

HEDGES AND BOOSTERS: L1 AND L2 SPEAKERS' COMPREHENSION AND  
PRODUCTION

A Thesis submitted to the faculty of  
San Francisco State University  
In partial fulfillment of  
the requirements for  
the Degree

Master of Arts

In

English: Linguistics

by

Minsun Lee

San Francisco, California

May 2020

Copyright by  
Minsun Lee  
2020

## CERTIFICATION OF APPROVAL

I certify that I have read Hedges and Boosters: L1 and L2 Speakers' Comprehension and Production by Minsun Lee, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Master of Art in English: Linguistics at San Francisco State University.

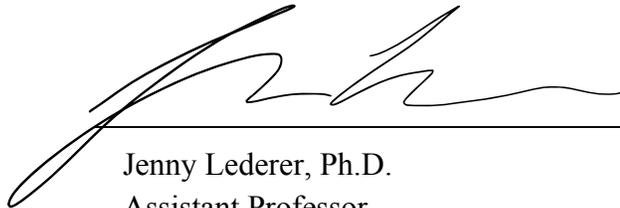
DocuSigned by:

Anastasia Smirnova

7DB2FBA96666439...

---

Anastasia Smirnova, Ph.D.  
Assistant Professor



---

Jenny Lederer, Ph.D.  
Assistant Professor

## Hedges and Boosters: L1 and L2 Speakers' Comprehension and Production

Minsun Lee  
San Francisco, California  
2020

To make a gentle, yet convincing, claim requires you to soften down or raise up your argument. Rhetorical devices are often used for following purposes: hedges are used to adjust one's statement in modest and polite way by expressing doubt and boosters are to strengthen one's argument by expressing certainty. Hedges and boosters are considered very important in English academic writing, so they have often been discussed by linguists concerning their definitions and distributions (Lakoff, 1972; Hinkel, 2002; 2005; Holmes, 1984; Hyland, 1994; 1998; 2005). A comparison of native (NS) and non-native speakers' (NNS) usage patterns has been one of many popular studies and corpus analysis was scholars' favorite method to study it. However, there has been only a few experimental studies done to measure their perceptions and productions of hedges and boosters. This study, therefore, carries out two studies: (1) a perception test that measures readers' perceptions towards hedges and boosters and (2) a computational corpus analysis of native and non-native speakers' argumentative writings to compare their production of the two devices. The test results show that native and non-native speakers generally have similar perceptive patterns besides some exceptional cases. In terms of production, however, non-native speakers show distinctive distributions from native speakers. The findings from the two studies invite us to consider a potential discrepancy between comprehension and production of hedges and boosters, regardless of one's native

language background. Thus, the importance of teaching epistemic markers for academic writing is raised. These comparative studies will be a weighty addition to previous conflicting research findings in this topic and the methodologies I used - experimental and computational - will gear up more robust research and practices to follow.

I certify that the Abstract is a correct representation of the content of this thesis.

DocuSigned by:

Anastasia Smirnova

7DB2FBA96666439

---

Chair, Thesis Committee

5/12/2020

---

Date

## ACKNOWLEDGEMENTS

I thank God my Father for His abundant grace and bountiful supply. Thanks to His arrangements, I was able to meet many wonderful people during my M.A. Linguistics program and finish the course with their help.

I would like to express my gratitude to Professor Anastasia Smirnova, my research advisor, for guidance and support through the process of producing this thesis. I would also like to show appreciation to Professor Jenny Lederer for the additional support in writing this thesis. I extend thanks to Yohan Jo, a Ph.D. candidate at Carnegie Mellon University, for helping me to write codes for Study 2. Also, much thanks to a professional development grant awarded to Anastasia Smirnova. Study 1 was able to be conducted by its provided fund.

I also give thanks to my colleagues in the M.A. Linguistics and the M.A. TESOL program for encouraging and helping me to complete the journey together. I am also grateful to my friends in South Korea who volunteered or helped me gather participants for the experimental study. It is never enough to say *Thank you* to my family and my spiritual families in Berkeley, Palo Alto, and Seoul for the cares and the prayers. Finally, no word can fully express my gratitude to my best companion, Hyung-sin Kim! I would like to give thanks to him for supporting me in every way as a husband and advising me as a professor in data science while conducting this research.

## TABLE OF CONTENTS

List of Table.....	vii
List of Figures.....	viii
List of Appendices.....	ix
1 Introduction.....	1
2 Hedges and Boosters: Theoretical Background.....	3
2.1 Evaluation and Interaction in Writing.....	3
2.2 Stance Markers: Hedges and Boosters.....	4
2.3 Stance Markers for Non-native Speakers.....	6
2.4 Corpus Studies on Non-native Speakers' Uses of Hedges and Boosters.....	7
2.5 Earlier Experimental Studies.....	9
2.6 Needs for Further Studies.....	14
2.7 Needs for Teaching Hedging Strategies to Non-native Speakers.....	17
3 Study 1 - Perception of Hedges and Boosters by native speakers of English and Korean English learners.....	18
3.1 Methodologies.....	18
3.1.1 Participants.....	18
3.1.2 Material.....	21
3.1.3 Design and Procedure.....	28
3.2 Results and Discussion.....	29
3.3 Limitations and Further Discussion.....	40
4 Study 2 - Corpus analysis of argumentative writing of native speakers of English and Korean English learners.....	43
4.1 Methodologies.....	43

4.1.1 Material .....	43
4.1.2 Design and Process .....	48
4.2 Findings & Discussion .....	49
4.2.1 Corpus Analysis Results .....	49
4.2.2 Accuracy Test Results .....	57
4.3 Limitations and Further Discussion .....	65
4.3.1 Topic dependency .....	65
4.3.2 The impact of teaching .....	66
5 Conclusion .....	68
References.....	72
Appendices.....	84

## LIST OF TABLES

Table	Page
1. English NS participants' background .....	19
2. English NNS participants' background .....	19
3. LOCNESS - Native speakers' writing corpus .....	48
4. YELC - Korean learners' writing corpus .....	48
5. Frequency of epistemic markers in L1 and L2 writings .....	50
6. Ten most common hedges and boosters .....	54
7. Most common part of speech .....	55
8. Accuracy test results .....	57
9. Accuracy scores for booster items .....	58
10. Accuracy scores for booster items .....	58

## LIST OF FIGURES

Figures	Page
1. Perception Test Sample Question .....	27
2. Recoded values for target-neutral ordered questions .....	30
3. Recoded values for neutral-target ordered questions .....	30
4. L1 and L2 perception scores for hedges and boosters .....	32
5. L1 and L2 perception scores for hedges .....	34
6. L1 and L2 perception scores for boosters .....	34
7. Perception scores for hedges and boosters between the L2 groups with and without living abroad experiences .....	39
8. Top Twenty hedges and boosters .....	54

## LIST OF APPENDICES

Appendix	Page
1. Survey Questions .....	84
2. List of Hedges and Boosters in Corpus Analysis .....	88
3. Statistical Results for Study 1 .....	90

## Chapter1. Introduction

“Apparently, he is funny.”

How would you interpret this utterance of describing one guy to be funny? Many Americans’ initial reaction would be that the guy is very humorous unlike the speaker has been told. The adverb, *apparently*, is almost like ‘evidently’ or ‘obviously’ in this interpretation. On the other hand, others may perceive some sarcasm here at first sight that the guy is not as funny as the speaker has heard. This short statement with a rhetorical device, *apparently*, can often cause misunderstandings between a speaker and listeners or a writer and readers. This rhetorical device known as a hedge contains much history and culture among English native speakers that people from other language backgrounds can hardly share, unless told.

Hedges are used for presenting claims in cautious, modest, and accurate ways. They soften down one’s voice so that they could meet listeners’ or reader’s expectations and gain acceptance from the discourse community (Hyland, 1996). On the other side is a booster which is as critical as a hedge in English. It strengthens, intensifies, and adds certainty to one’s statement.

Hedges and boosters are very deeply rooted in English discourse practice. They are especially important skills in writing as well as speaking. Normally, writers would want to attract readers’ agreement and thus, they use these means to place their new statements carefully into the existing community (Hyland, 1994). In other words, these

two different types of devices adjust the degree of probability of a statement and by doing so they reflect a relation between a writer and readers (e.g., Coates 1987; Holmes 1984; Myers, 1989; Skelton 1988b). Such pragmatic aspects of communication, however, are vulnerable to cross-cultural differences. First language and cultural backgrounds, lack of exposure to second language community, and shortage of early cultivated intuitions, they all often interrupt L2 speakers' understanding of discursive uses of hedges and boosters and cause issues with the application of them.

The view of hedges and boosters between NS and NNS in this paper emerges from two studies: (1) an experimental study that tests comprehension of how hedges and boosters are understood by native and non-native speakers and (2) a corpus study that analyzes how native and non-native speakers of English use hedges and boosters in argumentative writings. This paper will examine the critical functions of hedges and boosters, by carrying out comparative studies between native speakers and non-native speakers concerning their comprehension and production of hedges and boosters. The purpose of these studies is to explore non-native speakers' perceptions and usage patterns of the two rhetorical devices compared to the native speakers', so as to provide directions for more robust research in this topic and academic English curriculum development in practice.

## **Chapter 2. Hedges and Boosters: Theoretical Background**

### **2.1 Evaluation and Interaction in Writing**

The ways writers use language to express their opinions and interact with their readers have long been recognized as an important feature of language. To account for these linguistic resources, different terms have been used: Evaluation (Hunston & Thompson, 2000), attitude (Halliday, 1994), epistemic modality (Hyland, 1998), stance (Biber & Finegan, 1989; Hyland, 1999), meta-discourse (Crismore, 1989; Hyland and Tse, 2004), and interaction (Hyland, 2005) are several examples. However, what these terms all have in common is that they acknowledge the fact that successful writing depends on how writers establish their stance, conduct interpersonal negotiations with readers and balance their claims through rhetorical choices.

Encompassing all these concepts, Hyland (2005) devised a new taxonomy for academic interaction which consists of the two elements – stance and engagement. Stance, according to Hyland, is “an attitudinal dimension that includes features which refer to the ways writers present themselves and convey their judgements, opinions, and commitments (p.4).” Engagement refers to the “alignment dimension (p.4)”, to borrow Hyland’s term, which allows writers to relate to their readers with respect to their arguments. Stance includes:

1. *Hedges*: The writer's unwillingness to present proposition in confirmed tone, such as about, perhaps, etc.

2. *Boosters*: Language of certainty. Some examples are clearly and definitely.

3. *Attitude markers*: The writer's appraisal of proposition. For example, I agree, surprisingly, etc.

4. *Self-mentions*: The presence of author in terms of first person pronouns and possessives. Some examples are I, we, our, my, etc.

While they all play different roles in effective writing, the first two elements - hedges and boosters - were regarded as “requisites (p.2)” for Anglo-American English prose (Hinkel, 2005). Myers (1989) describes these two as to be conventional and necessary features in claim-making writing. As such, there were great attention to softening and intensifying devices and thus, research into various types and functions of them emerged in great numbers in the late 20th Century (e.g. Chafe, 1985, 1986; Chafe & Danielewicz, 1987; Hermeren, 1978; Huebler, 1983; Holmes, 1984).

