think* was used both with reduction and without reduction. However, the epistemic *think* was preferred to be used in a positive clause the most in four corpora. Therefore, the negation of epistemic *think* was utilized few in four corpora. JPICLE (n=82) used it the most; it was followed by NICLE (n=48), LOCNESS (n=18), and TRICLE (n=19) in sequence.

Moreover, the collocation of epistemic lexical verb *think* was found as the adverb *also* in the meaning of *in addition* in four corpora although it was rather few:

1. I **also think** the question is too extensive to simply answer yes or no. (NICLE)
2. Of course I am happy with their presents but now I **think** it is **also** a waste. (JPICLE)
3. **Also**, I **think** that exams do not reflect students real performances. (TRICLE)

As in the examples (1-2-3) from NICLE, JPICLE, and TRICLE, the adverb *also* was utilized with epistemic *think* as a collocation. It was used before and after the epistemic *think* in four corpora. Concerning this, it is NICLE (n=40) that used it the most; it was followed by JPICLE (n=32), TRICLE (n=24), and LOCNESS (n=19) in sequence.

Lastly, the epistemic *think* was used with the contrastive conjunction *but* saliently in four corpora. In this regard, it was used with the epistemic *think* to position the writer's opinion into a contrastive clause indicating contrast:

1. Modernization is good, **but I think** that women should not be pressured into this if this is not for them. (LOCNESS)
2. I myself, too, **thought** that his decision was best, **but** it didn't go well. (JPICLE)

As in the examples (1-2) extracted from LOCNESS and JPICLE, the conjunction *but* was utilized after or before the epistemic *think*. However, the conjunction *but* was usually used before the epistemic *think* in four corpora. JICLE (n=188) was the one that used this pattern the most; it was followed by NICLE (n=87), TRICLE (n=50), and LOCNESS (n=22) (Appendix 13).

4.4.4.2. The Case of Epistemic 'Believe'

As in the case of epistemic lexical verb *believe*, the use of subjective *I* and *we* was examined through the concordance line in four corpora. According to the qualitative analysis, it was found out while LOCNESS used the subjects *I* and *we* less than half of the frequency of the epistemic *believe*, the other three corpora used it as half time as the frequency or more than the frequency. It was used to convey the opinion of the writer in a subjective way:

1. **I believe** that the aspect of ethnicity is clearly portrayed in ethnic literature. (LOCNESS)
2. **I believe** it is these university degrees that mostly get the criticism of being too theoretical. (NICLE)
3. **We believe** we all will die but nobody can know when s/he will die or there is not a chance to say that "Let me die on . . ." as we don't have a choice on the time. (TRICE)

The examples (1-2-3) show the use of the subjects *I* and *we* with the epistemic verb *believe*. Concerning this, it was used to express the writer's visibility while conveying the beliefs of the writer. In four corpora, it was usually collocated with the use of epistemic *believe*. NICLE (n=121) was the one that used this pattern the most. JPICLE (n=71), LOCNESS (n=66), and TRICLE (n=58) come next in sequence. However, except LOCNESS, the other three corpora used the epistemic *believe* with the subjective *I* and *we* nearly half times as the frequency of the epistemic *believe*. It can mean LOCNESS preferred to use objective epistemic *believe* more than the subjective one among the four corpora.

As another salient pattern used with the epistemic *believe*, the use of that-clause with epistemic *believe* was found out. The epistemic *believe* was mostly used with that-clause:

1. So some people **believes that** online education is not being driven by educational needs or values, but by benefit. (JPICLE)
2. For example; the people in Turkey **believe that** women can never play football or men can never knit up. (TRICLE)
3. Furthermore, the French **believed** the work should take place within twenty-four hours. (LOCNESS)
4. **I believe** it's wrong that they all that luxury inside. (NICLE)

The examples (1-2) extracted from JPICLE and TRICLE present the use of that-clause with the epistemic verb *believe*. On the other hand, that-clauses can be also created by not using the word *that* as in the examples (3-4) from LOCNESS and NICLE. According to the qualitative analysis, most of the concordance lines of epistemic *believe* were created with the that-clause

saliently. It is LOCNESS (n=185) that used that-clause with epistemic *believe* the most among four corpora. It was followed by NICLE (n=109), TRICLE (n=108), and JPICLE (n=76) sequentially.

The other salient pattern used with the epistemic verb *believe* was discovered as the use of *it is believed* although it was not used more salient than the pattern with that-clause:

1. **It is believed** that men just provides economical comfort to their family. (TRICLE)
2. Instead of the flag flying in memory of our heritage and our ancestors, **it is now believed** that it flies for racial purposes. (LOCNESS)

The pattern *it was believed* was used in three corpora, except NICLE. Also, it was not used the same saliency as the pattern with that-clause. In this regard, TRICLE (n=9) was the one that used it the most while LOCNESS (n=4) and JPICLE (n=2) follow it in sequence.

Moreover, the other pattern that was used with the epistemic *believe* was found as the conjunction *but*. It was used to convey the contrastive *believe*:

1. We have probably changed our habits over the past hundred years, **but** I don't **believe** that this has changed for the worse. (NICLE)
2. **But** they **believed** that I need a proper Japanese education. (JPICLE)

As in the examples (1-2) reached from NICLE and JPICLE, the conjunction *but* was used with the epistemic *believe* to convey a belief in a contrastive clause indicating a contrast with another clause suggesting the opposite proposition. LOCNESS (n=28) used this pattern the most; it was followed by NICLE (n=23), TRICLE (n=23), and JPICLE (n=18).

Other salient collocations of epistemic verb *believe* were found as the adverbs *also* and *still* in four corpora. The adverb *also* was used to indicate a belief in addition to another thing. The adverb *still* was utilized to mention an ongoing belief in four corpora:

1. They believed that the Christianity or Shinto was right and the religion on the land was wrong, and **also believed** that the word in their own country was more excellent than the word in the land. (JPICLE)
2. **Also**, I **believe** that we can bring up a baby without spending too much money if we can be thrifty. (TRICLE)
3. I **still believe** she should have been forced to go through with the pregnancy. (LOCNESS)
4. But I **still believe** that I have the right to choose not to be a part of such a system. (NICLE)

The examples (1-2) extracted from JPICLE and TRICLE show the use of adverb *also* with the epistemic *believe*. These clauses express a belief in addition to another proposition. On the other hand, the adverb *still* was used with the epistemic *believe* as in the examples (3-4) from LOCNESS and NICLE to convey an ongoing belief of the writer. LOCNESS was the one that used the adverbs *also* (n=16) and *still* (n=14) the most in four corpora.

When it comes to the negation of epistemic verb *believe*, it was used in four corpora; however, to convey the belief of the writer, the positive clauses were preferred mostly in four corpora:

1. Personally, I do **not believe** one person has the right to take away another person's life, even if that person is not fully developed into what he or she will be. (LOCNESS)
2. In conclusion, I **don't believe** that euthanasia is the last choice and so it should be supported. (TRICLE)

As in the examples (1-2) reached from LOCNESS and TRICLE, the negation of epistemic *believe* was used either with reduction or without reduction. In this case, the beliefs of the writers were conveyed mostly via positive clause in four corpora instead of negative ones; the writers' beliefs were mentioned the most rather than the propositions that the writers do not believe. LOCNESS (n=19) used the negation of epistemic *believe* the most while it was followed by TRICLE (n=9), JPICLE (n=6), and NICLE (n=4) (Appendix 14).

4.4.4.3. The Case of Epistemic 'Show'

The case of epistemic lexical verb *show* was examined through the concordance lines of four corpora to see the salient patterns and collocations used with the epistemic *show*. The epistemic verb *show* was mostly used to convey the epistemic meaning involving uncertainty in an evidential reporting by hedging the proposition. In this regard, the epistemic verb *show* was mostly used with the pattern involving that-clause:

1. Third, voting **shows that** people have freedom of choice. (JPICLE)
2. Studies have **shown that** this method has decreased misbehavior and attendance is better because of it (Ryan 295). (LOCNESS)
3. But most of all **show** them there is more to life than being a criminal. (NICLE)
4. Then, they had important places in community and **showed** they were clever and adroit as boys. (TRICLE)

The examples (1-2-3-4) present the use of that-clause with the epistemic *show* in four corpora. It was used to report something that the writer deduces from evidence. That-clause was used with using the word *that* or without it in four corpora. The epistemic verb *show* was used with

that-clause saliently. LOCNESS (n=84) is the one that used this pattern the most among four corpora. It was followed by NICLE (n=47), TRICLE (n=34), and JPICLE (n=16) in sequence.

The other salient pattern used with the epistemic verb *show* was found out as the conjunction *but*. It was used for conveying the report of the writer involving the avoidance of certainty in a contrasting clause regarding another contrastive clause:

1. Perhaps, women don't have the same power **but** it doesn't **show** that they are weak. (TRICLE)
2. It is just an example, **but** it **shows** that to know other culture, its language is quite important as well. (JPICLE)

Regarding the examples (1-2) from TRICLE and JPICLE, the conjunction *but* was used with the epistemic verb *show* to convey the report in a proposition contrasting with another proposition. It was LOCNESS (n=27) that used this pattern the most; NICLE (n=13), TRICLE (n=8), and JPICLE (n=4) come next sequentially.

As another collocation of epistemic verb *show*, the adverb *also* was used saliently in four corpora. The report of the writer based on evidence was expressed in addition to another proposition via the use of adverb *also*:

1. The civil war in Spain in the late 1930s between the fascists led by General Franco and the communist, **showed also** that women took their part of the struggle , likewise in the Vietnam war, which became "a hot potato" for the US. (NICLE)
2. It has **also** been **shown** that people who have certain prejudices in one area change their prejudices when they move to a new area. (LOCNESS)

The examples (1-2) from NICLE and LOCNESS indicate the use of adverb *also* with the epistemic verb *show*. It was used before or after the epistemic verb *show* as in the meaning of *in addition*. LOCNESS (n=22) used this pattern the most among four corpora while NICLE (n=6), TRICLE (n=6), and JPICLE (n=4) come next in sequence.

As for the negation of epistemic verb *show*, it was used in four corpora; however, four corpora preferred to use the epistemic *show* to report the writer's deduction based on evidence mostly in the positive clauses, rather than the negative clauses:

1. Namely, a certificate does **not show** that a graduate is very good at the branch s/he is specialized. (TRICLE)
2. These facts do **not** accurately **show** that prayer lead to moral virtue. (LOCNESS)

Examples (1-2) from TRICLE and LOCNESS present the use of negation of epistemic verb *show*. However, the epistemic *show* was preferred to be used mostly in a positive clause instead of a negative one in four corpora. Therefore, while LOCNESS (n=7) used it the most and TRICLE (n=3) and JPICLE (n=2) come next, NICLE did not use the negation of epistemic verb *show* at all (Appendix 15).

4.4.4.4. The Case of Epistemic ‘Seem’

The epistemic lexical verb *seem* was examined through the concordance lines in four corpora to discover the salient patterns and collocations that was used with the epistemic verb *seem*. It conveys epistemic meaning involving uncertainty based on sensory evidence, especially visual evidence. Concerning this, as a subject of epistemic verb *seem*, the subject *it* was discovered as the most preferred one in four corpora:

1. **It seems** to me she would have gained much and lost little. (LOCNESS)
2. **It seems** to me a great paradox! (NICLE)

Regarding the examples (1-2) from LOCNESS and NICLE, the subject *it* was used saliently rather than the other subjects in four corpora. LOCNESS (n=52) was the one that used the subject *it* the most; it was followed with NICLE (n=48), JPICLE (n=42), and TRICLE (n=20) sequentially.