## **2.2 Stance Markers: Hedges and Boosters**

Hedges and boosters are major elements of positioning stance (Hyland, 2005). Hyland highlights that hedges and boosters are writer-oriented features of interaction, which project the possible accuracy or credibility of a writer's claim. Especially the term, ‘hedge’ first gathered attention since Lakoff's (1972) introducing it as “words whose job is to make things fuzzier or less fuzzy (p.195).” It is to mitigate the author's voice so that

it conveys politeness, modesty, tentativeness, and openness to the readers' views.

Boosters, on the other hand, express certainty of the writer's assertions. Holmes (1995) explains that the boosting devices function to intensify and emphasize a proposition with confidence. While it is relatively easier to understand that boosters are useful for opinionated statements, since they would help the authors to build up their positions strongly, hedges might not sound appropriate for making claims at firsthand. However, as they mark and adjust the stance of the author, many scholars have agreed on the uses of hedging to be necessary in texts that include claim-making (Myers, 1989). Hyland (2005) also stresses that the writers should not only be able to weigh up, but tone down their arguments to powerfully gain acceptance for claims. In his earlier study, Hyland (1998) highlighted the critical role of hedging in scientific writing to gain ratification for writers' claims. He found that hedges were especially denser in the discussion section, which reflected the need for claims to be presented tentatively rather than assertively in order to get approval from readers. Similar research has been conducted by other scholars (Abbuhl, 2006; Dafouz-Milne, 2008) and they also showed that an argument presented through hedging devices was considered more convincing and credible. In short, effective use of hedges and boosters can pave a way for the authors' claims to be made firmly, yet composedly.

### **2.3 Stance Markers for Non-native Speakers**

While we have seen that stance markers are important in English writing, it could be problematic for non-native English writers to convey statements with proper degree of doubt and certainty (Hyland & Milton, 1997). Many propose that expressing proper degrees of certainty is difficult for language learners (Allison, 1995; Hyland & Milton, 1997) and a great number of studies support them with their research findings that direct and unskilled writing was more typical among second language writers, even compared to native adults who were relatively poor at writing (Bloor & Bloor, 1991; Silva, 1993; Skelton, 1988a). Such lack of rhetorical techniques seem to lead ESL (English as a Second Language) or EFL (English as a Foreign Language) writers to make unjustifiably strong claims. For example, Hu et al (1982) and Allison (1995) found that students in China and Hong Kong to be more direct and authoritative in tone than native speakers. Students from the other side of the globe, such as Arab, Dutch, and French, also appear to have difficulties in employing hedging strategies in academic writing (Robberecht & Peteghem, 1982; Scarcella & Brunak, 1981). Such a linguistic limitation was often explained by the absence of teaching, cultural differences, limited range of vocabulary (Lee, 2006), and the fact that different languages use different linguistic means for expressing certainty (Hyland, 1998).

## **2.4 Corpus Studies on Non-native Speakers' Uses of Hedges and Boosters**

Although many have addressed non-native speakers' difficulties of using hedging strategies in written language as discussed above, corpus-based research that compared NSs' and NNSs' uses of hedges and boosters often have shown disagreement (e.g. Granger, 1998; Hinkel, 2002; Huh & Lee, 2016; Kim, 2009; Lorenz, 1998).

Granger (1998) analyzed French English learners' writings from the International Corpus of Learner English (ICLE) corpus and the writings collected from English speaking undergraduate students in order to examine non-native speakers' uses of English intensifiers in comparison with native speakers'. He counted the numbers of tokens and types of intensifiers from each language group's corpus and found out that both tokens and types of French corpus were fewer than those of native speakers'. Based on this numerical data, Granger highlights that non-native speakers' use of amplifiers is not only less frequent but also less diverse. He, thus, claims the need of teaching to overcome non-native speakers' underuses of amplifiers. Lorenz (1998), on the other hand, looked into German learners' corpus to compare with English native speakers' writings. In his studies, both hedging and intensifying in adjectival quality were investigated. In contrast to Granger's finding, Lorenz found that German learners of English use much more intensifiers than the native speakers of English. However, the patterns show that non-native speakers' intensifiers were mostly repetitive, meaning particular words were overused. The contrastive findings between Granger (1998) and

Lorenz (1998) in the frequency pattern suggests that cross-cultural differences between German and French might have influenced their L2 writing.

However, given the same national participants, Kim (2009) and Choi and Ko (2005) also arrived at different conclusions. Kim claims that Korean undergraduate L2 writers employ far less interactional meta-discourse features than the British journalists, but Choi and Ko argue that Korean postgraduate writers in the Applied Linguistics field use a greater number of hedges in their academic journals compared to L1 corresponding experts. Both these studies, however, display limitations in their methodologies. In Kim's studies, for instance, it is not surprising that British journalists (L1) would display a stark difference in their quality of the writings from Korean undergraduate students (L2). The L1 professional writers should be notably more competent in English argumentative writing than the L2 college students. In terms of equal educational level between L1 and L2, Choi and Ko's research is meaningful in that they compared the advanced L2 writers with the native speakers from the same Linguistics field. However, this L2 experimental group does not seem to represent average Korean English learners. These linguistics-majored Ph.D. students are likely to have much knowledge on using hedges in academic papers.

Accordingly, as an attempt to establish a more solid theory, Kang (2017) collected both the L1 and L2 writing samples from undergraduate students of diverse academic backgrounds. For L1's data, Kang chose the texts written by Seoul National University students, who should have gone through public English education pretty well but had not

learned advanced English academic writing yet. Kang manually coded hedges and boosters and meticulously analyzed each corpus to arrive at the conclusion that L2 groups were generally not equipped with epistemic rhetorical devices. Hedges were found to be especially rare in L2's writings when L1 students used even more numbers of hedges than boosters. This article draws on such important qualitative findings in Korean English learners' usage of rhetorical devices, but the number of essays she analyzed was relatively small to show consistent frequency patterns. Also, her randomly selected writings cannot reflect the relationships between the writers' English ability and their usage patterns of hedges and boosters. Therefore, Kang's (2017) study still leaves the need for more solid corpus studies to validate its finding.

### **2.5 Earlier Experimental Studies**

While there have been many corpus-based studies as seen above, there are relatively few experimental studies on this topic. One rare exception is Hyland's (2000) noticing modifiers study. His research re-invites the Low's (1996) Lexical Invisibility Hypothesis that readers often do not notice certain hedges or boosters, and this unawareness causes distortion of their understanding. Low designed his study in a way that he can see the participants' reactions to six boosters (*very, extremely, far, full, never, and consistently*) and two hedges (*seem and tend*) which were often found in questionnaires. Low measured readers' recognition of hedge and booster items by think-

aloud procedure, in which participants would report their thoughts for choosing answers while doing the task of reading and answering the questions. In this way, he found that most rhetorical devices were little or never noticed by the participants when reading the question statements. Only four boosters (*very, extremely, far, and full*) were well recognized by the readers. Through this discovery, he has claimed that such participants' unawareness of certain rhetorical devices might lead to misinterpretations of a question statement, deviated from the question designers' intentions. This is seen a critical matter as it may possibly lead to readers' incorrect answers.

Hyland (2000) supports Low's hypothesis and explores this more with some revised methodologies. He discontinued using Low's think-aloud procedure, where the participants would report their thoughts while working on a given task. While this method has been widely used by many researchers (eg. Afflerbach & Johnson, 1984; Lennon, 1989; Smagorinsky, 1994; Talbot, 1992), Hyland noticed some drawbacks from the past studies that such reports only provided a partial record of processes (Hayes & Flower, 1983), and that the act of verbal reporting may have caused the participants' cognitive process to be distorted (Stratman & Hamp-Lyons, 1994). Also because concurrent think-aloud tasks might be too difficult or distracting for the second language students to carry out their tasks and report their trains of thoughts in L2 at the same time, Hyland switched the method to the retrospective one.

Hyland (2000) gathered fourteen Cantonese speakers from Hong Kong, who were fluent in English. They were given a text to read, and during the interview they were

asked to answer 15 questions based on the reading. The interviewer interacted with the participants to elicit subjects' awareness of hedges and boosters. Twenty minutes were given to complete the questions then were asked to explain how they arrived at the answers. Having a twenty-minute alone time seems to give sufficient time for them to reflect on each question and reduce cognitive overload. Unlike think-aloud procedure, this way may help avoid some risks of either task or report being interfered with by simultaneous performance. Through this way, Hyland arrives at the conclusion that non-native speakers are likely to fail to interpret the writer's assessment of certainty. While Hyland asked students questions to assess their recognition of hedges and boosters, their successful attention to the devices was only limited to 20%, 50 out of the 210 cases.

While this is a meaningful attempt to improve the interview environment for non-native speakers to fully display their understanding, this way still does not seem to be the best option as it does not eliminate all the possibilities of memory loss, leaving some risks of distortion of the report. As the questions are given after the passage reading, participants are most likely to be focused on reading itself than how they arrived there. They might not be able to fully give their thoughts that led to their interpretations. Moreover, Šeškauskienė (2008) brings in a contradicting view in her qualitative research of Lithuanian learners a few years later. She claims that Lithuanian English learners who were in their advanced levels, in fact, noticed hedges and boosters well, unlike what was reported in earlier studies (Hyland, 2000; Low, 1996). Furthermore, Hyland's (2005) study might be more informative if the same test was conducted to native speakers as

well. If this is really the noticing matter, there is always the case that native speakers result in similar consequences. It could be just strategic reading comprehension skills that made the learners intentionally or unconsciously miss those relatively smaller items to more effectively process the reading task. Therefore, this same phenomenon can happen to native speakers in this experimental environment, where they are supposed to focus on the content of the reading passage. They are likely to approach the text from top (broader understanding) to down (details) in order to get the gist of the story. In other words, it is an open question whether the result could have been compatible between native and non-native speakers. Nevertheless, despite such limitations, this study seems to fulfill what the author intended to examine, and it provides meaningful finding that some words, such as *definitely* and *clearly*, are more salient than other words like *suggest*, to non-native speakers.

Low's (1996) and Hyland's (2000) experimental studies focus on whether or not the subjects notice the rhetorical devices while reading. They do point out that non-native speakers often fail to recognize rhetorical devices, which in fact play quite significantly in shaping the meaning. However, this finding can be true only based on the assumption that non-native speakers are knowledgeable with these items, but this assumption was not addressed in any way. They do not explain whether the L2 learners actually 'know' that these devices make differences in their interpretations, given more time. Do the interviewees have general knowledge about hedging and boosting in English writing? If the interview were designed for them to pay more attention to these items than merely to

focus on reading comprehension, would they have shown different outputs? While noticing is critical in many fast-driven tasks like a job interview, tests, or surveys, these unresolved questions above seem to enlighten what needs to be discussed further to make this study even worthier.

There are other useful studies which investigated people's perception of hedges and boosters. Lewin (2005)'s study is one example. It was conducted in two steps. First, Lewin asked the writers of articles - eight native and two non-native speakers - to identify hedges in their own texts and to address their motivation for using them. These ten academicians' fields of study were physical sciences (neuroscience, medicine, physics, chemistry, biochemistry, atmospheric science, and geophysics) or social sciences (sociology and educational psychology). Then, the same articles were presented to Ph.D. candidates in an EAP (English for Academic Purpose) class and they were requested to identify hedges as well.

In this study, a very interesting finding was discovered that the authors and the readers showed a great divergence in their identification of hedges and boosters. In general, the readers identified more hedges than the author hypothesized. In addition, there was also a big gap shown between the authors' choices and general linguists' definitions of hedges. The scholars' choices of hedges did not agree with common definitions of hedges in the previous literature. For instance, some authors (both native and non-native speakers) often reported that their motivation for using hedges was to heighten claims rather than to tone them down. It was common among those authors that

politeness was not cited to be a general motivation of hedges. Overall, it seems that their selection of hedges does not match with what linguists have defined.

As a result, Lewin's (2005) research gives the implication that hedges or boosters could be even 'fussier' than Lakoff's (1972) claim. It might be that people have different perceptions of hedges and boosters, when not explicitly taught. Ph.D. candidates are expected to be relatively knowledgeable in writing academic papers, but they reported different interpretations than the intended meanings. Given that even the writers showed discrepancy in their using hedging skills from what have been commonly known linguistically, people without academic writing backgrounds might use the rhetorical devices even less in number and inaccurately. Moreover, considering the fact that EAP course is dominantly preferred by non-native speakers, the result could reflect more of L2 speakers' interpretations. This evokes the need for further studies to examine both native and non-native speakers' perceptions of hedges and boosters. Since many native academicians also presented their motivations differently from linguists' interpretations in the study above, investigating native speakers' perception would be also meaningful.

## **2.6 Needs for Further Studies**

Overall, the findings from those studies discussed above (Hyland, 2000; Lewin, 2005; Low, 1996) urge the needs for more exploratory research on the people's comprehension as well as production of hedges and boosters. Let alone noticing, the

comprehension of hedging skills and the actual usage of them are very critical matters since improper understanding and uses may distort the meaning and lead to divergent interpretations. This may operate as obstacles for successful communication with the journal reviewers as well as general readers.

Therefore, more experimental studies that measure readers' comprehension of epistemic markers can be considered worthy to the linguistics fields because there have been very few attempts to explore people's perceptions towards hedges and boosters. They can also contribute to ESL or EFL instructions, as they could extend the practitioners' understanding of the important roles of hedging skills and might give confirmation on the need of teaching them. Even more so, the perception studies are valuable for future compilation of hedges and boosters. As Lewin (2005) points out, the disagreement among linguists, authors, and readers imply the inconsistency in the definition of hedging. Farrokhi & Emami (2008) compiled an inventory of hedging and boosting lexical items from traditional studies, yet those items have not been examined one by one to be confirmed as functioning as genuine hedges or boosters. There can be some items that are poly-pragmatic and context-dependent (Hyland, 1998). These items can switch their functions from hedges to boosters or vice-versa depending on its context. Therefore, it would be very critical and useful to carry out a perception test over the list to see if each item truly fits their rhetorical categories.

Not only is there a need for a comprehension test, but also it seems to require more in-depth understanding of learners' production of hedges and boosters. There have

been many attempts to compare L1 and L2's uses of rhetorical devices (e.g. Granger, 1998; Huh & Lee, 2016; Kim, 2009; Lorenz, 1998). However, they would either show limitations in the design of the studies or the scope of data. For example, some studies (e.g. Kim, 2009) compared participants of unparalleled writing proficiencies or ages between native and non-native speakers while others (Kang, 2017; Mirzapour & Mahand, 2012) used too small a size of corpus to generalize. The number of texts they used remain very low as much as they could analyze them manually. There is a definite lack in computational approach that enables us to explore millions of words. Computational methods will require completely different settings than by human classification, but this attempt surely widens quantitative research opportunities to produce more statistically meaningful results.

Last but not least, more thorough investigation of Korean English learners should be contributive, too. There needs more robust research that examine specifically Korean English learners' corpus to analyze their patterns of using stance marking devices in their English writing. Firstly, most previous studies focusing Korean students mainly observe undergraduate students from well-named schools, but it does not reflect their English writing proficiencies. Even in the same college, the students' English fluency can vary and the writing skills could be distinctive from other perceptive skills like reading and listening. However, most English proficiency tests (KSAT, TOEIC, TEPS, and FLEX) evaluate reading and listening and this is usually how students are screened in Korea for college entrance, internships, and study abroad programs per se. It is, therefore,

problematic to refer to participants from the same college as a same advanced level group without extra proof of writing abilities (e.g. Kang, 2017). Therefore, it is worthy to divide the text data by its proficiency levels and study abroad experiences to see if there is any connection between non-native speakers' overall English fluency and exposure to authentic English and their employment of hedges and boosters. Secondly, previous corpus analysis that compares native speakers with Korean English learners was carried out manually. While this can be qualitatively more accurate, quantitative research with massive data is still necessary to gain statistically valid results. The purpose of computational analysis, therefore, is to populate statistics from a high volume sample size of Korean learners' writings. As this study is the first attempt to carry out computational analysis of the uses of hedges and boosters from extensive Korean students' data, there will be much meaningful discussion on methodologies and results. This will be of great value of itself for future research on this topic.

### **2.7 Needs for Teaching Hedging Strategies to Non-native Speakers**

Revisiting Lewin (2005)'s study, it is important to teach rhetorical skills for academic writing. In general, readers identified greater instances of hedges than the authors' intention. This suggests the need for formal instructions both to NSs and NNSs, if they are going to pursue advanced degrees. Especially, research into other cultures has claimed that non-native speakers are found to be lacking intuition on hedging if not

taught (Kobayashi, 2016; Lewin, 2005; Tan & Eng, 2014; Ventola, 1997). Therefore, it is deemed necessary to explicitly teach EAP students rhetorical strategies for better written communication. The importance of teaching will be more enlightened by the following studies.

Study 1 focuses on the comprehension of hedges and boosters among native and non-native speakers. Study 2 investigates the production of hedges and boosters by native and non-native speakers. The studies will focus on Korean English learners who have the same educational level, college degree, and have decent English reading and writing skills at college level. This will give us a general understanding of L1 and L2's attitudes towards the two rhetorical tools and how close they are to linguists' descriptions. In addition, this will provide Korean English practitioners with informative results to understand their students and apply them to developing their teaching practices.

### **Chapter 3. Study 1 – Perception of Hedges and Boosters by native speakers of English and Korean English learners**

#### **3.1 Methodologies**

##### **3.1.1 Participants**

The survey was conducted to American English native speakers and non-native speakers whose first language was Korean. Native speakers' survey was conducted

through Amazon Mechanical Turk (AMT). They were redirected to a study that was hosted on Qualtrics. Korean native speakers were invited to participate in the experiment through academic snowballing. They were also redirected to a study that was hosted on Qualtrics. 47 English native speakers and 55 Korean native speakers participated in total.

Education			Gender		Age				
High School	College	Grad School	Male	Female	20s	30s	40s	50s	60s
32%	51%	17%	57%	43%	36%	31%	21%	6%	6%

*Table 1 English NS participants' background*

Gender		Age	KSAT Score				Living Abroad	
Male	Female	17~26	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	NA	3m~48m	None
38%	62%	100%	45%	18%	22%	15%	25	30

*Table 2 English NNS participants' background*

For the native group, the survey was open to diverse backgrounds to see general native speakers' intuitions. 24 participants have graduated from college, 8 finished graduate school and 15 completed up to high school. 57 percent of participants identified themselves to be male and 43 percent as female. Their average age was 36, varying from their 20s to 60s. The majority of the participants were in their 20s, the second most fell into 30s, and the 40s came next. The least numbers were in their 50s and 60s, 3 and 2 respectively.

Korean English learners were limited to college students who have fluent English in academic reading. It is not only because the survey is conducted in English, but also because they can represent well of Korean English learners who have successfully passed through public English curricula for secondary education in South Korea. The college entrance exam known as KSAT (Korean Scholastic Aptitude Test) includes an English subject and it is to measure the minimum expectancy for academic English at college level. The total 55 Korean students participated and their ages varied from 17 to 26 (19 to 28 in Korean age). The gender ratios are male 38 percent and female 62 percent. Their grades of English subject in Korean SAT range from the top 1st to 3rd out of 9 curved grades. 25 participants hold the 1st grade, 10 the second, 12 the third, and 8 reported as not taken. Since KSAT English test is highly focused on reading, this reflects their relatively stronger reading skills than average Korean English learners. Many of the participants were gathered from high-profile colleges in Korea, including Busan University Medical School, Hankuk University of Foreign Studies, KAIST, Korea

University, Seoul National University, and Yonsei University and some from high school graduates planning on going to colleges abroad. In terms of their living abroad experience, 25 have lived at minimum 3 months and at maximum 4 years while 30 have never been to English speaking countries.

### **3.1.2 Material**

#### **A. Stimuli**

The stimuli are based on the Farrokhi and Emami's (2008) inventory of hedges and boosters. The inventory was compiled from previous literature that were influential in this topic (Holmes, 1988; Hyland, 1996; 1998; Hyland and Milton, 1997; Quirk et al., 1985; Varttala, 2001). However, not every item was appropriate for my survey, for some were difficult to find their neutral counterparts or more frequently used as content words than rhetorical devices. Also, some had to be reclassified according to other literature or linguists' intuitions. A few items appeared in the test for three times considering the possibility of having different functions depending on their locations. More details to be discussed below.

### **(1) Syntactic limitation to find neutral counterparts**

It is often difficult to make a contrasting sentence with the item which is neutral in meaning. For example, *hypothetical* is an adjective, which could be either attributive or predicative. It modifies the noun that follows or describes its subject. See the following examples.

a. The proposition that men are healthier than women is hypothetical.

(Predicative)

b. This is a hypothetical proposition that men are healthier than women.

(Attributive)

In the first sentence, the word *hypothetical* was used as a predicate, not as a rhetorical device. In contrast, the second sentence could be considered to function as a hedge. However, it is hard to find a neutral sentence that contrasts the second sentence rhetorically, without changing the meaning. There is no word that can replace *hypothetical* that keeps the same interpretation, yet not diminishing certainty of the proposition. Completely omitting the word would not be helpful as well, since it breaks the grammaticality of a sentence. One of the good ways to avoid this problem is to make an adjective in a main clause form as 'It is (Adjective) that...'. However, the word *hypothetical* is not found to be used in this form according to Corpus of Contemporary American English (COCA).