According to the qualitative analysis, it was discovered the pattern *seem + to* was used saliently in four corpora. Thus, a verb comes after the pattern *seem + to* for indicating uncertainty based on visual evidence with epistemic *seem* and to infinitive:

1. **It seems to take** an important role in restricting the crime. (JPICLE)
2. The equality **seems to be** effective in the application of rules. (TRICLE)

As in the examples (1-2) from JPICLE and TRICLE, the epistemic verb *seem* was used with to infinitive. Four corpora used this pattern saliently; LOCNESS (n=142) was the one that used it the most. It was followed by NICLE (n=127), JPICLE (n=48), and TRICLE (n=36) in sequence.

For the other collocation of epistemic *seem*, it was discovered that the adjectives were used after the epistemic verb *seem* saliently in four corpora:

1. Rehabilitation of the criminal instead of punishment may **seem unfair** and insulting to the victim. (NICLE)
2. This approach **seems profitable**, but it is not completely adequate. (TRICLE)

As in this case (1-2) shown from NICLE and TRICLE, the epistemic verb *seem* was used with the adjectives. In LOCNESS (n=44), it was saliently used; however, NICLE (n=25), JPICLE (n=8), and TRICLE (n=12) used it less than the use of pattern *seem + to infinitive*. Thus, visual evidence conveyed with uncertainty was expressed via the adjectives.

The other collocations were found as the words *like* and *very* although they were not as salient as the other collocations and patterns. The preposition *like* were used after the epistemic verb *seem* in the meaning of *similar to*. For the adverb *very*, it was used to emphasize the adjectives:

1. Sudden riches overnight **seems like** a dream come true, but in reality could provide many problems. (LOCNESS)
2. Everything **seems very** convenient in the city. (JPICLE)

As in the examples (1-2) extracted from LOCNESS and JPICLE, the collocations of epistemic verb *seem* were used as the preposition *like* and the adverb *very* after the epistemic verb *seem*. For the preposition *like*, it was used saliently in NICLE (n=20) and LOCNESS (n=11); on the other hand, LOCNESS (n=8) and JPICLE (n=7) used the adverb *very* saliently.

Lastly, it was discovered the pattern *seem + that-clause* was used in four corpora. So, the epistemic verb *seem* were complemented with a clause:

1. It **seems that** social equality is virtually impossible to achieve. (NICLE)
2. It **seems that** there are laws which are running, In fact there aren't. (TRICLE)

The examples (1-2) reached from NICLE and TRICLE show the use of that-clause with the epistemic verb *seem*. It was used mostly with the subject *it* in four corpora. In this regard, LOCNESS (n=19) and JPICLE (n=19) were the ones that used it the most whereas NICLE (n=11) and TRICLE (n=5) come next in sequence (Appendix 16).

4.4.4.5. The Case of Epistemic 'Know/n'

When it comes to the case of epistemic lexical verb *know/n*, qualitative analysis was conducted to see the salient lexico-grammatical patterns and collocations used with the epistemic *know/n*. According to the qualitative analysis, the epistemic use of *know/n* was preferred mostly in a positive clause rather than a negative one. Moreover, the epistemic verb *know/n* was also used with that-clause saliently in four corpora:

1. We **know that** discrimination is occurring, and we know that it is both morally, and legally wrong, by the Americans With Disabilities Act of 1990. (LOCNESS)
2. It is **known that** nuclear energy has the lowest impact on the environment including air, water, land, habitat and animals, in all of energy resources. (TRICLE)
3. Now I **know** some are thinking we are losing a lot of quality time doing these things manually. (NICLE)
4. I **know** some students don't like English, especially speaking in English. (JPICLE)

The examples (1-2) from LOCNESS and TRICLE present the epistemic verb *know* was used with *that*-clause. Moreover, four corpora utilized epistemic verb *know* with *that*-clause both by using *that* and by not using *that*, as in the examples (3-4) extracted from NICLE and JPICLE. Thus, TRICLE (n=72) used the epistemic verb *know* with *that*-clause the most among four corpora; it was followed by LOCNESS (n=58), NICLE (n=53), and JPICLE (n=30) in sequence.

Another pattern used with the epistemic verb *know* was the pattern *as + pronoun + know(s)*. It was used saliently other than the pattern with *that*-clause in four corpora:

1. **As everybody knows** if there were nothing called money, there would be no order in the world. (TRICLE)
2. **As you know**, English is spoken almost the world. (JPICLE)

Regarding the examples (1-2) from TRICLE and JPICLE, the epistemic verb *know* was used with the pattern *as + pronoun* to indicate the proposition based on the knowledge of someone. This pattern was used by TRICLE (n=32) and JPICLE (n=15) the most while it was followed by NICLE (n=3) and LOCNESS (n=1).

Lastly, the collocation *be known as* was saliently used by LOCNESS and TRICLE whereas the other corpora did not prefer to use it that much:

1. This **is known as** the brutalization effect. (LOCNESS)
2. Physical strength **is known as** a male characteristic. (TRICLE)

The examples (1-2) from LOCNESS and TRICLE show the use of collocation *be known as*. It was mostly used in LOCNESS (n=17) and TRICLE (n=9). Moreover, the epistemic verb *know* was hardly used in a negative clause; instead, it was mostly used in a positive clause (Appendix 17).

4.4.4.6. The Case of Epistemic 'Claim'

The case of epistemic lexical verb *claim* was investigated with qualitative analysis through the concordance lines of four corpora to discover the salient patterns and collocations used with the epistemic verb *claim*. According to the findings, it was found that *that*-clause was used as a pattern with the epistemic verb *claim* saliently in four corpora:

1. On the other hand, some scientists **claim that** they use animals for the sake of science. (TRICLE)
2. People **claim that** computer games don't need a brain to be played. (LOCNESS)
3. People **claim** it has broken up homes. (NICLE)

As the examples (1-2-3) from TRICLE, NICLE, and LOCNESS suggest, the epistemic verb *claim* was used with *that*-clause. This pattern was used saliently in four corpora. It was used either with *that* or without *that* as in the third example from NICLE. In this regard, NICLE (n=70) was the one that used this pattern the most among four corpora whereas TRICLE (n=49), LOCNESS (n=43), and JPICLE (n=3) come next in sequence.

Moreover, another pattern was discovered although it was only used TRICLE and LOCNESS saliently. The pattern *it + passive + that-clause* was used in LOCNESS, TRICLE, and NICLE, but not in JPICLE. It may be because of the low frequency of the epistemic verb *claim* in JPICLE:

1. **It has been claimed** that it is so addictive that people will spend all their available cash on lottery tickets, only to be disappointed. (LOCNESS)
2. **It is claimed** that money is the best friend of your bad days. (TRICLE)

The examples (1-2) from LOCNESS and TRICLE present the use of the *pattern it + passive (claimed) + that-clause*. It was only used in three corpora; TRICLE (n=6) used this pattern the most. LOCNESS (n=5) and NICLE (n=2) follows it in sequence. Lastly, it was discovered that the epistemic verb *claim* was mostly used in a positive clause instead of a negative one; therefore, the negation of epistemic verb *claim* was not preferred by the four corpora (Appendix 18).

4.4.4.7. The Case of Epistemic 'Feel'

According to the qualitative analysis of the epistemic lexical verb *feel*, it was discovered that the epistemic verb *feel* was used mostly in a positive clause; thus, the negation of *feel* was hardly preferred in four corpora. For the salient patterns, the most used pattern was the pattern *feel + that-clause* in four corpora. Moreover, mostly all concordance lines were consisted of this pattern in four corpora:

1. For example, many people **feel that** all of the welfare recipients are black. (LOCNESS)
2. Once they have left school or college, many **feel that** they never want to open a book again. (TRICLE)
3. Many women **felt** they were both marginalized and invisible as both subjects and objects, especially in academic studies. (NICLE)
4. Now, I am one of the university students and I sometimes **feel** it is true. (JPICLE)

As in the examples (1-2) from LOCNESS and TRICLE, the epistemic verb *feel* was used with that clause by using the word *that*. On the other hand, there were other concordance lines that were used without *that*, as in the examples (3-4) from NICLE and JPICLE. When used with that-clauses, the epistemic *feel* was followed with a clause. In this regard, that-clauses were used both by using *that* and by not using *that* in four corpora; however, that-clause was preferred to be used mostly with *that*. LOCNESS (n=130) was the one that used the pattern *feel* + *that-clause* the most while it was followed by NICLE (n=66), JPICLE (n=14), and TRICLE (n=13) sequentially.

For the subjects that were used with the epistemic verb *feel*, the subject *I* was used saliently in LOCNESS, NICLE, and JPICLE although TRICLE used it less than the half of the frequency of the epistemic verb *feel*. The use of subject *I* with the epistemic *feel* can indicate the visibility of the writer and opinion of the writer more than the other subjects do:

1. **I feel** that although I take a practical degree the lack of practical topics are severe. (NICLE)
2. **I feel** it is important for Britain to join fully with its European partners to adapt to these changes in demography. (LOCNESS)

As in the examples (1-2) extracted from NICLE and LOCNESS, the epistemic verb *feel* was used with the subject *I*. It was used to convey the writer's opinion more directly by indicating the visibility of the writer. It was LOCNESS (n=69) that used the subject *I* with the epistemic verb *feel* the most. It was followed by NICLE (n=27), JPICLE (n=13), and TRICLE (n=2) in sequence (Appendix 19).

4.4.4.8. The Case of Epistemic 'Argue'

As for the case of epistemic lexical verb *argue*, the most salient lexico-grammatical patterns and collocations of epistemic verb *argue* were examined through the concordance lines of four corpora with a qualitative inquiry. First of all, the epistemic *argue* was preferred to be used mostly in a positive clause; thus, the negation of epistemic verb *argue* was hardly used in four corpora.

Moreover, as a salient pattern used with the epistemic verb *argue*, the pattern *argue* + *that-clause* was preferred mostly in four corpora:

1. Some people would **argue that** such children need to learn in their own language at first. (JPICLE)
2. Quite a few people would probably **argue that** this is true. (NICLE)
3. For instance, some **argue** it is a way to bring the family together. (LOCNESS)
4. First of all, they argue that they have not got any modern lab as a visual aids. (TRICLE)

Concerning the examples (1-2-3-4) from the four corpora, the epistemic verb *argue* was used with *that-clause*. It was used with or without the word *that* in four corpora; however, it was mostly used with the word *that*. LOCNESS (n=96) was the one that used it the most among four corpora. It was followed by NICLE (n=5), TRICLE (n=21), and JPICLE (n=9) sequentially.

Another pattern that was used with the epistemic verb *argue* was the pattern *it* + *passive* + *that-clause* in four corpora. In this case, the epistemic verb *argue* was used in its passive form followed by a *that-clause*:

1. **It can be argued that** an 8 hour written exam on 3000 pages is outrages. (NICLE)
2. **It has been argued that** when a father no longer occupies the role of breadwinner in the welfare family, he feels a loss of self-esteem that can lead to marital disruption. (LOCNESS)
3. Yet, **it is argued that** university degrees are theoretical and have little value. (TRICLE)

The examples (1-2-3) from NICLE, LOCNESS, and TRICLE show the epistemic verb *argue* was used with the pattern *it* + *passive* + *that-clause*. LOCNESS (n=24) was the one that used this pattern the most; it was followed by NICLE (n=6) and TRICLE (n=6). On the other hand, JPICLE did not use this pattern at all.