Some auxiliary verbs like *must* and *should* also encounter similar limitations. The sentence 'I must go home' cannot be compared with 'I go home' as it totally takes away the meaning of obligation. Employing other auxiliaries, such as *may* or *should*, can be also problematic. *May* could drastically change the tone of the sentence, not keeping it neutral. The verb *should* holds the claim stronger than the neutral sentence. Moreover, *should* could be too subtle to discern from *must* in its boosting function as Van der Hoek and Meyer (1997) state that both *should* and *must* could make the writer's claim strong. *Must* and *should* could be even complementary to one another as Leech (2003) claims that the uses of *must* and *should* have reversely fluctuated over time, according to corpora of American and British English. Therefore, such items as *hypothetical* and *must*, which do not always have ideal counterparts had to be removed from the survey questions.

## **(2) More than a rhetorical device**

There are many words in the Farrokhi and Emami's (2008) list that are used as content words rather than function words more dominantly. Otherwise, they could be very subtle in their meanings, making it hard to grasp the author's motivation for employing certain words: whether they were literally put or rhetorically put. Therefore, those items which are more often used as content words are eliminated from the list. Those items are *find*, *interpret*, *establish*, and *necessarily*, just to name a few. I checked with these items in COCA one by one and found out that they appeared as boosters only

in a few sentences. See the example with the verb *establish*. Compared with (d), sentence (c) with the word choice of *establish* is more strengthened.

c. Early cases, however, *establish* two basic principles.

d. Early cases, however, *suggest* two basic principles.

### **(3) The effect of the location**

There are a couple of items that appear in the survey three times, instead of once. They are words that clearly express one's thoughts: *believe* and *think*. In general, *think* and *believe* are considered to express personal doubt (e.g. Afshar et al., 2013). However, some Korean speaking scholars, including Kang (2017) and Kim (2009), claim that *think* and *believe* can express commitment and firm will of the author when it comes in a topic sentence. Kang claims that Korean students use a great amount of *I think* phrases in the head position of sentences, but far less or none in other positions unlike native speakers. Not an instance of the rear location of *I think* was found in L2 students' writings. Kim even suggests that many L2 writers would not understand and make use of the hedging function of *I think* or *I believe*. Considering these claims, it would be meaningful to see how native speakers and Korean speakers perceive and produce the two items in different positions. When it comes to corpus analysis, however, including these items could be precarious since they are very context dependent. Therefore, the two words with different

locations will only be cited on the list for the survey, not for the corpus analysis.

However, we will look at the concordances to examine their productions of those opinion phrases.

#### **(4) Re-classification based on other literature**

There are a few items that are categorized differently from other major studies. For example, *greatly*, *prove*, and *highly* are listed under a hedge category in Farrokhi and Emami (2008) but many other scholars consider them as amplifiers (Quirk et al., 1985; Bolinger, 1972). The verb *prove* is especially known in academic writing as to give excessive confidence to the author's statement, thereby blocking other opinions from coming in. It is even described to be a very strong verb that parallels with *warn*, *threaten*, and *promise* by the University of Adelaide academic writing guide. Therefore, for some items like those, I will relocate them according to more dominant descriptions.

#### **(5) Re-classification based on native linguists' intuitions**

A more complicated instance would be the modal verb *would*. Farrokhi and Emami (2008) assigned *would* in both lists of hedges and boosters, yet others have had different views towards the function of *would*. Whereas it was regarded as extreme likelihood by Palmer (1979), it was deemed to connote tentativeness and politeness by

Coates (1983). However, three candidates of M.A. Linguistics manually assessed 30 samples of concordance output and found almost all as hedges. Therefore, we decided to categorize *would* as a hedge first, and then verify it by carrying out the two experimental studies to follow: a perception test and an accuracy test for the corpus study. This will be discussed in design and process.

## **B. Questionnaires**

The survey consists of 85 questions of the same patterns: 42 on hedges and 43 on boosters. Each question has two statements given: one neutral statement and one with a target rhetorical device. See the two example pairs below.

< A pair with and without a hedge item >

(1) Target: Global warming is serious, *I believe*.

(2) Neutral: Global warming is serious.

<A pair with and without a booster item >

(1) Neutral: New York is a crowded city.

(2) Target: New York is *the most* crowded city.

Participants are asked to read each pair of statements as above and to evaluate the author's confidence level in each statement. It is a multiple-choice type to choose from four options that describe one sentence to be 'more confident' or 'somewhat more confident' than the other. Below is a sample question. The question statement and the four options apply to every question.



**SAN FRANCISCO  
STATE UNIVERSITY**

Please read and evaluate the following two statements.

(1) Computers have made life easier, I think.  
(2) Computers have made life easier.

In which sentence is the speaker more confident?

- In sentence (1) the speaker is more confident.
- In sentence (1) the speaker is somewhat more confident.
- In sentence (2) the speaker is somewhat more confident.
- In sentence (2) the speaker is more confident.

*Figure 1 Perception Test Question Sample*

The word *confident* in the questions was chosen based on the definitions of hedging by many scholars. According to Coates (1987), Holmes (1984), and Hyland (1998), a booster and a hedge are devices that show confidence or the lack of confidence, indicating that the statement is an opinion rather than a fact. Hedges are to limit the confidence invested in the claims while boosters are to express strong conviction and

certainty and to assert a proposition with confidence. The word *confidence* reflects well of the author's attitude, certainty, tentativeness, and responsibility towards his or her proposition.

Meanwhile, there are a few items that clearly signify that the following statement would be the writer's opinion, and they will appear in questions three times each, in different positions. As mentioned earlier, these seem to be used by some non-native speakers to rather emphasize their opinions than it is commonly known to soften down the voice by increasing subjectivity. Therefore, the two phrases, *I think* and *I believe*, will be asked for its certainty three times and their locations will differ each time: the beginning, the middle, and the end.

### **3.1.3 Design and Procedure**

The survey provides an introduction at the initial stage. The survey is voluntary, so it is only proceeded under their agreement. The survey takes on average 20 minutes for non-native speakers and 15 minutes for native speakers, but it may vary depending on the individuals. The order of stimuli and question sentences are completely randomized. By Excel program's randomizing tool, stimuli were randomly assigned into two groups: Group 1 where a target sentence comes before a neutral sentence and Group 2 where statements come in the reverse order. Half of the stimuli for each batch have a neutral sentence first, and the other half have the sentence with a hedge / booster first. After

recording the questions in Qualtrics in such an order, the order of questions is also set to be randomized for every trial. As a result, every participant gets to see questions in different orders. Moreover, participants cannot see the questions ahead or go back to earlier questions to modify. They can only answer one item at a time. After the questions, they are asked of several demographic questions: age, gender, education, and first language. Non-native speakers get extra questions, such as Korean SAT score on English, experience of living in the English speaking countries, and the length of stay, if any. The survey was open for two months.

### **3.2 Results and Discussion**

The raw data that was extracted from the survey results were re-coded into constant values. Considering the order of the statements (neutral sentence + target sentence or vice versa) and the target words being hedges or boosters, the answer codes were re-coded in order for all the answers to have the same values as '4' the most confident and '1' the least confident.



Please read and evaluate the following two statements.

- (1) Computers have made life easier, I think.
- (2) Computers have made life easier.

In which sentence is the speaker more confident?

- In sentence (1) the speaker is more confident.
- In sentence (1) the speaker is somewhat more confident.
- In sentence (2) the speaker is somewhat more confident.
- In sentence (2) the speaker is more confident.

Figure 2 Recoded values for target-neutral ordered questions



Please read and evaluate the following two statements.

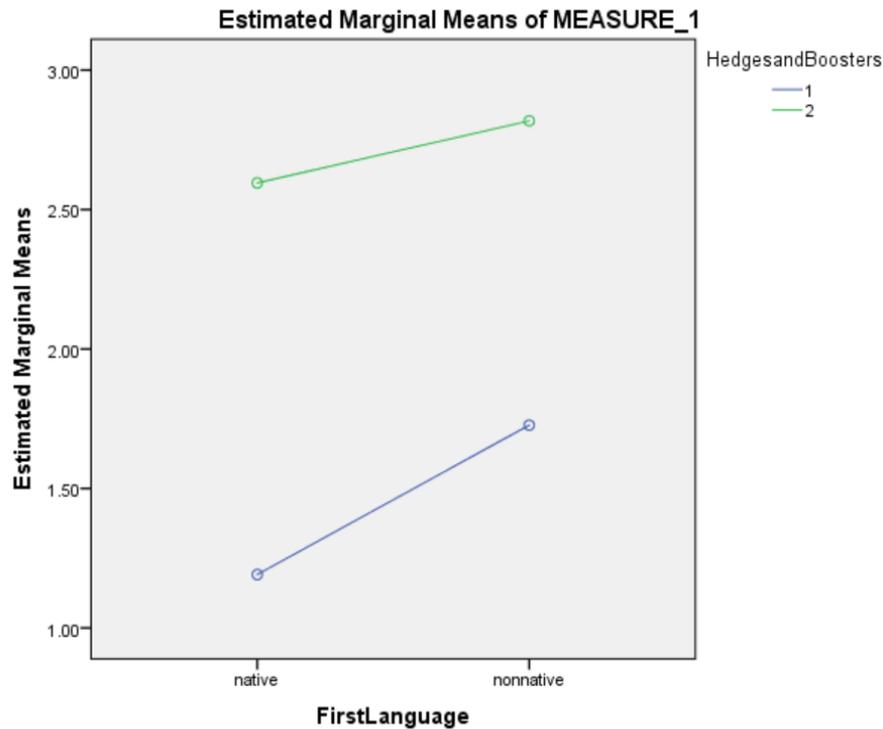
- (1) John will come over for dinner.
- (2) Perhaps John will come over for dinner.

In which sentence is the speaker more confident?

- In sentence (1) the speaker is more confident. → 1
- In sentence (1) the speaker is somewhat more confident. → 2
- In sentence (2) the speaker is somewhat more confident. → 3
- In sentence (2) the speaker is more confident. → 4

Figure 3 Recoded values for neutral-target ordered questions

The re-coded values were analyzed with Mixed Anova (2x2 design) and their results turned out to be significant. The items of hedges and boosters are confirmed to be very different in nature as can be seen from their high F value as  $F(1,100) = 269.21$  and low significance value as  $p < 0.001$ . As of native and non-native speakers, they also display a great difference, resulting in  $p < 0.001$  and  $F = 3342.95$ . Therefore, both items within subjects and between subjects are proved to be highly significant. However, there is found to be no interaction between native and non-native speakers in their usage patterns of hedges and boosters. To put it another way, they have similar patterns in their use of hedges and boosters. Their F score is as low as 2.005 and its significance value is  $p = 0.16$ . The graph below shows how closely the patterns of L1 and L2 move together. (1) The lower line is hedges and (2) the higher line displays boosters.



*Figure 4 L1 and L2 perception scores for hedges and boosters*

Both L1 and L2 marked less confidence towards hedges than boosters. NSs' average score on hedges is 1.318 while that on boosters is 2.591. NNSs show a similar discrepancy, 1.777 on hedges and 2.853 on boosters. Their perceptions correspond to what hedges and boosters are defined to function as rhetorical devices: Hedges making an argument less certain and boosters making it more certain.