Furthermore, different modal auxiliaries were used with the epistemic verb *argue* in four corpora. The salient ones were detected as the modal auxiliary *would* in LOCNESS and NICLE and the modal auxiliary *can* in NICLE:

1. I **would argue** that many people take this calculated risk, people with low incomes all the way up to the incredibly wealthy. (NICLE)
2. Many **would argue** that the weaknesses and unfairness of this system are apparent and that it should be changed. (LOCNESS)
3. Nobody **can argue** that there is a great difference from reading about something and do it for real. (NICLE)

Regarding the examples (1-2) from NICLE and LOCNESS, the modal auxiliary *would* was used before the epistemic verb *argue*. LOCNESS (n=11) used this pattern the most while it was followed by NICLE (n=7), JPICLE (n=1), and TRICLE (n=1) in sequence. On the other hand, NICLE (n=8) used the pattern *can argue* the most saliently; LOCNESS (n=2) and TRICLE (n=1) come next sequentially. JPICLE did not use the pattern *can argue* at all.

Lastly, the epistemic verb *argue* was used with different subjects; especially with the subject *I*, *(many) people*, *(some) people*, and *one*:

1. **Many people argue** that only God has the right to take a life, not a judge or a prison warden. (LOCNESS)
2. **Some** might **argue** this is true democracy. (JPICLE)
3. **One** can also **argue** that stronger punishments will help decreasing criminality, but is this really so? (NICLE)
4. **I argue** that people must not be stoned to death before the people by rule. (TRICLE)

As in the examples (1-2-3-4) from the four corpora, different subjects were used with the epistemic verb *argue*. However, the objective subjects were mostly preferred rather than the subjective *I* and *we* in four corpora (Appendix 20).

4.4.4.9. The Case of Epistemic ‘Consider’

As for the qualitative analysis of the concordance lines of epistemic lexical verb *consider* in four corpora, the salient lexico-grammatical pattern and collocations of epistemic verb *consider*. According to the findings, the epistemic verb *consider* was preferred to be used mostly in a positive clause in four corpora; it means the negation of epistemic verb *consider* was not used that much in four corpora. In addition, the most used pattern in four corpora was discovered as the pattern *be + considered* in four corpora; the passive form of epistemic *consider* was used saliently in four corpora:

1. In such cases, a university degree **cannot be considered** as a criterion for a position. (TRICE)
2. I believe when a female **is considered** equal to a male in the physical and emotional sense, it distorts reality. (LOCNESS)
3. Now, however, milieu **is widely considered** to be responsible. (NICLE)
4. Sometimes cell phones **are considered** to be annoying. (JPICLE)

As the examples (1-2-3-4) from the four corpora suggest, the epistemic verb *consider* was used in its passive form. Regarding this, it was used as the most salient pattern among the other

patterns and collocations of epistemic *consider* in four corpora. LOCNESS (n=45) was the one that used this pattern the most; it was followed by TRICLE (n=35), NICLE (n=24), and JPICLE (n=14) in sequence.

The other pattern of epistemic verb *consider* that was discovered through the concordance lines of four corpora was found as the pattern *it + be + passive*. The four corpora used this pattern saliently:

1. For this reason, **It is considered** the most awe-inspiring and dangerous weapon in the world. (LOCNESS)
2. In other words **it is considered** a criminal act not to do as you are told by your superior officer for this period of time! (NICLE)
3. **It is** still **considered** to be the way to save or prolong the flesh and organs. (JPICLE)
4. Besides, **it's considered** that physically, men are more powerful than women. (TRICLE)

The examples (1-2-3-4) from the four corpora show the use of *it* with the passive form of epistemic *consider*. It was discovered that when the subject *it* was used with the passive form of epistemic *consider*, simple present tense was utilized in four corpora. Therefore, LOCNESS (n=10) used the pattern *it is considered* the most. TRICLE (n=8), NICLE (n=2), and JPICLE (n=2) come next sequentially.

Furthermore, after the epistemic verb *consider*, it was followed mostly by that-clause, the pattern *to be*, the pattern *as*, or an adjective in four corpora:

1. The more terrible experiences we gave and took, the more I **consider that** we should change the national flag. (JPICLE)
2. A criminal's life is generally **considered to be** of less value than the life of an innocent man. (LOCNESS)
3. First of all, men are **considered as** the superior class and good qualities are brought to them. (TRICLE)
4. This may be **considered discriminatory** against both parties... (NICLE)

On the other hand, four corpora utilized these patterns with different frequencies. For the pattern with that-clause, TRICLE (n=13), LOCNESS (n=11), and NICLE (n=8) used it the most saliently. The pattern *consider(ed) to be* was mostly used in LOCNESS (n=11) and TRICLE (n=9). Moreover, the pattern *consider(ed) as* was mostly preferred by TRICLE (n=25). Lastly, the pattern *consider(ed) + adjective* was used mostly by LOCNESS (n=15) and NICLE (n=11) (Appendix 21).

4.5. Discussion

The present study showed there were differences in the use of epistemic modal auxiliaries and lexical verbs in argumentative essays of those who have different L1 backgrounds involving Norwegian, Japanese, Turkish, and native speakers of English. The first research question was to investigate the frequency and diversity of epistemic modal auxiliaries used in four corpora. According to the results of frequency analyses, NICLE and LOCNESS used the epistemic modal auxiliary *would* the most and *will* at the second rank whereas TRICLE and JPICLE preferred to use the epistemic modal auxiliary *will* the most and *may* at the second rank. On the other hand, it was found the epistemic *must* and *should* was used the least in four corpora. In this regard, except the epistemic *must* and *should*, the other epistemic modal auxiliaries were used a great deal in four corpora. Nevertheless, JPICLE and TRICLE underrepresented the epistemic *would* and *could* while NICLE overrepresented the epistemic *might*. Moreover, LOCNESS was the one that used epistemic modal auxiliaries the most and followed by NICLE. However, TRICLE and JPICLE were the ones that used it the least in sequence. Regarding this, while the use of epistemic modal auxiliaries by LOCNESS and NICLE was found to be nearly the same, the use of it by TRICLE and JPICLE was nearly the same.

As for the answer of the second research question, it was concluded that while LOCNESS used the epistemic lexical verb *show* the most, NICLE, TRICLE, and JPICLE overrepresented the epistemic lexical verb *think* the most. NICLE, JPICLE, TRICLE underrepresented other lexical verbs while LOCNESS used them without huge gaps. JPICLE and TRICLE underrepresented epistemic *believe* and *seem* while NICLE and LOCNESS used them nearly the same. Therefore, non-native corpora used a limited range of epistemic lexical verbs while LOCNESS used a wide range.

When it comes to the epistemic strengths that these epistemic devices express, for epistemic modal auxiliaries, NICLE used nearly the same epistemic strengths, but mostly weak epistemic strength with a few more frequencies. JPICLE and TRICLE preferred to use strong epistemic strength the most and medium epistemic strength the least. On the other hand, LOCNESS used the medium strength the most and used others the same. However, a log-likelihood analysis was applied to see if there is a statistically significant difference between the four corpora concerning the expressions of epistemic strengths that epistemic modal auxiliaries carry. However, not all of them were concluded as statistically significant. For the ones that were concluded as statistically significant, it was found LOCNESS and NICLE used medium epistemic strength more than JPICLE and TRICLE and used weak epistemic strength more than TRICLE. Furthermore, NICLE preferred to use epistemic modal auxiliaries *may* and *might* the most to express weak epistemic strength. LOCNESS used epistemic *could* and *may* while TRICLE and JPICLE preferred to use epistemic *may* the most to express weak epistemic strength. For medium epistemic strength, four

corpora used epistemic *would* instead of epistemic *should*. Lastly, four corpora preferred to use epistemic *will* instead of epistemic *must* for expressing strong epistemic strength.

For the results of epistemic strengths expressed by epistemic lexical verbs in four corpora, it was revealed four corpora used medium epistemic strength the most. While LOCNESS, NICLE, and JPICLE used weak and strong epistemic strengths nearly the same, TRICLE underrepresented weak epistemic strength expressed by lexical verbs. On the other hand, it was found JPICLE overrepresented the medium epistemic strength by using it the most among the four corpora. To see if there is a statistically significant difference between them, a log-likelihood test was applied. According to the results, LOCNESS and NICLE used weak epistemic strength more than TRICLE. LOCNESS used strong epistemic strength more than JPICLE. Also, JPICLE used medium epistemic strength more than LOCNESS and TRICLE. As for the preferences of four corpora on the use of epistemic lexical verbs carrying varying degrees of epistemic strength, four corpora used the epistemic lexical verb *seem* the most to convey weak epistemic strength while JPICLE and TRICLE used with fewer frequencies than LOCNESS and NICLE. For medium epistemic strength, NICLE, JPICLE, and TRICLE overrepresented epistemic lexical verb *think*; however, LOCNESS used epistemic *believe* the most. Lastly, NICLE, JPICLE, and TRICLE preferred to use epistemic *know* the most for strong epistemic strength whereas LOCNESS used epistemic *show* the most.

As a result, quantitative analysis showed that although there were small differences between NICLE and LOCNESS, they showed a great deal of similarity in terms of the use of epistemic devices and the expression of epistemic strengths. On the other hand, TRICLE and JPICLE showed significant differences when compared to LOCNESS regarding the use of epistemic devices.

When it comes to the fourth question related to the qualitative analysis of epistemic devices used by four corpora, epistemic modal auxiliaries were used mostly with the objective subjective rather than the subjects *I* and *'we'* in four corpora. As a collocation of epistemic *may*, the conjunction *but* was utilized saliently in four corpora. The conjunctions *also* and *or* with the epistemic *may* were used in LOCNESS frequently; however, the other three corpora underrepresented these collocations of epistemic *may*. Moreover, if-clauses as a pattern used with the epistemic *may* was used by LOCNESS the most and used by NICLE and TRICLE saliently as well whereas JPICLE underrepresented this pattern. For the collocations and patterns used with the epistemic modal auxiliary *might*, it was discovered that the conjunction *but* was overrepresented by NICLE while the other three corpora used it fewer. Also, four corpora used if-clause with the epistemic *might* saliently. On the other hand, LOCNESS utilized the negation of epistemic *will*, the adverbs *also*, *never*, *eventually*, *soon*, *perhaps*, and *then* with the epistemic *will* saliently while the other three underrepresented these patterns. In addition, NICLE used the adverbs *probably* and *always* with the epistemic *will* the most among four corpora while the other three corpora used it fewer. As for the epistemic *would*, LOCNESS used if-clauses with the epistemic *would* frequently

while the others used it fewer. The conjunction *when* and the adverbs *probably*, *also*, and *then* with the epistemic *would* was utilized by LOCNESS and NICLE saliently whereas JPICLE and TRICLE underrepresented these collocations. The most salient collocation of epistemic *must* in four corpora was discovered as the verb *to be*; it was overrepresented by NICLE, JPICLE, and TRICLE whereas it was used less in LOCNESS. In the case of epistemic *should*, the verb *to be* was used as a collocation of epistemic *should* saliently in four corpora. Lastly, epistemic *could* was mostly used with the conjunction *but*, the verb *to be*, and the pattern *have + past participle* by LOCNESS and NICLE.