However, those two language groups of participants display slight differences as well. Native speakers of English gave weaker confidence level to hedges than Korean participants, making 0.459 difference. Most L1 respondents have given the minimum score 1 to almost all hedges, given that their score is 1.318 which is only slightly above

the least number they can give. It gives us the assumption that native speakers have no argument over hedges as to be softening down their claims. Only exception we could find is *quite*. While non-native speakers remained below 2.5 on this item, native speakers locate it slightly above the neutral line. As Merriam-Webster (2002) defines it as *to an extreme* with positive connotation, *quite* can be sometimes considered as an intensifier than a softener. However, this same word holds the opposite nuance of hedging as in ‘not *quite* finished’. In short, *quite* can be seen to be sensitive to its context and it functions as either a hedge or a booster depending on its surrounding environment.

The Korean groups also perceived overall hedges to be weakening an author’s confidence and it was shown by their scores remaining below 2. Meanwhile, there are two items - *apparent* and *apparently* - which hit 2.5 baseline. While they connote *seeming* or *seemingly* often times, non-native speakers’ evaluation seems to reflect their recognition of the items more similar to *obvious* or *obviously*. This may be a consequence from Korean learners’ common habits of memorizing English vocabulary with its first definition given in a vocabulary book or a dictionary. Usually, the first matched Korean words on dictionaries for these two word families are bunmyeonghan / bunmyeonghagae (분명한 / 분명하게) which are literal senses of manifestness or obviousness. This points out the problem in the way of studying / teaching English vocabulary in which prevails the neglect of pragmatic meanings of English words.

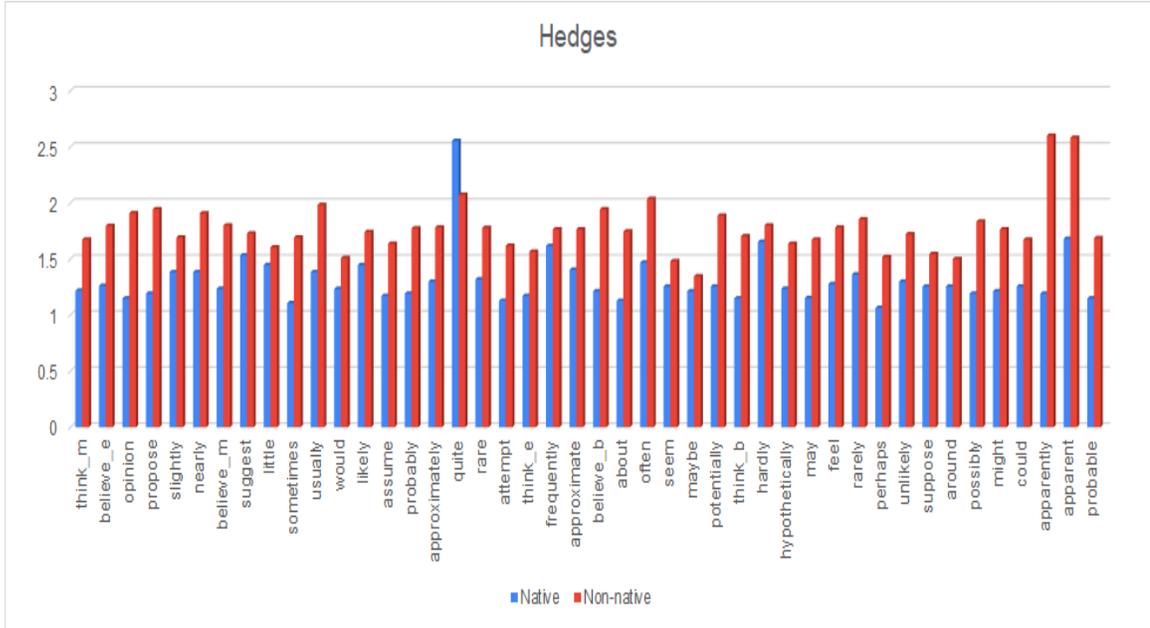


Figure 5 L1 and L2 perception scores for hedges

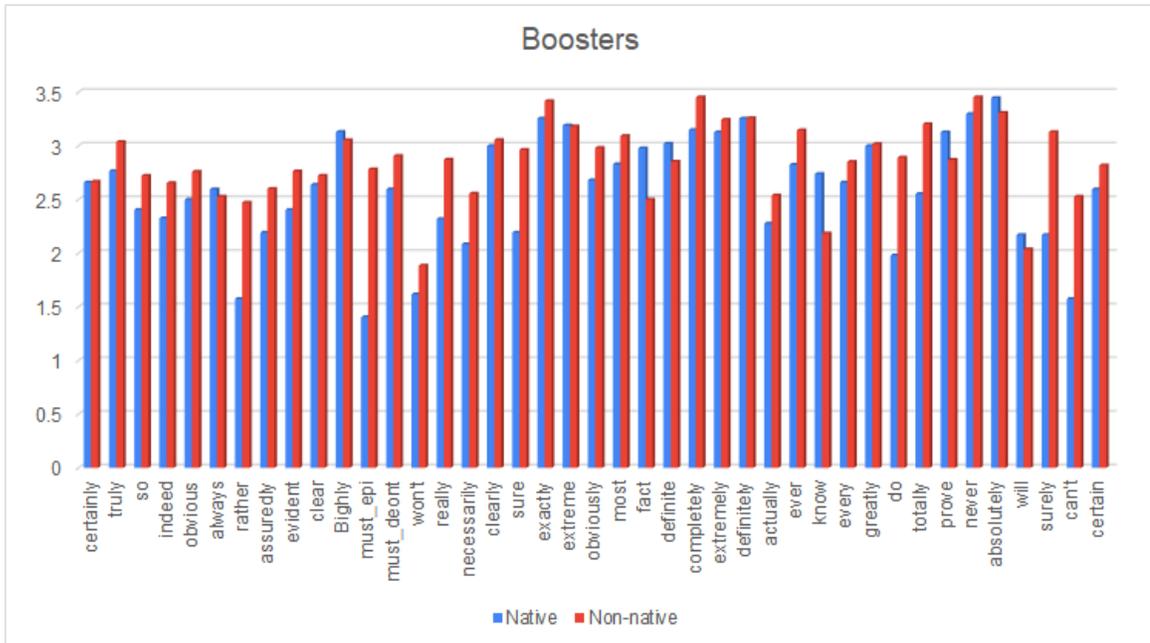


Figure 6 L1 and L2 perception scores for boosters

For boosters, in contrast, Korean students perceived them as expressing greater certainty compared to native speakers. It could be explained by two possible reasons. First, it could be the case that the native speaker participants are equipped with so many different colloquial meanings of booster items in their background knowledge, so they might have affected their comprehension of a statement which is relatively more open to many interpretations than a sentence in a longer passage with more contexts given. Second, it also invites us to think about the importance of teaching hedges and boosters in written context to native speakers as well as non-native speakers. Despite native intuition, academic writing or argumentative writing sometimes requires teaching and training since academicians share some rules and format in their writings for effective communication. The fact that L2 speakers outscored the L1 speakers' evaluation of the boosters' strength imply that they are better trained in the uses of such intensifiers in written context because they first encountered those items in the academic setting. Also, considering that the Korean participants have top grades on KSAT, they must have well trained in academic reading, not much in other skills. Consequently, they usually lack pragmatics and cultural backgrounds compared to native speakers and, thus, answering the survey could have been less complicating to them, but rather very straightforward.

Now we will see some booster items that received particularly lower evaluation than expected. Above everything else, modal verbs are evaluated weaker than 2 by native speakers. They are *will*, *won't*, *can't*, *do*, and epistemic *must*. They all are in the boosters' list, but they are among the items that gained the weakest scores from the L1s.

Many of their choices can be explained by their epistemic properties. Epistemic uses of modality are to express possibility, allowing a chance for a statement not to be true. Although every modal verb expresses different degrees of certainty, even the epistemic modal verb with the strongest certainty among others has a chance for its proposition to turn out to be false. Despite the degree of certainty, as a result, given those items compared with neutral statements without any epistemic modal verbs, evaluators may well assess the statements with the auxiliary verbs to be weaker than the neutral ones. If these booster items, *will*, *won't*, *can't*, *do*, and epistemic *must*, were compared with relatively less certain modal verbs such as *might* or *could*, readers will definitely find the booster modal verbs to be stronger. However, because modality is paired with a neutral sentence in this study, they will still be considered to be less confident, as they contain the meaning of probability. Considering the design of the test, native speakers' choices can be rather taken as reasonable judgements. If they were told to evaluate those items alone, the participants might have agreed that they sounded very strong for academic purpose.

On the other hand, it is worth referring to the emphatic use of *do* which gained similar results. One possible explanation is that emphatic *do* might have added some nuances of subjectivity. Although the purpose for the use of *do* in writing is clearly to give strength to the writer's argument (Biber, 2010), when it comes to evaluating its statement with the neutral counterpart, there is always a chance for the reader to bring in their subjective or pragmatic interpretation that its overwhelmingly stronger claim could

rather clinch its opinion than delivering a fact. This assumption may apply to other examples such as *really* and *surely*, which literally show great emphasis, but barely passed the borderline of boosters in NSs' responses.

Overall, the native speakers' perceptions of boosters provide three important suggestions. First is that even native speakers of English might need to be taught English rhetoric for academic writing purpose. It seems very crucial to instruct them the functions of boosters in academic writing and to guide them that some items are more preferred over others in academic settings as they could prevent from overstating the assertion. Secondly, we cannot help acknowledging that every reader could have a different interpretation. We may want to re-invite Lewin's (2005) study at this point, where there was a big difference shown in the interpretations of rhetorical devices between readers and writers, both of whom had advanced degrees. Our survey data of native speakers in which they perceived less confidence from several boosters seem to be supporting the idea that some words sporadically direct readers to an opposite direction from the author's original intention and boosters tend to do this job more seriously, resulting in alienating readers from writers. Finally, this finding triggers us to think that maybe these rhetorical devices are much fuzzier than we thought, so they cannot be one or the other. Therefore, it is very difficult to make a list of hedges and boosters that is free from contexts or individual perspectives.

Back to modality, non-native speakers are found to have unstable responses. It has been said that many English textbooks used at schools in South Korea failed to either

maintain or clarify the distinction between epistemic and deontic modality, thereby confusing possibility with necessity (Hyland, 1998). This was well proved by the Koreans' answers on *must*. Unlike native speakers, they display similarly strong perceptions of certainty between epistemic and deontic uses of *must*. Since the difference of these two functions are definite, this implies the shortage of proper understanding of modal verbs among NNSs. When modality is so important in writing, this seems to evoke the importance of teaching modality in academic writings.

Other factors did not have much effect on the results. In terms of opinion marking phrases *I think* and *I believe*, there was no significant difference by their locations: beginning, middle, and end. Neither Korean English learners nor Americans were affected by the positions of the phrases. The results did not have much association with the participants' backgrounds either. Gender, KSAT score, and living abroad experience all did not have any significance on the recognitions of stance markers. The significance value between the responders' genders and the rhetorical items is 0.681 and between their KSAT scores and the items is 0.324. However, the  $p$  value for hedges and boosters and living in the English-speaking countries is worthy to take a closer look. Its value is  $p = 0.068$ , still higher than 0.05, which is considered significant, but it is only marginally above it. It appears to stay very close to the 0.05 level. The graph below also suggests that living in the English-speaking countries does have influence on their perceptions of hedges and boosters. Students with experience of studying abroad gave weaker scores to hedges and higher scores to boosters than those without. The means of those participants'

answers on boosters is 2.94 and that on hedges is 1.66. This is the most adjacent pattern to what linguists describe about hedges and boosters. This may imply that studying abroad experiences positively reinforce NNSs' comprehension knowledge on the rhetorical devices that may have been taught in classrooms.

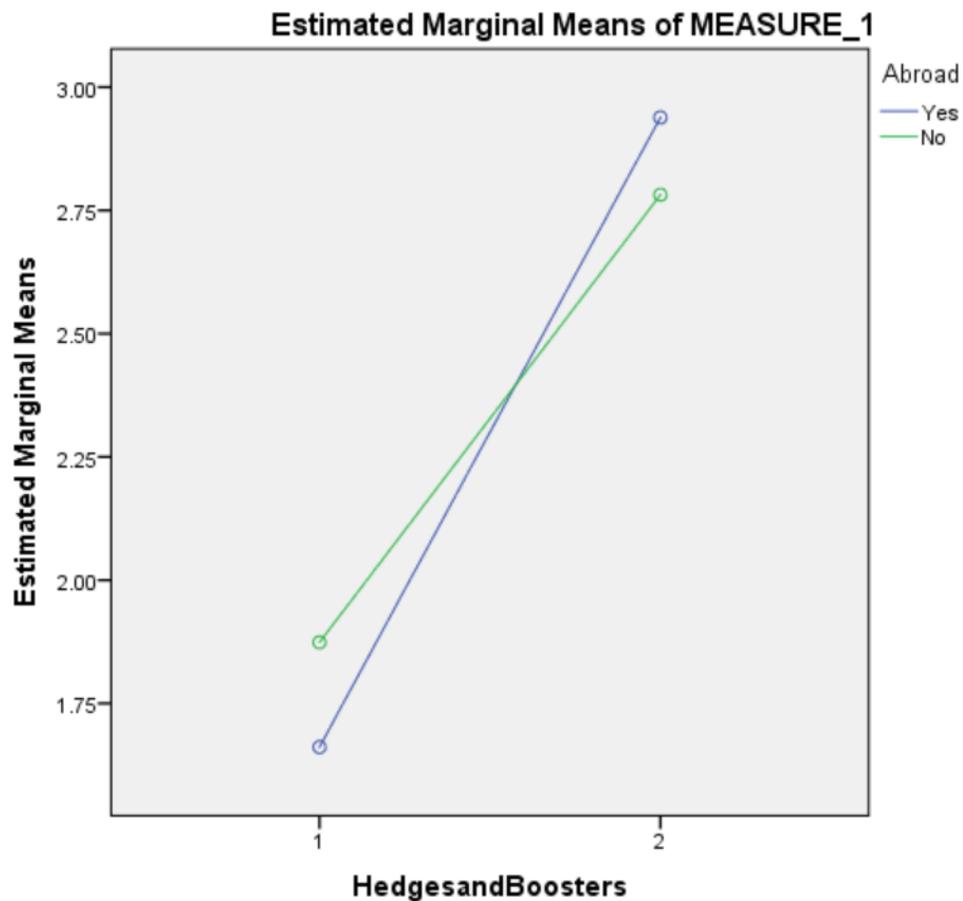


Figure 7 Perception scores for hedges and boosters between the L2 groups with and without living abroad experiences

### **3.3 Limitations and Further Discussion**

We have seen that booster modal verbs were marked lower than the boosters' average confidence level. As mentioned above, one big reason could be its relativity with neutral counterparts. Normally in research discourse community modals of necessity and obligation are considered very strong for academic writing. Attam (2014) and Farrokhi and Emami (2008) found in their studies that participants used more epistemic than deontic modality in academic papers. However, in the experimental study 1 above, native speakers of English felt it much less confident to use epistemic modal verbs than simple statement without them. Deontic use of *must* was clearly separated as a booster by native speakers, but the same *must* in epistemic use, which can still deliver relatively strong certainty to readers, was marked much lower. Because of the structure of the survey questions, where they were asked to compare the sentences with target items with the neutral statements, we cannot stop the participants from perceiving *must*, *can't*, or *won't* to be less certain than the simple factual statements. For example, the following two statements from the question,

a. Mr. Jones must have been sick.

b. Mr. Jones has been sick.

it is very likely that readers will choose (b) over (a) to be more certain, because (b) sounds more factual than with *must* which marks one's prediction. If (a) was compared with some other statements such as:

c. Mr. Jones could have been sick.

d. Mr. Jones may have been sick.

(a) will be chosen over them unanimously. Likewise, if the statement were to be evaluated by itself, (a) could have received higher scores than the current values.

Otherwise, it seems reasonable to assume that participants could have perceived it as more confident if the statements were more academic or argumentative, so the participants had consensus in context reading.

e. Based on the results of this study, I suggest that lockdown must be the only way to stop pandemics.

f. Based on the results of this study, I suggest that lockdown is the only way to stop pandemics.

While (f) can sound more factual even between this pair, the result may have been more controversial. Meaning, with more backgrounds of the statements given to the participants, they might have been able to perceive (e) just as strong as (f) for academic or argumentative writings.

Not only epistemic modality, but several other boosters also barely passed the minimum score to be classified as boosters. Overall, it indicates the need for qualitative studies to support the results of the survey. Follow-up interviews might have helped clarify and understand the participants' answers more accurately. More interactions

during or after the survey could have also improved the survey environment by compensating the lack of contextual information.

Finally, we focused on the comprehension of rhetorical devices in the study above. We have seen that American/British native speakers and Korean English learners display similar patterns in their perceptions of hedges and boosters. Both speakers' comprehension tend to correspond to the linguists' descriptions of hedges and boosters. Nevertheless, this finding does not explain their productions of stance markers. Many scholars in the second language acquisition fields claim that comprehension and production are different matters. Despite the different propositions, one premise remarked by almost all SLA theories nowadays is that it requires extra steps to arrive at production of language (Swain, 2005). In this sense, it is now worthwhile to turn our attention to see L1 and L2's productions of hedges and boosters. Therefore, in the next chapter, I will discuss the experimental study 2 which was designed to explore L1 and L2's use patterns of hedges and boosters in academic written contexts. Native and non-native college students' opinion essays will be analyzed to compare the frequencies of hedges and boosters and will discuss any distinctive findings in relation to their survey results. The next study will be a good addition to earlier research as well because many studies have shown disagreement in their analysis.

## **Chapter 4. Study 2 - Corpus analysis of argumentative writing of native speakers of English and Korean English learners**

A corpus analysis of Korean learners' and native speakers' usage patterns of hedges and boosters in their English argumentative writing was conducted computationally to see the distributions of hedges and boosters in L1 and L2 written data. This study will run massive corpus data with codes to produce the general patterns of productions by NSs and NNSs. Additionally, this study will examine how much the results agree with the previous literature findings and whether or not they reflect the perception test results found from Study 1.

### **4.1 Methodologies**

#### **4.1.1 Material**

##### **A. Stimuli**

Stimuli was also selected from Farrokhi and Emami's (2008) list of hedges and boosters, yet with the items occurring more as content words removed (see earlier discussion on pp.22-24). It consists of 41 hedges and 50 boosters in total. Although there have been attempts to identify hedges and boosters, setting up a list of hedges and boosters has often been found difficult, because they could be used in different meanings. I also encountered several difficulties in trimming the entry as follows.

### (1) Lexical Ambiguity

There are some words on the list which hold lexical ambiguity and interrupt writing codes for corpus analysis. For example, a modal verb, *can*, carries three different meanings: ability, permission, and speculation (Biber, 2010). Only when it expresses speculation, can it function as a hedge. However, these meanings cannot be discerned by their surrounding words or grammar. The sentence (a) below, for example, can be interpreted as ability (Tom is able to be at home by himself), permission (Maybe Tom's mom allows him to be home alone), and speculation (It is possible that Tom is at home by himself).

- a. Tom *can* be at home by himself.

Similarly, Stevenson (2010) says that word *certain* can have a few different meanings. It could address the author's confidence or signify a specific amount and an unspecified person or a thing. Only the first definition seems to operate as a booster. This kind of lexical ambiguity makes the computational tasks trickier.

Another example is *complete*. An adjective *complete* can also become an emphaser in some cases like (b), but in other cases, it means 'full' or 'entire' as in (c). Compare the two below. While in (b) *complete* is an optional word that can be omitted without changing the message, the same word in (c) cannot be so, as this adds the meaning of entirety.

b. She is a *complete* stranger.

c. I've collected the *complete* set.

## (2) Syntactic Ambiguity

There are some adjectives that function only as a content word without functioning as a booster or a hedge in certain structures while the same word can be used as a rhetorical device in other positions. For instance, *inevitable* emphasizes the noun preceding it as in (d) while the same word is more descriptive of fact in (e).

d. Air pollution is an *inevitable* crisis to human beings.

e. Traffic jam is *inevitable* during rush hour.

When this adjective is attributive to describe its following noun, this seems to function as a booster to amplify the factuality of the noun. Therefore, in (d) the crisis is not a simple crisis, but an “inevitably” serious one. On the other hand, when *inevitable* is a complement as in (e), it only describes traffic jam, but does not make a claim stronger.

For those items that apply to lexical or syntactic ambiguity, I found it very limited to include them in the computational analysis as they would be run through a code.

Therefore, I will remove those items that are ambiguous from the list and only operate

those that are relatively constant in their function and meaning or able to be sorted out by extra rules through codes.

### **(3) Functioning both as a hedge and booster**

Some, but fortunately few, can imply both stronger and weaker claims depending on the context. One instance is an argumentative phrase *I think*. The verb *think* or *believe* together with the first person pronoun *I* is very often discussed to show doubt rather than certainty and only some degree of assurance attributed to the claim (Algi, 2012).

However, some scholars, including Kang (2017), claim that *I think* construction is used as a booster more dominantly by non-native speakers. Since it is more common to consider *think* and *believe* as hedging devices (Farrokhi & Emami, 2008), I will firstly categorize these two items as hedges in the code dictionary, but will leave them for discussion later.

Likewise, a modal verb *would* will be assigned as a hedge, not a booster.