As for the qualitative analysis of grammatical structures of epistemic lexical verbs, the epistemic *think* was mostly used with that-clauses in four corpora. The conjunction *also* was utilized as a collocation of epistemic *think* in four corpora. On the other hand, JPICLE overrepresented the conjunction *but* with the epistemic *think*. The epistemic *believe* was used mostly with that-clause in four corpora; however, while LOCNESS overrepresented that-clauses, the other three corpora used it less. It was used with the conjunctions *but* and *also* in four corpora. When it comes to the epistemic *show*, LOCNESS overrepresented that-clause with the epistemic *show*; it was followed by NICLE. The conjunctions *but* and *also* were saliently used in LOCNESS whereas the other three corpora used it less. The epistemic *seem* was used with the subject 'it' saliently in four corpora. While LOCNESS and NICLE utilized the patterns *seem to + verb* and *seem + adjective* saliently, TRICLE and JPICLE used these patterns less. For the epistemic *know*, that-clause was used saliently in four corpora. Also, TRICLE used the pattern *as... know(s)* saliently while the other three corpora used it less. Similarly, that-clause was used with the epistemic *feel* in four corpora frequently. However, LOCNESS overrepresented this pattern and it was followed by NICLE while JPICLE and TRICLE underrepresented the pattern. Furthermore, the epistemic *claim* was used with that-clause in LOCNESS, NICLE, and TRICLE saliently while JPICLE underrepresented this pattern. It was the same in the use of epistemic *argue*. Moreover, LOCNESS utilized the pattern *it + passive + that-clause* with the epistemic *argue* saliently whereas the other did not. Lastly, the epistemic *consider* was used in its passive form saliently in four corpora.

All in all, it was found there is a similarity between LOCNESS and NICLE in terms of the use of epistemic modal auxiliaries. Differently, JPICLE and TRICLE underrepresented epistemic modal auxiliaries in total and epistemic *would*; instead, they preferred to use epistemic *will* more than the others. The results are in line with the previous studies (Yang & Zeng, 2021; Bouhlal et al., 2018; Vandenhoek, 2018; Karahan, 2017; Milton & Hyland, 1999; Hyland & Milton, 1997). Differently, Vold (2006) concluded in his study that while native speakers of English preferred to use epistemic modal auxiliaries *may* and *might* the most, native speakers of Norwegian used epistemic *can* and *may* the most. Similarly, McEnery and Kifle (2002) concluded from their study that non-native speakers of English preferred to use epistemic modal *may* the most while native

speakers of English used *would* the most. Furthermore, Yang (2018) found out in his study that both native and non-native speakers of English preferred to use modal auxiliary *can* the most. Hu and Li (2015) found in their study that while both native and non-native speakers of English used the modal auxiliary *will* the most. When it comes to epistemic strengths, the epistemic modal auxiliary *may* was preferred for the weak epistemic strength by the four corpora. For medium and strong epistemic strengths, four corpora preferred to use *would* and *will* instead of epistemic *should* and *must*. Overall, LOCNESS used the medium strength the most whereas JPICLE and TRICLE used the strong epistemic strength the most and the medium epistemic strength the least. Similarly, Karahan (2017) concluded that native speakers of English preferred to use the medium epistemic strength the most while those who have Turkish as L1 background used the medium strength the least while they preferred to use strong epistemic strength the most. NICLE used the epistemic strengths nearly the same; however, it used the weak epistemic strength the most.

For epistemic lexical verbs, four corpora used epistemic modal auxiliaries more than epistemic lexical verbs as in the study of Hyland and Milton (1997), Vandenhoeck (2018), and McEnery and Kifle (2002). On the other hand, JPICLE used it the most while TRICLE used it the least. LOCNESS preferred to use epistemic lexical verb *show* the most whereas the other three corpora overrepresented epistemic *think*. JPICLE and TRICLE underrepresented epistemic lexical verbs *believe* and *seem* while NICLE and LOCNESS used them nearly the same. These results were in line with the study of Karahan (2017) in terms of the overrepresentation of epistemic *think* by TRICLE and underrepresentation of epistemic *seem* by TRICLE. Similarly, in the study of Yang and Zeng (2021), it was found native speakers of English preferred to use lexical verb *seem* the most while non-native speakers of English used lexical verb *think* the most. Differently, in her study (2017), it was found that native speakers of English also used the epistemic *think* the most. Moreover, non-natives speakers' narrow range of epistemic lexical verbs were in line with the study of Hyland & Milton (1997) while non-native speakers of English used the epistemic lexical verb *think* the most. The same results were also obtained in their other study (Milton & Hyland, 1999). Also, Hu and Li (2015) concluded in their study that both native and non-native speakers of English preferred to use the lexical verb *think* the most. In the study of Vandenhoeck (2018), it was found non-native speakers of English used lexical verbs *suggest* and *think* the most whereas native speakers of English used lexical verbs *seem* and *expect* the most. Regarding the epistemic strengths, four corpora used the medium epistemic strength the most; however, JPICLE overrepresented the medium epistemic strength. LOCNESS used more strong epistemic strength than JPICLE and more weak epistemic strength than TRICLE. Differently, Karahan (2017) found in her study that while native speakers of English used the medium epistemic strength the most, those who are the native speakers of Turkish used strong epistemic strength the most.

Regarding these differences and similarities among the four corpora, it may be noticeable to discuss the possible reasons for these differences and similarities. These differences and similarities

may stem from the specific features of epistemic devices, language instruction, learner factors, language and nationality, pedagogical materials, intercultural pragmatic competence, and communicative competence.

For the inherent features of epistemic devices in English, learners of English in this study preferred to use modal auxiliary *will* and lexical verb *think* more than the others. In this regard, it seems that the use of lexical verb *think* with its epistemic meaning can be easier because the inherent features of lexical verb *think* suggest a way to express the belief of the writer by positioning himself/herself into his/her proposition by expressing his/her opinion without giving the proposition a full credence. The same rhetorical aspect is applicable to the use of modal auxiliary *will*. It may be suggested that the inherent features of modal auxiliary *will* suggest the use of modal auxiliary *will* for prediction more than for volition. In this regard, unlike epistemic *must* and *should* and other lexical verbs, epistemic *will* and *think* may seem easier to access and convey epistemic meaning for Turkish, Norwegian, and Japanese undergraduates thanks to the inherent features of these epistemic devices. Moreover, the results were in line with the other studies' suggestions; the inherent features of epistemic devices were discussed by Hyland and Milton (1997) in terms of their complexity both semantically and pragmatically. They (1997: 185) suggested that “modal expressions are complex for novice writers because they are polypragmatic, that is, they can simultaneously convey a range of meanings”. Hu and Li (2015) give the example *would* to signify the several meanings that it conveys. Moreover, the different meanings conveyed by modal expressions were discussed also by Coates (1983). Therefore, modal expressions do not carry only one meaning; thus, the context to which the modal expression belongs decides the meaning of the modal expression related to pragmatic interpretation (Hyland & Milton, 1997). On the other hand, while modal expressions can carry different meanings, epistemic meanings can be expressed through different modal verbs and lexical devices including epistemic verbs, which can make the use of epistemic devices more difficult for novice writers (Hyland & Milton, 1997). Not only semantic and pragmatic complexity, but also form complexity can be discussed related to the features of epistemic devices, which can make the use of epistemic devices more difficult. The form complexity is in relation with the ease of articulation and the lengths of the epistemic devices (Hu & Li, 2015). In this case, if the epistemic device is shorter than the other equivalent epistemic device, easier than the other one, and having easier articulation than the other one, learners of English can use it in their writings more than the other one (Hu & Li, 2015: 26). Overall, the inherent features of epistemic *will* and *think* may lead a way to use them more than the other epistemic devices by making epistemic *will* and *think* easier to use.

In addition to the complexity of epistemic devices, frequency of epistemic devices and the saliency of them are the other features that can have a role in the use of epistemic devices (Karahan, 2017; Hu & Li, 2015; Gass et al, 2013). These two are related to each other because “the frequent form is normally the one easily to get noticed, thus salient” (Hu & Li, 2015: 27). On the

other hand, every frequent epistemic device may not be as salient as the less frequent epistemic device. Therefore, the epistemic device that was used in the initial of the sentence may be more salient, but not that frequent (Hu & Li, 2015). In this regard, it can be the reason why four corpora used these epistemic devices and used the adverb *probably* with the epistemic *will* as a collocation.

Another possible reasons for the differences and similarities among four corpora can be the pedagogical instruction given to the learners of English. In this study, it was revealed that non-native learners of English overrepresented the epistemic lexical verb *think*. This overrepresentation may stem from the lack of knowledge on the argumentative essay and variety of lexical verbs that can carry epistemic meanings. The lack of instruction on academic and argumentative essay and epistemic modality markers can be a reason for this overrepresentation (Khoshshima et al., 2016). The learners of English may not notice the different variations of epistemic devices in the pedagogical instructions that had been given to them. Based on the Noticing Hypothesis proposed by Schmidt (1990), a linguistic item should be noticed by the learners to make the input intake. The pedagogical instructions given to the learners of English should enhance the notice and awareness of the learners regarding the various epistemic devices (Karahana, 2017; Hu & Li, 2015). Regarding this, the epistemic devices that had not been emphasized in the pedagogical instructions may not be noticed by the learners of English; therefore, it may be one of the reasons that the learners of English in this study may not use them saliently in their corpora and overrepresented only one or two epistemic devices.

Moreover, the instructors' supports to the learners of English can be one of the possible reasons of these differences and similarities. According to the sociocultural theory proposed by Vygotsky (1978), learning is a social process that can be enhanced with the parents' and instructors' support to the learner. In this case, scaffolding can promote the learning process of the learners (Vygotsky, 1978). Thus, it may be possible that if the support of the instructor and scaffolding was not given to the learners enough, their learning of epistemic devices may be finite (Hu & Li, 2015). As this study shows, there were overrepresentation and underrepresentation of several epistemic devices with a finite range of use of epistemic devices by non-native speakers of English.

In addition to the instructors, the pedagogical materials that are used in teaching of English may affect the use of epistemic devices and can be one of the explanations of the differences and similarities resulted in this study. Since the essays written by undergraduates have different nationalities, it is possible to argue on the use of pedagogical materials for English teaching throughout their school period. Thus, pedagogical materials for English teaching can be one of the determinants of the overrepresentations and underrepresentations in the argumentative essays of Turkish, Norwegian, and Japanese undergraduates. As in line with the study of Bouhlal et al. (2018), it was found that textbooks overrepresented several modal verbs. Also, it was revealed that

the textbooks that was used in teaching of English used these modal verbs more than the native speaker corpus (Bouhlal et al., 2018; Römer, 2004; McEnery & Kifle, 2002). Therefore, some of the epistemic devices can be ignored by the textbooks and coursebooks of learners of English (Hyland & Milton, 1997). This overrepresentation and underrepresentation of epistemic devices in textbooks can affect the notice of the learners regarding other epistemic devices resulting the overrepresentation of only the epistemic devices that are salient in their textbooks and coursebook while they underrepresent the others.

As well as the external factors, there may be internal factors about the learners that might be reasons for these differences and similarities in this study regarding the overrepresentation of epistemic *think* and *will* and the expression of more strong epistemic strength. One of the most known factors may be the L1 transfer. The modal system that the learners acquired in their L1 may be transferred into the knowledge on L2 (Karahana, 2017; Hu & Li, 2015; Gass et al., 2013). As in the behaviorist theory, there might be L1 interference that resulted in the transfer of L1 habits to the knowledge on L2 by the learners of English (Gass et al., 2013). Thus, the learners of English in this study may transfer their L1 habits into their argumentative essays with overrepresentation or underrepresentation of several epistemic devices. In addition to L1 transfer, L1=L2 hypothesis can explain these results. L1=L2 hypothesis can be defined as the process of learning L2 is similar to the process of acquiring L1 (Ellis, 1994). Thus, those who have different L1 backgrounds can show differences in the use of epistemic devices.

Concerning the overrepresentation of epistemic *will* and *think*, Hu and Li (2015) maintained when a modal auxiliary or a lexical verb is chosen to convey the epistemic meaning, the other modal auxiliaries and lexical verbs can seem unnecessary by suggesting that One-to-One Principle can be another factor for the use of epistemic devices by the learners of English. Although it can be diminished as the proficiency of the learner improves, it can still be a factor that can affect the use epistemic devices explaining the overrepresentation of some epistemic devices in this study.

When it comes to the L2 proficiency of the learners, it can be one of the reasons of why non-native speakers of English stuck to the epistemic *will* and *think* and ignored the other epistemic modal auxiliaries and lexical verbs. Since more proficient learners of English can be more interested in the stylistic aspect of the writing and register knowledge, they may use epistemic modal auxiliary *must* less frequently as in this study (Hu & Li, 2015). Moreover, less proficient learners of English may not notice the epistemic devices that are not as salient as the other while the more proficient ones can notice and be aware of them (Hu & Li, 2015). Moreover, the learners' less knowledge on the register and argumentative essay may result in the limited range of use of epistemic devices in their argumentative essays (Khoshshima et al., 2016). Thus, it can explain the underrepresentation of modal auxiliaries *would*, *could*, *might*, *must*, and *should* along with the lexical verbs except *think*.

As for the other reasons of the overrepresentation of epistemic will and think, underrepresentation of the other epistemic devices, and the expression of either weak or strong epistemic strength, one of them can be related to the language family and nationality that the learners of English have. The language family and nationality can affect the use of epistemic devices indicating various epistemic strengths. As in line with this study, Vold (2006) found out that English and Norwegian articles are the same and explained it with the similarity between the language family that they both have, which is Germanic. He (2006) adds that English and Norwegian speakers may tend to express themselves clearer. Related to cautiousness, positive face, and humbleness, expressing uncertainty can occur more in the cultures having more humbleness (Vold, 2006). In this regard, the L2 writers from different academic cultural backgrounds may choose the epistemic devices differently with different epistemic strengths (Yang & Zeng, 2021; Yang, 2018; Chen, 2010; Vold, 2006; Milton & Hyland, 1999). Therefore, different cultural conventions and educational backgrounds may affect the choose of epistemic devices (Khoshshima et al., 2016). In this regard, those who have different language family and nationality may use the epistemic devices differently and express their L1 academic cultural backgrounds in their argumentative essays by expressing different degrees of commitments. It may explain the reason why Japanese learners of English preferred to use medium epistemic strength the most while Turkish learners of English used the medium strength the least and used strong epistemic strength the most in this study.

Hereby, having a cultural knowledge of the target language regarding pragmatic competence can be requirement for the learners of English who have different L1 backgrounds. It can be called intercultural pragmatic competence, which mainly means the cultural knowledge of the target language's situation (Chen, 2010). When target discourse community is understood by the learners, it may be possible to use linguistic items that express different degrees of directness appropriately based on the intercultural pragmatic competence the learners have built (Chen, 2010). The pragmatic nuances that are not understood may be avoided by the learners (Khoshshima et al., 2016). The intercultural pragmatic competence can be a determinant on the choice on rhetorical strategies enabling the learners of English to use linguistics items that preserve positive face and avoid face-threatening acts (Chen, 2010; Hyland & Milton, 1997). In this regard, having intercultural pragmatic competence can be the other reason why non-native speakers of English used the epistemic devices expressing medium epistemic strengths the most in lexical verbs. As for the TRICLE and JPICLE using the strong epistemic strength the most in modal auxiliaries, they may not be as much as competent in intercultural pragmatics and may not understand the target community's discourse in expressing uncertainty by using modal auxiliaries.

Lastly, communicative competence can be another explanation of different uses of epistemic devices by non-native speakers of English. Communicative competence can be divided into two:

academic and interpersonal competence. Academic competence is more related to writing skills (Yang & Zeng, 2021). Through writing, there will be communication between the writer and the reader. This communication can be preserved via face-saving epistemic markers; thus, this pragmatic aspect of writing can enable the efficient use of communicative competence in argumentative essays of L2 learners by engaging with the reader communicatively in their argumentative essays (Chen, 2010). Thus, the use of epistemic markers can depend on the communicative competence of the learners of English.

All these factors can be determinants in different uses of epistemic devices with different epistemic strengths in four corpora. Therefore, each of these factors are complex issues in nature and should be examined separately to find out the relationship between them. However, there is still reliable evidence about them being possible reasons for the differences and similarities between native speakers of English and non-native speakers of English.

CONCLUSION AND RECOMMENDATIONS

This study aims at investigating epistemic modal auxiliaries and epistemic lexical verbs used in the argumentative essays of Norwegian, Japanese, Turkish, and English undergraduate students along with the epistemic strengths they express. To this end, the frequencies of epistemic devices and epistemic strengths of epistemic devices were analyzed. Moreover, a log-likelihood test was applied to the epistemic strengths of four corpora to see if there is a statistically significant difference between them. Apart from the quantitative analysis, qualitative analysis was conducted to investigate the most used epistemic lexical verbs, all epistemic modal auxiliaries which were included in this study, and their grammatical structures for discovering the lexico-grammatical patterns and collocations used with them.

In conclusion, LOCNESS and NICLE showed similarities in the use of epistemic modal auxiliaries and epistemic strength expressed with epistemic modal auxiliaries and lexical verbs. They both used the epistemic *would* the most while TRICLE and JPICLE used the epistemic *will* the most among the modal auxiliaries. Moreover, while TRICLE and JPICLE expressed strong epistemic strength the most and medium epistemic strength the least, NICLE used weak epistemic strength more with small differences and LOCNESS preferred medium epistemic strength the most. Therefore, it was concluded that TRICLE and JPICLE showed similarities in the use of modal auxiliaries, epistemic strength expressed with epistemic modal auxiliaries. For the epistemic lexical verb that was used the most, NICLE, TRICLE, and JPICLE stuck to epistemic *think* while they hardly used the other epistemic lexical verbs. For LOCNESS, the epistemic *show* was used the most; however, the other epistemic lexical verbs were also used with high frequencies. In this regard, NICLE, TRICLE, and JPICLE used a limited range of epistemic lexical verbs. Four corpora used medium epistemic strength expressed with epistemic lexical verbs the most. The reason for the use of medium epistemic strength via lexical verbs may come from the desire to avoid direct and certain propositions since lexical verbs are used as main verbs in the clauses. For the patterns and collocation of epistemic modal auxiliaries and lexical verbs used in four corpora, four corpora showed similarities in common patterns that helps to hedge the propositions. Lastly, subjects *I* and *we* were hardly used by four corpora; they preferred to use objective subjects with epistemic devices.

Overall, the results from this study seem to provide an insight into the use of epistemic modal auxiliaries and lexical verbs in Turkish, Japanese, and Norwegian non-native argumentative essays and the argumentative essays of native speakers of English. The preliminary results indicate that

while native speakers of English and Norwegian learners of English utilized the epistemic modal auxiliaries and lexical verbs nearly the same in terms of frequencies, epistemic strengths, and grammatical structures despite small differences, TRICLE and JPICLE used these with noticeable differences. The results may suggest that Norwegian learners of English try to involve in the target language's discourse community by utilizing the epistemic devices in their argumentative essays for both expressing uncertainty and making an argumentation like the native speakers of English. On the other hand, TRICLE and JPICLE expressed their propositions by using epistemic modal auxiliaries indicating strong epistemic strength, which means they are making their argumentations in a more certain way than native speakers of English instead of keeping the positive face by adding uncertainty in their propositions. In this regard, Norwegian speakers of English seem to position themselves into their proposition more efficiently because they tend to express uncertainty in their argumentative essays; thus, it can enable efficient communication between the reader and the writer giving the reader more credence. On the other hand, since JPICLE and TRICLE was concluded as using more certain propositions, it may distort the credence of the writer in terms of argumentative essay. However, although NICLE have several similarities with the native speakers of English, NICLE, TRICLE, and JPICLE did not use wide range of epistemic devices; instead, they used a limited range of epistemic lexical verbs.

Limitations of the Study

Although this study sheds on light the use of epistemic modal auxiliaries and lexical verbs in the argumentative essays of the Turkish, Norwegian, and Japanese learners of English and native speakers of English, there is still a room for further studies. Regarding this, future studies may investigate the use of other epistemic devices, such as epidemic adverbials, adjectives, and nouns. Moreover, this study investigated the argumentative essays of only Turkish, Norwegian, and Japanese learners of English; therefore, future investigations can involve the argumentative essays of those who have different L1 backgrounds. Not only those who have different L1 backgrounds, but also those who have different L2 proficiency levels can be included in future studies. Gender differences in the use of these epistemic devices may be examined in future studies. A focus group interview can be conducted with the participants to see their intentions behind using these devices. In addition, different writing genres and registers may be examined in terms of the use of epistemic devices. However, the reason why this study investigated argumentative essays is argumentative essays can be one of the most suitable writing genres for making argumentation by using epistemic devices. Moreover, the use of epistemic devices may be examined within the framework of the possible determinants mentioned in the discussion section, such as intercultural pragmatic competence and cultural conventions. In this regard, future studies may also investigate these factors. Lastly, epistemic devices can be investigated in terms of the topics of argumentative essays because the use of epistemic devices can change according to the topics that were written on.

Implications

Based on the results, there is a need that teachers of English should be given several implications to involve learnings of English into the target language's discourse community in terms of argumentative essay. First of all, teachers of English should mention about the nature of epistemic devices including the complexity that they convey in nature to make the English learners aware of the complexity of them. Moreover, the required scaffolding should be given to the English learners to support them for involving them in the efficient use of language in different genres, especially in argumentative essays. Regarding frequency and saliency that can affect the use of epistemic devices by the learners of English, coursebook can be better to benefit from corpus linguistics in order to reflect the authentic usage of target language. Therefore, the coursebooks that involve only a limited range of epistemic devices can add the elements of target language's discourse community into their contents. By doing so, the learners of English may notice a wide range of epistemic devices and authentic usages of them. Moreover, the coursebooks for academic writing can focus on the genre writing, especially argumentative essays. In this way, it can enable the learners of English genre knowledge on how they can use the linguistic items proper to the specific writing genre. These implications can allow the learners to gain intercultural pragmatic competence by making them aware of the authentic use of epistemic devices to express uncertainty and make argumentations in argumentative essays. Overall, both teachers and learners of English and coursebooks writers can benefit from the corpus linguistics to make English learners aware of the language and its authentic use in the target language's discourse community.