Although there were some perspectives that *would* can be a booster to express extreme likelihood (Farrokhi & Emami, 2008; Hyland, 1998 ;Palmer, 1979), as seen from Study 1, the majority of linguists and the survey participants perceived it as a hedge in more occurrences. Therefore, it will also be categorized as a hedge in Study 2 and we will discuss it by its concordance output.

Considering the three kinds of ambiguities referred to above - (1) Lexical ambiguity, (2) Syntactic ambiguity, (3) Functioning both hedge and booster, I will first adopt Farrokhi and Emami's (2008) list which is based on well noted studies (Holmes, 1988; Hyland, 1996; 1998; Quirk et al., 1985; Varttala, 2001) as the baseline. Then, some of the items, such as *complete* and *inevitable*, which are more frequently used as content words will be removed from the list. A revised compiled list can be found in the Appendix 2.

## **B. Corpus**

The L1 corpus is The Louvain Corpus of Native English Essays (LOCNESS) and the L2 corpus is Yonsei English Learner Corpus (YELC). YELC was collected from a freshmen placement test for incoming students' English language class at Yonsei University. As one of the named universities with a long history in Korea, Yonsei University students are normally expected to have abilities to read and write academic English. LOCNESS, on the other hand, is not based on one particular school, but includes undergraduate level British and American students' expository and argumentative writing. The YELC corpus also has two types of writing – free and argumentative – and it has a list of students with their gender and proficiency levels identified based on the CEFR (Common European Framework of Reference) standard. The size of the cleaned corpus is as below. The texts that are to be compared are argumentative writings.

British Argumentative	American Argumentative	British Literature/Expository	American Literature / Expository
70,902	131,844	67,601	16,302
Total number of texts: 209		Total number of texts: 113	
Total number of words: <b>202,746</b>		Total number of words: <b>83,903</b>	
Total Native Speakers' text: <b>286,649 words</b>			

*Table 3 LOCNESS - Native Speakers' writing corpus*

Free	ARG(A1)	ARG(A2)	ARG(B1)	ARG(B2)
<b>340,671 words</b> <b>3,286 texts</b>	25,700	144,165	295,888	140,164
Argumentative: <b>823,531 words / 3,286 texts</b>				
Total YELC Corpus: <b>1,079,849 words / 6,572 texts</b>				

*Table 4 YELC - Korean learners' writing corpus*

#### 4.1.2 Design and Process

Both written corpora of native speakers and non-native speakers were analyzed to compare their frequency patterns of hedges and boosters. The hedges and boosters are from the revised inventory of Farrokhi and Emami (2008) as mentioned above. Both corpora were used as bags of words in the Python coding environment. YELC was divided by students' levels to see if the levels of their English have any influence on their uses of rhetorical devices. While the levels were originally divided into five groups in

YELC ranging from A1 (the lowest) to C2 (the highest), since the size of C1 and C2 are so small, I joined them to B2 in order to make its results more valid.

I used Python2 through Jupyter Notebook to process the data. The corpus was cleaned, the list of hedges and boosters was formed into a dictionary, and then the list of items were searched in both corpora. Their distributions in respective corpus was mainly explored and we looked into the most common items and the most common part of speech from each corpus.

## **4.2 Findings & Discussion**

### **4.2.1 Corpus Analysis Results**

To assess the significance of the results, the proportions have been compared. As recommended by Campbell (2007) and Richardson (2011), Chi-squared test was conducted. Firstly, both language groups' proportions of hedges were tested, and their Chi-square value was  $X^2(1, N=1,026,277) = 190.145, p < 0.0001$ . The proportions of boosters from both groups were assessed next, resulting in  $X^2(1, N=1,026,277) = 101.990, p < 0.0001$ . It means that there is a significant difference in how native and non-native speakers use both hedges and boosters. There was no variation by non-native speakers' proficiency levels. From the least proficient group (A1) to the most proficient

group (B2), all levels remained significantly different from native speakers' proportions of hedges and boosters.

	Native		A1		A2		B1		B2		Total	
	Token	1,000	Token	1,000	Token	1,000	Token	1,000	Token	1,000	Token	1,000
Hedge	5556	27	789	30	4716	33	9911	34	4144	33	19560	33
Booster	4738	23	791	30	4135	29	7941	27	3337	26	16204	27
	B	H	B	H	B	H	B	H	B	H	B	H
Type	42	39	23	24	38	29	41	35	37	31	42	35

*Table 5 Frequency of epistemic markers in L1 and L2 writings*

As can be seen from Table 3 above, NNSs use greater number of rhetorical devices for their stance marking than NSs. Non-native speakers show the tendency to employ many hedges and boosters in their argumentative writing, regardless of their English fluency. It suggests that they overuse stance markers for stating opinions than native speakers would. Excessive or redundant uses of hedging strategies, however, could cause readers to frown. This perspective is most clearly indicated in the style guide literature where hedging devices have been described as “unnecessary words” (Yarber, 1985: p. 188) or “padded expressions” (Muller, 1985: p. 328). This warns us to beware

that rhetorical devices could be used positively if used well, but could be a poison to one's writing if used recklessly.

Their uses of stance markers, however, seem not only excessive, but also very repetitive. It can be seen by the number of types. Non-native speakers show a discrepancy in its number of types compared to native speakers. L1 speakers used 39 types of hedges and 42 types of boosters out of 41 hedges and 50 boosters on the list. On the other hand, L2 speakers used 35 types of hedges and 42 types of boosters from the same list.

Although their numbers of boosters tied, Korean English learners adopted fewer types of hedges than native speakers. L2 speakers used both devices with greater frequency but the number of types of devices they used were only equal or fewer than native speakers. This demonstrates that English learners as a second language use the same markers more repeatedly and more frequently than needed. Greater frequencies yet fewer types, overall, imply that non-native speakers lack in vocabulary. This echoes the Kang's (2017) finding that non-native speakers' lack of vocabulary causes them to use the limited types of devices repeatedly, which makes their writings less dynamic and less effective. As vocabulary diversity has always been one of the key constructs to measure the text quality (Olinghouse & Wilson, 2013), the number of types must give a significant effect on the impression of the text.

Kang's (2017) suggestion above that non-native speakers' fewer types is because of their lack of vocabulary, which is verified by a correlation between NNSs' English proficiency and the number of types they used. L2 speakers' English proficiency levels

had a great influence on their number of types. The lowest level group (A1) displays conspicuously lower numbers of kinds, which is almost a half of native speakers' variety. In contrast, the other higher groups show compatible numbers of types with native speakers. The number of types for hedges and boosters increase dramatically as the level advances until B2 shows a slight decrease. The reason B2 has fewer types than B1, despite the former group being more advanced, could be its small size of corpus. The number of words of B1 is 295,888 and that of B2 is 140,164, almost a half-size.

Looking back at the result from Study 1, non-native speakers perceived hedges and boosters identically with native speakers, but in Study 2 we see that they use them distinctively from native speakers. It suggests that they comprehend the notions and the functions of hedge and booster lexical items, but they employ them inefficiently compared to native speakers. As some (Hyland, 1998; Lee, 2006) pointed out, L2 speakers' heavy dependence on limited range of rhetorical devices could be caused by the absence of English instruction focused on productive skills, lack of vocabulary, and cultural differences. This is also an evidence that comprehension does not necessarily lead to production (Swain, 2005). Therefore, the need for teaching productive skills and guiding how to effectively apply hedging and boosting strategies in their academic writing are indispensable (Hyland, 1996).

Meanwhile, Korean English learners have displayed coinciding pattern in the ratios of hedges and boosters with native speakers. Both groups employed more hedges than boosters in their argumentative writings. Native speakers used 27 hedges and 23

boosters per 1,000 words; and non-native speakers used 33 hedges and 27 boosters per 1,000 words. While non-native speakers' usage patterns of the devices, in general, agreed with that of native speakers, the numbers varied according to writers' proficiency levels. The more proficient their English was, the closer their patterns were to the native speakers'. The least proficient students (A1) used the same number of hedges and boosters per 1,000 words, both being very high. As their level advances, the gap between hedges and boosters becomes greater and the ratios become more like the native speakers'. Although most Yonsei University students are expected to have relatively advanced or intermediate high levels of English skills, the most advanced writing proficiency group (B2) among these freshmen shows the most similar distributions of rhetorical devices to native English writers'.

Now that we have a general idea of NSs' and NNSs' distributions of hedges and boosters, we will next look at more detailed patterns. First, many of the most common items overlap between native and non-native speakers, yet there are a few that have quite a difference in their counts. For example, a modal verb *would*, which is the most frequent hedge among native speakers, ranked only sixth in Korean learners' texts. A regular verb *think* came the second from the top among non-native speakers, yet it appeared only third from the last among native speakers. *Believe* and *little* which made into the top ten list for NS did not even come in top ten for NNS and *might* and *opinion* filled those slots instead.

In terms of boosters, NSs tend to use more *fact* and *never* than NNSs; on the other hand, the latter group favors *really* and *always* over those two.

	L1 Hedge	Count	L2 Hedge	Count	L1 Booster	Count	L2 Booster	Count
1	Would	1,249	Can	5633	Will	923	Should	4,124
2	Can	794	Think	4,228	Should	653	Do	2,218
3	Could	512	Some	1,982	Do	492	So	2202
4	Some	456	About	1975	So	452	Will	2138
5	About	427	Opinion	989	Most	381	Must	1681
6	May	397	Would	804	Must	249	Know	875
7	Feel	216	Could	730	Fact	166	Most	645
8	Think	196	May	589	Every	160	Really	370
9	Believe	160	Feel	505	Know	138	Every	328
10	Little	121	Might	309	Never	125	Always	326