REFERENCES

- Aikhenvald, Alexandra Y. (2004), **Evidentiality**, Oxford University Press.
- Anthony, Laurence (2013), "A Critical Look at Software Tools in Corpus Linguistics", **Linguistic Research**, 30(2), 141-161.
- Auwers, Johan Van Der and Plungian, Vladimir A. (1998), "Modality's Semantic Map", **Linguistic Typology**, 2(1), 79-124.
- Barton, Ellen L. (1993), "Evidentials, Argumentation, and Epistemological Stance", **College English**, 55(7), 745-769.
- Bıkmaz Bilgen, Özge and Doğan, Nuri (2017), "The Comparison of Interrater Reliability Estimating Techniques", **Journal of Measurement and Evaluation in Education and Psychology**, 8(1), 63-78. doi: 10.21031/epod.294847
- Biber, Douglas (1993), "Representativeness in Corpus Design", **Literary and Linguistic Computing**, 8(4), 243-257.
- _____ (2006), **University Language: A Corpus-Based Study of Spoken and Written Registers**, John Benjamins Publishing Company.
- Biber, Douglas et al. (1998), **Corpus Linguistics: Investigating Language Structure and Use**, Cambridge University Press.
- Biber, Douglas et al. (1999), **Longman Grammar of Spoken and Written English**, Longman.
- Bouhlal, Fatma et al. (2018), "Modality in ESL Textbooks: Insights from a Contrastive Corpus-Based Analysis", **Canadian Modern Language Review**, 74(2), 227-252.
- Boye, Kasper (2012), **Epistemic Meaning: A Crosslinguistic and Functional-Cognitive Study**, De Gruyter Mouton.
- Breyer, Yvonne Alexandra (2011), **Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges**. *English Corpus Linguistics*, Peter Lang.
- Btoosh, Mousa A. (2019), "Modals in Arab EFL Learners' Composition: A Corpus-Based Approach", **Linguistics and Literature Studies**, 7(3), 100-109. doi: 10.13189/lis.2019.070302
- Chen, Hsin-I (2010), "Contrastive Learner Corpus Analysis of Epistemic Modality and Interlanguage

- Pragmatic Competence in L2 Writing”, **Journal of Second Language Acquisition and Teaching**, 17, 27-51.
- Cheng, Winnie and Lam, Phoenix (2021), “What Can a Corpus Tell Us About Language Teaching?”, Anne O’Keeffe and Michael J. McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (299-312), Routledge, London.
- Clancy, Brian (2021), “Building a Corpus to Represent a Variety of a Language”, Anne O’Keeffe and Michael J. McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (62-74), Routledge, London.
- Coates, Jennifer (2015), **The Semantics of the Modal Auxiliaries**, Routledge.
- Cohen, Jacob (1960), “A Coefficient of Agreement for Nominal Scales”, **Educational and Psychological Measurement**, 20(1), 37-46.
- Conrad, Susan (2000), “Will Corpus Linguistics Revolutionize Grammar Teaching in the 21st Century?”, **Tesol Quarterly**, 34(3), 548-560.
- Corder, Stephen Pit (1967), “The Significance of Learners’ Errors”, **International Review of Applied Linguistics**, 5, 161-170.
- Cornillie, Bert (2009), “Evidentiality and Epistemic Modality: On the Close Relationship Between Two Different Categories”, **Functions of Language**, 16(1), 44-62. doi 10.1075/fol.16.1.04cor
- Crompton, Peter (1997), “Hedging in Academic Writing: Some Theoretical Problems”, **English for Specific Purposes**, 16(4), 271-287.
- Davies, Alan (2003), **The Native Speaker: Myth and Reality**, Multilingual Matters.
- Desagulier, Guillaume (2017), **Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics**, Springer.
- Ellis, Rod (1994), **The Study of Second Language Acquisition**, Oxford University Press.
- Ericsson, Tina (2008), “It Is Certain That It Can Be Argued A Million Times Over” – Expressions of Epistemic Modality in L1 and L2 Writing”, http://www.diva-portal.org/smash/resultList.jsf?dswid=-1754&language=en&searchType=SIMPLE&query=It+Is+Certain+That+It+Can+Be+Argued+A+Million+Times+&af=%5B%5D&aq=%5B%5B%5D%5D&aq2=%5B%5B%5D%5D&aqe=%5B%5D&noOfRows=50&sortOrder=author_sort_asc&sortOrder2=title_sort_asc&onlyFullText=false&sf=all (15.05.2023).
- Facchinetti, Roberta et al. (2003), **Modality in Contemporary English**, Mouton de Gruyter.
- Flowerdew, John (1996), “Concordancing in Language Learning”, Martha C. Pennington (Ed.), **The Power of CALL**, In (97-113), Athelstan, Houston, TX.

- Flowerdew, Lynne (2011), **Corpora and Language Education**, Springer.
- _____ (2012), **Corpora and Language Education**, Palgrave Macmillan.
- Gass, Susan M. and Selinker, Larry (2008), “Reexamining the Role of Explicit Information in Processing Instruction”, **Studies in Second Language Acquisition**, 30, 277-305.
- Gass, Susan M. et al. (2013), **Second Language Acquisition: An Introductory Course**, 4th Ed., Routledge, London.
- Granger, Sylviane (1998a), “Prefabricated Patterns in Advanced EFL Writing: Collocations and Lexical Phrases”, Anthony Paul Cowie (Ed.), **Phraseology: Theory, Analysis and Applications**, In (145-160), Clarendon Press, Oxford.
- _____ (1998b), “The Computer Learner Corpus: A Versatile New Source of Data for SLA Research”, Sylviane Granger (Ed.), **Learner English on Computer**, In (3-18), Addison Wesley Longman.
- _____ (2002), “A Bird’s-Eye View of Learner Corpus Research”, Sylviane Granger, Joseph Hung, and Stephanie Petch-Tyson (Eds.), **Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching**, In (3-33), John Benjamins Publishing.
- _____ (2015), “Contrastive Interlanguage Analysis: A Reappraisal”, **International Journal of Learner Corpus Research**, 1(1), 7-24. doi: 10.1075/ijlcr.1.1.01gra
- Granger, Sylviane et al. (2020), **International Corpus of Learner English**, 3rd Volume, Presses Universitaires de Louvain.
- Greenbaum, Sidney (1996), **The Oxford English Grammar**, Oxford University Press.
- Guziurová, Tereza (2016), “The Integrative Approach to Metadiscourse Revisited: Hedging in Academic Writing”, Tomášková, Renáta et al. (Eds.), **Professional Genres from an Interpersenoal Perspective**, In (112-139), Cambridge Scholars Publishing.
- Halliday, M.A.K. (1994), **An Introduction to Functional Grammar**, 2nd Ed., Edward Arnold.
- Halliday, M.A.K. et al. (2004), **Introduction to Functional Grammar**, 3rd Ed., Hodder Arnold.
- Hu, Chunyu and Li, Xuyan (2015), “Epistemic Modality in the Argumentative Essays of Chinese EFL Learners”, **English Language Teaching**, 8(6), 20-31.
<http://dx.doi.org/10.5539/elt.v8n6p20>
- Huddleston, Rodney and Pullum, Geoffret K. (2017), **The Cambridge Grammar of the English Language**, Cambridge University Press.
- Hunston, Susan (2022), **Corpora in Applied Linguistics**, 2nd Ed., Cambridge University Press.
- Hyland, Ken (1998), **Hedging in Scientific Research Articles**, John Benjamins Publishing.

- _____(2005a), **Metadiscourse: Exploring Interaction in Writing**, Continuum.
- _____(2005b), "Stance and Engagement: A Model of Interaction in Academic Discourse", **Discourse studies**, 7(2), 173-192.
- Hyland, Ken and Milton, John (1997), "Qualification and Certainty in L1 and L2 Students Writing", **Journal of Second Language Writing**, 6(2), 183-205.
- Johns, Tim (1991), "Should You Be Persuaded - Two Examples of Data-Driven Learning Materials", Tim Johns, & Phillip King (Eds.), **Classroom Concordancing: English Language Research Journal**, In (1-16), Birmingham, UK: Birmingham University.
- Karahan, Pınar (2017), **Expression of Epistemic Modality in Argumentative Writing: A Study Based on Learner Corpora**, Unpublished Doctoral Dissertation, Anadolu University – Department of English Language Teaching.
- _____(2022), "Comparative Analysis of Epistemic Modal Verbs in the Three Sub-Corpora of Humanities & Social Sciences", **International Journal of Educational Spectrum**, 4(1), 43-63. <https://doi.org/10.47806/ijesacademic.1070806>
- Khamkhien, Attapol (2021), "Adjectives in the Expression of Epistemic Modality in Applied Linguistics Research Articles", **Language Literacy: Journal of Linguistics, Literature, and Language Teaching**, 5(2), 431-452. doi: 10.30743/ll.v5i2.4534
- Khoshsima, Hooshang et al. (2016), "A Cross-Linguistic and Cross-Cultural Study of Epistemic Modality Markers in Linguistics Research Articles", **Iranian Journal of English for Academic Purposes**, 5(2), 102-122.
- Kibler, Amanda K. and Hardigree, Christine (2017), "Using Evidence in L2 Argumentative Writing: A Longitudinal Case Study across High School and University", **Language Learning**, 67(1), 75-109. doi: 10.1111/lang.12198
- Kilgarrieff, Adam (2005), "Language is Never, Ever, Ever, Random", **Corpus Linguistics and Linguistic Theory**, 1(2), 263-276.
- Kilgarrieff, Adam et al. (2014), "The Sketch Engine: Ten Years on", **Lexicography**, 1(1), 7-36.
- Larsen-Freeman, Diane and Long, Michael Hugh (2014), **An Introduction to Second Language Acquisition Research**, Routledge, New York.
- Lennon, Paul (2008), "Contrastive Analysis, Error Analysis, Interlanguage", Stephan Gramley and Vivian Gramley (Eds.), **Bielefeld Introduction to Applied Linguistics**, In (51-60), Aisthesis, Bielefeld.
- McEnery, Tony and Kifle, Nazareth Amselom (2002), "Epistemic Modality in Argumentative Essays of Second-Language Writers", John Flowerdew (Ed.), **Academic discourse**, In (182-215), Longman.

- Meyer, Charles F. (2002), **English Corpus Linguistics: An Introduction**, Cambridge University Press, Cambridge.
- Milton, John and Hyland, Ken (1999), "Assertions in Students' Academic Essays: A Comparison of L1 and L2 Writers", Roger Stephen Berry et al., (Eds.), **Language Analysis, Description and Pedagogy**, In (147-161), HKUST.
- Ngula, Richmond Sadick (2015), **Epistemic Modality in Social Science Research Articles Written by Ghanaian Authors: A Corpus-Based Study of Disciplinary and Native vs. Non-Native Variations**, Unpublished Doctoral Dissertation, Lancaster University – Department of Linguistics and English Language.
- Nuyts, Jan (2001), **Epistemic Modality, Language and Conceptualization: A Cognitive-Pragmatic Perspective**, John Benjamins.
- _____(2017), "Evidentiality Reconsidered", Juana Isabel Marín Arrese et al. (Eds.), **Evidentiality Revisited: Cognitive Grammar, Functional and Discourse-Pragmatic Perspectives**, In (57-83), John Benjamins Publishing Company.
- Oh, Sun-Young and Kang, Suk-Jin (2013), "The Effect of English Proficiency on Korean Undergraduates' Expression of Epistemic Modality in English Argumentative Writing", **The Journal of Asia TEFL**, 10(4), 97-132.
- O'Keeffe, Anne et al. (2007), **From Corpus to Classroom: Language Use and Language Teaching**, Cambridge University Press.
- Palmer, Frank (1986), **Mood and Modality**, Cambridge University Press
- _____(2001), **Mood and Modality**, Cambridge University Press.
- _____(2003), "Modality in English: Theoretical, Descriptive and Typological Issues", Facchinetti, Roberta et al. (Eds.), **Modality in Contemporary English**, In (1-17), Mouton de Gruyter, Berlin, New York.
- Panocová, Renatá and Lukačín, Lukáš (2019), "Epistemic Modal Markers in Two Domains Of Academic Research Papers in English", **Brno Studies in English**, 45(2), 121-138. <https://doi.org/10.5817/BSE2019-2-6>
- Pérez-Llantada, Carmen (2010a), "The 'Dialectics Of Change' As a Facet of Globalisation: Epistemic Modality in Academic Writing", Miguel F. Ruiz-Garrido et al. (Eds.), **English for Professional and Academic Purposes**, In (25-41), Rodopi.
- _____(2010b), "The Discourse Functions of Metadiscourse in Published Academic Writing: Issues of Culture and Language", **Nordic Journal of English Studies**, 9(2), 41-68.

- Pérez-Paredes, Pascual and Mark, Geraldine (2021), "What Can Corpora Tell Us about Language Learning?", Anne O'Keeffe and Michael J. McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (313-327), Routledge, London.
- Quirk, Randolph et al. (1985), **A Comprehensive Grammar of the English Language**, UK: Longman Group Ltd.
- Reppen, Randi (2021), "Building A Corpus: What are Key Considerations?", Anne O'Keeffe and Michael J. McCarthy (Eds.), **The Routledge Handbook of Corpus Linguistics**, In (13-20), Routledge, London.
- Roberts, Carl W. et al. (2010), "Modality Analysis: A Semantic Grammar for Imputations of Intentionality in Texts", **Quality & Quantity**, 44, 239-257.
- Rozumko, Agata (2017), "Adverbial Markers of Epistemic Modality across Disciplinary Discourses: A Contrastive Study of Research Articles in six Academic Disciplines", **Studia Anglica Posnaniensia**, 52(1), 73–101. doi: 10.1515/stap-2017-0004
- Römer, Ute (2004), "Textbooks: A Corpus-Driven Approach to Modal Auxiliaries and Their Didactics", John M. Sinclair (Ed.), **How to Use Corpora in Language Teaching**, In (185–199). John Benjamins. [https://doi.org/ 10.1075/scl.12.14rom](https://doi.org/10.1075/scl.12.14rom)
- Rubin, Victoria L. (2010), "Epistemic Modality: From Uncertainty to Certainty in the Context of Information Seeking as Interactions with Texts", **Information Processing and Management**, 46, 533-540.
- Rustipa, Katharina (2011), "Contrastive Analysis, Error Analysis, Interlanguage and the Implication to Language Teaching", **Ragam Jurnal Pengembangan Humaniora**, 11(1), 16-22.
- Salager-Meyer, Françoise (1997), "I think That Perhaps You Should: A Study of Hedges in Written Scientific Discourse", **Functional approaches to written text: Classroom applications**, 1, 127-143.
- Schachter, Jacqueline (1974), "An Error in Error Analysis", **Language Learning**, 24, 205-214. doi: 10.1111/j.1467-1770.1974.tb00502.x
- Schmidt, Richard W. (1990), "The Role of Consciousness in Second Language Learning. Applied Linguistics", **Applied Linguistics**, 11, 129-158.
- Selinker, Larry (1972), "Interlanguage", **International Review of Applied Linguistics**, 10, 209-231.
- Sinclair, John (1991), **Corpus, Concordance, Collocation**, Oxford University Press.
- Tognini-Bonelli, Elena (2001), **Corpus Linguistics at Work**, John Benjamins Publishing.

- Vandenhoeck, Tim (2018), "Epistemic Markers in NS and NNS Academic Writing", **Journal of Academic Writing**, 8(1), 72-91. <http://dx.doi.org/10.18552/joaw.v8i1.359>
- Vass, Holly (2017), "Lexical Verb Hedging in Legal Discourse: The Case of Law Journal Articles and Supreme Court Majority and Dissenting Opinions", **English for Specific Purposes**, 48, 17-31. <https://doi.org/10.1016/j.esp.2017.07.001>
- Vold, Eva Thue (2006), "Epistemic Modality Markers in research Articles: A Cross-Linguistic and Cross-Disciplinary Study", **International Journal of Applied Linguistics**, 16(1), 61-87. doi: 10.1111/j.1473-4192.2006.00106.x
- Vygotsky, Lev S. (1978), **Mind in Society: The Development of Higher Psychological Processes**, Cambridge, MA: Harvard University Press.
- Yang, Xiaowan (2018), "A Corpus-Based Study of Modal Verbs in Chinese Learners' Academic Writing", **English Language Teaching**, 11(2), 122-130. doi: 10.5539/elt.v11n2p122
- Yang, Demei and Zeng, Xiangmin (2021), "Epistemic Modality and Projection in Argumentative Writing: A Corpus-driven Comparison of Chinese English Majors and Native English Speakers", **Theory and Practice in Language Studies**, 11(4), 412-421. DOI: <http://dx.doi.org/10.17507/tpls.1104.11>
- Wentzel, Arnold (2018), **A Guide to Argumentative Research Writing and Thinking: Overcoming Challenges**, Routledge.
- Willett, Thomas (1988), "A Cross-Linguistic Survey of the Grammaticization of Evidentiality", **Studies in Language**, 12(1), 51-97.
- Zhang, Hong-yan (2015), "Comparative Analysis of Modal Auxiliary Verbs in English and in Chinese", **Sino-US English Teaching**, 12(2), 128-136. doi:10.17265/1539-8072/2015.02.007



APPENDICES

Appendix 1: Moderate and High Inter-rater Reliability for ELVs in LOCNESS

Moderate and High Inter-rater Reliability for ELVs in LOCNESS

LOCNESS	Cohen's Kappa (κ) value	p
Appear	,816	<,001
Assume	,481	<,001
Assure	1	,025
Believe	,802	<,001
Claim	,887	<,001
Conclude	1	<,001
Confirm	1	,046
Consider	,873	<,001
Convince	1	<,001
Demonstrate	,870	<,001
Describe	,811	<,001
Doubt	,718	<,001
Emphasize	1	<,001
Establish	,869	<,001
Estimate	1	,025
Expect	,676	<,001
Feel	,890	<,001
Guess	1	<,001
Hope	1	<,001
Identify	,879	<,001
Indicate	1	<,001
Infer	1	<,001
Interpret	,780	<,001
Know	,722	<,001
Look (like)	,895	<,001
Look as if	1	<,001
Presume	1	<,001
Propose	,692	,001
Prove	,513	<,001
Realize	,685	<,001
Reckon	1	<,001
Recognize	1	<,001
Regard	1	<,001
Reveal	,921	<,001
Seem	1	<,001
Show	1	<,001
Speculate	1	<,001
Stress	1	<,001
Suggest	,791	<,001
Suppose	,921	<,001
Suspect	,696	,053
Tend	,658	<,001

Appendix 2: Moderate and High Inter-rater Reliability for ELVs in NICLE

Moderate and High Inter-rater Reliability for ELVs in NICLE

NICLE	Cohen's Kappa (κ) value	p
Appear	,823	<,001
Argue	,580	<,001
Assume	1	,003
Assure	1	<,001
Believe	,878	<,001
Claim	,901	<,001
Conclude	,789	,003
Confirm	1	,025
Consider	,941	<,001
Convince	1	<,001
Demonstrate	1	<,001
Describe	1	<,001
Doubt	1	<,001
Emphasize	,806	,003
Establish	1	<,001
Expect	,885	<,001
Feel	,963	<,001
Guess	1	<,001
Highlight	1	<,001
Hope	,959	<,001
Identify	1	,083
Indicate	1	<,001
Interpret	1	,014
Judge	1	<,001
Know	,795	<,001
Look (like)	,956	<,001
Point out	,890	<,001
Presume	1	<,001
Propose	1	,046
Prove	,886	<,001
Realize	,841	<,001
Reckon	1	<,001
Recognize	1	,005
Regard	1	<,001
Reveal	1	,008
Show	,858	<,001
Speculate	1	<,001
Stress	1	<,001
Suggest	1	<,001
Suppose	1	<,001
Tend	1	<,001
Think	,985	<,001

Appendix 3: Moderate and High Inter-rater Reliability for ELVs in JPICLE

Moderate and High Inter-rater Reliability for ELVs in JPICLE

JPICLE	Cohen's Kappa (κ) value	p
Appear	,769	<,001
Argue	,643	,002
Assume	,707	<,001
Believe	1	<,001
Claim	1	,003
Conclude	,800	,010
Confirm	,588	,088
Consider	,933	<,001
Convince	1	<,001
Demonstrate	1	<,001
Describe	1	,003
Doubt	,841	<,001
Emphasize	1	<,001
Establish	1	<,001
Estimate	1	,014
Expect	,868	<,001
Feel	,861	<,001
Guess	,627	,003
Hope	,924	<,001
Identify	1	,014
Indicate	1	,025
Interpret	1	,083
Judge	,733	<,001
Know	,910	<,001
Look (like)	,976	<,001
Point out	1	,014
Presume	1	,046
Propose	1	<,001
Prove	1	<,001
Realize	,955	<,001
Reckon	1	<,001
Recognize	1	<,001
Regard	1	<,001
Reveal	1	,046
Seem	1	<,001
Show	,959	<,001
Stress	1	<,001
Suggest	1	<,001
Suppose	1	<,001
Suspect	1	,025
Tend	1	<,001

Appendix 4: Moderate and High Inter-rater Reliability for ELVs in TRICLE

Moderate and High Inter-rater Reliability for ELVs in TRICLE

TRICLE	Cohen's Kappa (κ) value	p
Appear	1	<,001
Assume	1	<,001
Assure	1	,083
Believe	,937	<,001
Claim	,932	<,001
Conclude	1	,014
Confirm	1	<,001
Consider	,952	<,001
Convince	1	,046
Demonstrate	1	,025
Describe	1	,005
Doubt	1	<,001
Emphasize	1	<,001
Establish	1	<,001
Estimate	1	,083
Expect	,892	<,001
Feel	,822	<,001
Guess	,714	,035
Highlight	1	<,001
Hope	1	<,001
Indicate	,783	,011
Identify	1	,157
Infer	1	<,001
Interpret	1	<,001
Know	,904	<,001
Look (like)	,906	<,001
Point out	1	<,001
Propose	1	,157
Prove	1	<,001
Realize	1	<,001
Recognize	1	<,001
Regard	1	<,001
Reveal	1	,014
Seem	1	<,001
Show	,810	<,001
Speculate	1	,157
Stress	1	<,001
Suggest	1	<,001
Suppose	1	<,001
Tend	1	<,001

Appendix 5: All Epistemic Lexical Verbs Used in the Four Corpora

All Epistemic Lexical Verbs Used in the Four Corpora

LOCNESS		NICLE		JPICLE		TRICLE	
Verbs	<i>nf</i>	Verbs	<i>nf</i>	Verbs	<i>nf</i>	Verbs	<i>nf</i>
Show	9,2	Think	25,55	Think	52,82	Think	20,41
Believe	8,95	Seem	10,46	Believe	4,62	Know/n	6,41
Seem	8,12	Believe	7,86	Seem	4,14	Believe	5,72
Think	7,76	Know/n	4,79	Know/n	3,92	Show	3,59
Feel	5,93	Feel	3,97	Tend	2,86	Seem	3,45
Know/n	4,07	Suppose	3,76	Show	2,55	Consider	2,95
Argue	3,52	Claim	3,63	Hope	1,8	Claim	2,54
Prove	2,96	Show	3,45	Consider	1,54	Argue	1,45
Consider	2,66	Argue	3,02	Suppose	1,49	Suppose	1,4
Realize	2,46	Look (like)	2,59	Feel	1,36	Expect	0,9
Claim	2,27	Consider	2,03	Look (like)	1,19	Feel	0,81
Appear	1,77	Tend	2,03	Regard	1,14	suggest	0,81
suggest	1,41	Guess	0,99	Realize	1,05	Regard	0,54
Tend	1,35	Prove	0,86	Argue	0,83	Tend	0,54
Expect	1,13	Expect	0,77	Expect	0,79	Realize	0,4
Assume	1,1	Hope	0,77	Assume	0,66	Prove	0,36
Suppose	1,1	suggest	0,73	Guess	0,66	Assume	0,31
Point out	0,8	Regard	0,6	Judge	0,66	Indicate	0,31
Describe (as)	0,69	Indicate	0,51	Doubt	0,61	Emphasize	0,27
Hope	0,69	Point out	0,47	suggest	0,48	Guess	0,27
Look (like)	0,63	Realize	0,47	Prove	0,39	Look (like)	0,27
Reveal	0,63	Appear	0,43	Recognize	0,39	Point out	0,27
Conclude	0,61	Conclude	0,38	Propose	0,39	Describe (as)	0,22
Regard	0,58	Convince	0,38	Emphasize	0,35	Hope	0,18
Demonstrate	0,55	Emphasize	0,38	Conclude	0,26	Recognize	0,18
Emphasize	0,55	Stress	0,34	Point out	0,22	Conclude	0,13
Recognize	0,52	Assume	0,3	Claim	0,17	Convince	0,13
Propose	0,47	Doubt	0,25	Indicate	0,17	Demonstrate	0,13
Convince	0,36	Reveal	0,17	Appear	0,13	Appear	0,09
Establish	0,3	Confirm	0,12	Convince	0,13	Estimate	0,09
Stress	0,3	Describe (as)	0,12	Reveal	0,13	Judge	0,09
Indicate	0,27	Presume	0,12	Suspect	0,13	Infer	0,08
Interpret	0,27	Speculate	0,08	Confirm	0,08	Assure	0,04
Guess	0,19	Propose	0,08	Describe (as)	0,04	Deduce	0,04
Doubt	0,13	Assure	0,04	Establish	0,04	Highlight	0,04
Identify	0,13	Establish	0,04	Estimate	0,04	Identify	0,04
Suspect	0,13	Identify	0,04	Identify	0,04	Reveal	0,04
Highlight	0,08	Interpret	0,04	Interpret	0,04	Speculate	0,04
Judge	0,08	Judge	0,04	Stress	0,04	Propose	0,04
Presume	0,08	Recognize	0,04	Presume	0,04	Attest	0
Infer	0,08	Attest	0	Reckon	0,04	Confirm	0
Assure	0,05	Demonstrate	0	Attest	0	Doubt	0
Confirm	0,05	Deduce	0	Assure	0	Establish	0
Estimate	0,05	Estimate	0	Demonstrate	0	Look as if	0
Look as if	0,02	Highlight	0	Deduce	0	Interpret	0
Speculate	0,02	Look as if	0	Highlight	0	Stress	0
Reckon	0,02	Reckon	0	Look as if	0	Presume	0
Attest	0	Suspect	0	Speculate	0	Reckon	0
Deduce	0	Infer	0	Infer	0	Suspect	0

Appendix 6: The Case of Epistemic Modal Auxiliary *May*

The Case of Epistemic Modal Auxiliary *May*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Also	37	11	12	13
But	90	66	49	69
On The Other Hand	1	0	0	2
While	7	7	0	4
Whilst	2	1	0	0
Whereas	3	2	0	0
Although	19	8	3	0
As Well	4	0	0	1
Objective	456	254	266	296
Subjects	3-I	4-I	12-I	1 -I
	16-We	14-We	22-We	10-We
Negation	65	20	38	30
Even	13	1	2	1
Of Course	1	1	0	0
Only	4	0	1	0
If	50	23	2	30
Well	12	0	0	0
Possibly	1	0	1	1
Yet	1	1	1	0
In Fact	2	0	0	0
Or	26	5	9	8

Appendix 7: The Case of Epistemic Modal Auxiliary *Might*

The Case of Epistemic Modal Auxiliary *Might*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Also	2	8	1	2
But	11	59	17	5
On The Other Hand	0	3	0	0
While	1	2	0	1
Whilst	0	2	0	0
Although	0	5	0	0
As Well	3	5	2	0
Objective	79	260	118	54
Subjects	4-I	2-I	6-I	0-I
	2-We	9-We	6-We	4-We
Negation	4	29	15	7
Even	0	15	0	1
Of Course	0	1	0	0
If	21	25	26	11
Well	0	1	0	1
Possibly	0	1	0	0
Yet	0	1	2	0
Or	10	13	6	5

Appendix 8: The Case of Epistemic Modal Auxiliary *Will*

The Case of Epistemic Modal Auxiliary *Will*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Subjects	1-I	9-I	18-I	7-I
	32-We	39-We	52-We	20-We
Negation	103	52	46	37
If	179	81	125	96
Probably	13	36	10	16
Always	23	62	3	24
Perhaps	13	4	1	5
Eventually	26	8	1	1
Surely	7	1	5	1
Also	68	20	11	28
Soon	20	4	10	3
Never	30	19	7	16
No Longer	5	0	0	0
Then	40	19	20	23
Only	25	9	4	1

Appendix 9: The Case of Epistemic Modal Auxiliary *Would*

The Case of Epistemic Modal Auxiliary *Would*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Subjects	60-I	79-I	42-I	8-I
	23-We	25-We	25-We	5-We
Negation	136	70	24	41
If	346	207	104	77
When	54	60	24	25
Probably	45	50	5	4
Always	2	2	0	0
Perhaps	14	2	2	2
Eventually	8	2	2	0
Also	108	55	8	7
Soon	4	2	4	0
Never	11	14	6	1
No Longer	15	0	1	0
Then	93	41	4	9
Only	20	1	2	6

Appendix 10: The Case of Epistemic Modal Auxiliary *Must*

The Case of Epistemic Modal Auxiliary *Must*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Be	7	30	35	24
Must Have P3	5	10	4	2
But	2	3	4	5
Not	0	0	0	7



Appendix 11: The Case of Epistemic Modal Auxiliary *Should*

The Case of Epistemic Modal Auxiliary *Should*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Be	13	7	11	14
Must Have P3	1	2	0	2
But	4	1	2	4
Not	2	2	2	4
If	4	2	3	4

Appendix 12: The Case of Epistemic Modal Auxiliary *Could*

The Case of Epistemic Modal Auxiliary *Could*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Be	186	66	21	25
Have P3	25	30	2	5
But	57	30	9	13
Not	21	9	14	12
Also	23	7	6	2
Then	24	9	2	2
If	74	42	29	20



Appendix 13: The Case of Epistemic Lexical Verb *Think*

The Case of Epistemic Lexical Verb *Think*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Subjects	I-111	I-339	I-1062	I-171
	We-10	We-18	We-21	We-18
That	118	156	460	207
Also	19	40	32	24
Not	18	48	82	19
But	22	87	188	50

Appendix 14: The Case of Epistemic Lexical Verb *Believe*

The Case of Epistemic Lexical Verb *Believe*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Subjects	I-63	I-118	I-70	I-57
	We-3	We-3	We-1	We-1
That	185	109	76	108
Not	19	4	6	9
But	28	23	18	23
Also	16	4	11	5
It is believed	4	0	2	9
Still	14	11	1	7

Appendix 15: The Case of Epistemic Lexical Verb *Show*

The Case of Epistemic Lexical Verb *Show*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
That	84	47	16	34
It shows	4	5	8	7
As ... shown	5	0	1	0
Not	7	0	2	3
But	27	13	4	8
Also	22	6	4	6
It... shown	11	2	0	0



Appendix 16: The Case of Epistemic Lexical Verb *Seem*

The Case of Epistemic Lexical Verb *Seem*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
It	52	48	42	20
That	19	11	19	5
Seem to + Verb	142	127	48	36
Seem + Adj	44	25	8	12
Not	18	11	5	1
Would	16	2	0	0
May	15	17	2	4
Like	11	20	2	2
Very	8	3	7	3
As	9	11	3	4



Appendix 17: The Case of Epistemic Lexical Verb *Know*

The Case of Epistemic Lexical Verb *Know*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Known	37	6	5	28
I and We	23 - 5	19 – 10	22 – 7	10 – 19
That	58	53	30	72
As	17	0	1	9
As ... know(s)	1	3	15	32
Not	5	4	6	8



Appendix 18: The Case of Epistemic Lexical Verb *Feel*

The Case of Epistemic Lexical Verb *Feel*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
I and We	69 - 3	27 – 3	13 - 1	2 - 1
Not	4	5	2	0
That	130	66	14	13
Like	5	0	0	0



Appendix 19: The Case of Epistemic Lexical Verb *Claim*

The Case of Epistemic Lexical Verb *Claim*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
It + Passive	5	2	0	6
To	13	2	0	0
Subjects	They - 5	Some - 6	They - 0	I - 6
That	43	70	3	49
Not	3	2	0	1

Appendix 20: The Case of Epistemic Lexical Verb *Argue*

The Case of Epistemic Lexical Verb *Argue*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
It + Passive + That-clause	24	6	0	6
Would	11	7	1	1
Can	2	8	0	1
People	9	3	3	2
Not	3	1	1	1
That	96	55	9	21



Appendix 21: The Case of Epistemic Lexical Verb *Consider*

The Case of Epistemic Lexical Verb *Consider*

Collocations	LOCNESS	NICLE	JPICLE	TRICLE
Passive	45	24	14	35
Not	6	7	1	9
It is considered	10	2	2	8
That	11	1	8	13
To Be	11	4	3	9
As	3	5	1	25
Adjectives	15	11	3	6



CURRICULUM VITAE

She began to study in the English Language and Literature of the Faculty of Letters, Karadeniz Technical University. In 2018-2019, she attended an Erasmus exchange program and studied at the Pedagogy Faculty at Jan Evangelista Purkyně University in Usti and Labem, Czechia. In 2019-2020, she worked as an English teacher at Trabzon Manchester Language Schools. She holds her BA in English Language and Literature of Faculty of Letters, Karadeniz Technical University. After her graduation, she began to study for her MA in the Applied Linguistics Program in Karadeniz Technical University. Between the July of 2022 and September of 2022, she did her internship as a content writer for two months at MBC Home, an agency in Lisbon, Portugal. Her main areas of interest include applied linguistics, corpus linguistics, learner corpora, ESP, EAP, EFL students' writing, academic writing, language anxiety, and phraseology.