Table 6 Ten most common hedges and boosters

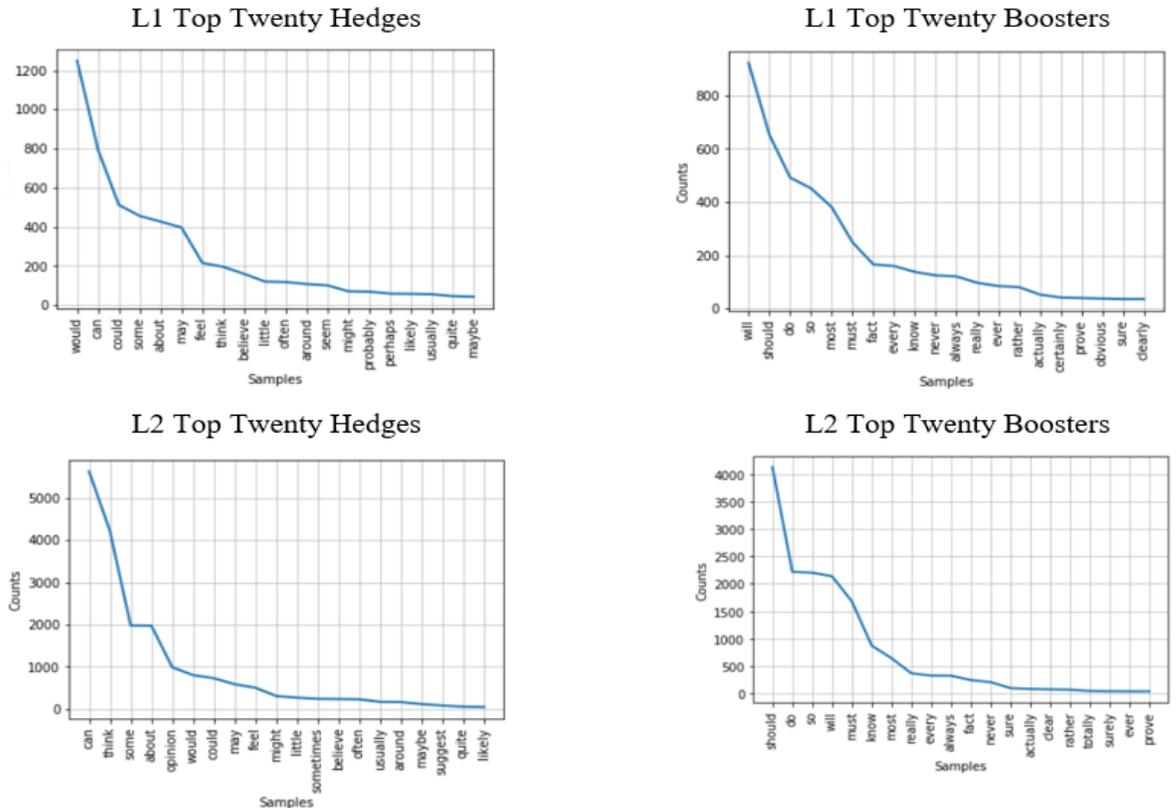


Figure 8 Top Twenty hedges and boosters

At this point, we might question if those more frequent items are from the same part of speech group. Meaning, is there certain part of speech that is more common than others for hedges and boosters, respectively. Do the ratios show any distinction between NSs and NNSs? Based on these questions, most common part of speech was measured, and it revealed that the distributions of part of speech were somewhat different between the two language groups. While both groups most preferred modal verbs to express tentativeness and certainty, their favor towards others did not always match.

However, non-native speakers' English proficiency seems to matter in their preferences of part of speech for hedges. With the two groups on the very extreme compared, B2 - the higher levels - has almost the same pattern in part of speech choices with the native speakers while the opposite group A1 has little resemblance with it. For example, the first three most common part of speech among the native speakers and the B2 students were modals, adjectives, and verbs whereas the A1 group put verbs as the first rank. Regardless of their native language, however, the part of speech that writers preferred the most to use for hedges and boosters was modal verbs.

	L1_Hedge	L2_Hedge	L2_B2_H	L2_A1_H	L1_Booster	L2_Booster
1	Modal (3023)	Modal (8065)	Modal (2021)	Verb (294)	Modal (1825)	Modal (7943)
2	Adjective (1284)	Verb (5125)	Adjective (906)	Modal (219)	Adjective (1401)	Adverb (3704)
3	Verb (633)	Adjective (4480)	Verb (836)	Adjective (173)	Adverb (676)	Verb (3133)
4	Adverb (542)	Noun (991)	Noun (204)	Adverb (65)	Verb (670)	Adjective (1178)
5	Noun (74)	Adverb (899)	Adverb (177)	Noun (38)	Noun (166)	Noun (246)

*Table 7 Most common part of speech*

Although modality is common in English writing as the high native speakers' frequencies prove, non-native speakers are known to sometimes overuse modality depending on the topic. Many of the results of previous studies (e.g. Hinkel, 1995; Kwachka & Basham, 1990) have found that NNSs employ modals of obligation and necessity at much higher rates than NSs. They attribute this for their cultural influence and topic dependence. In her later study in 2009, Eli Hinkel argues that such excessive uses of obligation / necessity modals in East Asian students' writings can possibly downgrade their text quality, bringing in lower evaluation from standardized tests. In addition, other research indicates that EFL learners differ from native speakers in marking epistemic modality (Hu & Li, 2015). This implies that some modals non-native speakers chose to use could interpret differently from their intentions. In short, simple ratios of modal verbs do not guarantee that they have similar patterns to native speakers. The quality of their modality also needs to be considered.

This matter not only belongs to modality, but to general results of my corpus study. For example, some items can be lexically or syntactically ambiguous, but it cannot be sorted out through codes. As a result, there is room for criticism without its quality discussed. This requires the need for the accuracy test to follow in order to identify the reliability of the corpus analysis results. Therefore, in the next few pages, I will discuss the accuracy of the analysis with the means of native and non-native speakers' usage examples of the most common ten items.

### 4.2.2 Accuracy Test Results

Compared to previous literature in which the distribution analysis has been done manually, the computer programming-based research could bring in some restrictions in accuracy. Therefore, I carried out an accuracy test with ten most common hedges and ten most common boosters among native and non-native speakers. The ten most common items were chosen from the intersection results from native and non-native speakers' data. Accordingly, the hedges selected are *would, can, could, some, about, may, feel, think, believe, and little*; and boosters are *will, should, do, so, most, must, fact, every, know, and never*. Among concordance outputs, ten sentences were randomly selected to go through an accuracy test, making it up to 400 samples in total for both hedges and boosters. Each sentence was assigned a value by its predictability: zero if unpredictable and one if predictable. Then the average was computed. This follow up test allowed us to consider the context dependency of rhetorical devices and increased the reliability of the computational study.

Native speakers		Non-native speakers	
Hedges	Boosters	Hedges	Boosters
74%	72%	70%	70%
<b>Total 73%</b>		<b>Total 70%</b>	

Table 8 Accuracy test results

Concerning native speakers, the accuracy score for hedges was 74% and that for boosters was 72%, making it up to 73% in total. Non-native speakers displayed marginally lower yet very similar accuracy as 70%, with hedges and boosters alike. This statistic can be interpreted that in 7 out of 10 sentences hedges or boosters were predicted correctly according to context. There are items that are more predictable than the others. Hedges such as *think* and boosters like *most*, *never*, *must*, and *should* are some examples. In contrast, predicted meanings were rarely found in the occurrences of *so*, *do*, and *about*. We will discuss more details in the following paragraphs.

	<b>Hedges</b>									
<b>Items</b>	About	May	Could	Would	Can	Some	Think	Feel	Little	Believe
<b>Native</b>	0	0.7	0.8	1.0	0.6	0.9	1.0	0.8	0.8	0.8
<b>Non-native</b>	0	0.9	0.7	1.0	0.6	1.0	1.0	0.4	0.7	0.7

*Table 9 Accuracy scores for hedge items*

	<b>Boosters</b>									
<b>Items</b>	Must	So	Know	Do	Most	Never	Will	Should	Every	Fact
<b>Native</b>	1.0	0.4	0.2	0.1	1.0	1.0	1.0	1.0	0.8	0.7
<b>Non-native</b>	1.0	0	1.0	0	1.0	1.0	0.9	1.0	0.9	1.0

*Table 10 Accuracy scores for booster items*

First, there were some items that were more used as other than rhetorical devices. Verb *do* is a great example. More cases were regular verbs as in sentence (b) below or an auxiliary verb to assist another verb as in sentence (a). Its amplifying function as in (c) was found only once in a while.

- a. They *do* not use cellular phones.
- b. Parliament could *do* this at any time.
- c. However, we *do* have a particularly low reputation.

Another case would be the items that describe proximity, *about* and *around*. These two can be used differently in different contexts.

- d. ...talks *about* subsidies...
- e. It weighs *about* 6 lbs.
- f. The dog turned *around* and sit.
- g. Should we meet *around* 1pm?

(e) and (g) are common examples of the two items used to show approximate measures of weight and time, respectively. (d) and (f), on the other hand, are prepositions attached to a preceded verb, which is used for different meanings: *about* in (c) is to bring up a topic and *around* in (e) describes the motion in a circle.

An auxiliary *can* is even more tricky. It is not a surprise that *can* was the only modal verb that had inconsistent results in the accuracy test. Only six out of ten sentences were predictable in both language groups. This may be because it expresses ability and possibility and its interpretation is very context dependent. There is no definite environment where this functions only as an epistemic marker. It is often the case that modal verbs' interpretation is hard to tell even by reading the entire sentence as seen from lexical ambiguity cases above (pp.36-37). Especially for some languages, their corresponding ability markers could be translated into not just *can*, but also *could*, *may*, and *might* into English (Algi, 2012) and Korean is one of these languages. Accordingly, students of these languages as their mother tongue tend to use *can* for various meanings including ability, possibility, and impossibility. In other words, it could imply that Korean speakers' epistemic uses of *can* could be even harder to distinguish from other uses. The below are the examples of *can* from non-native speakers' data.

h. Parents *can* not see their sons or daughter doing...

i. I think if someone *can* know about them...

j. However there *can* exist...

k. They feel physical touch *can* get education effect easily.

Sentences (h) and (i) are quite clear what the authors intended to say with *can*. In sentence (h) with negation, it is most likely to suggest parents' non-permission or inability to

see their children. This case is obviously not used as a hedge that expresses uncertainty. In case of (i), a stative verb *know* is put together with an auxiliary verb of ability and this is a common error made by EFL / ESL students. We cannot refer to one's ability 'to know' because to know is a complete state of knowledge acquired, not including the action of acquiring it. However, in Korean language, verb *know* is not a stative and thus can even be progressive as in 'I am knowing the fact.' As a result, from Korean speakers' perspectives, this *can* could be both ability and possibility depending on its context. Nevertheless, the structure of this sentence makes it clear that this *can* in (i) is ability not a possibility. An if-conditional-clause already connotes possibility, so the modal *can* will be redundant if it refers to possibility as well. Given that, it is more likely to carry the meaning of ability. Therefore, setting aside its grammaticality, *can* in (i) is considered an ability marker. While (h) and (i) are quite apparent in the function of the modal *can*, most other cases are wishy-washy. (j) and (k), for instance, could have two different interpretations either of ability or possibility. Sometimes, it is hard to classify unless the writers reveal what it is.

Earlier studies also claim that *can* is poly-pragmatic and conveys different meanings. Biber (2010) indicates that "*could, may and might* are used almost exclusively to mark logical possibility... and *can* is often ambiguous with a logical possibility meaning (p.179)." In this sense, Hyland and Milton (1997) concern L2 writers' employment of *can* due to its complicated nature in pragmatics. Similarly, Holmes (1982) argues that to express and interpret epistemic modality is a challenging task for learners of English because of the interaction of different types of meaning in different contexts.

Because of the difficulty of separating the two meanings of *can*, many earlier research pairs possibility and ability together in contrast to obligation and necessity (Algi, 2012).

For this reason, I considered discarding the item from the list in corpus analysis in the first place, yet I found that *can* as a hedge is critical in the epistemic modality (Hu & Li, 2015). Fraser (1975) emphasizes the function of modals in writing that certain performative verbs preceded by some modals are attenuated in their illocutionary force. Even the number of modals used in writings explains the important roles of modal verbs in Rhetorics. This is well shown in Kang's (2017) study that epistemic modal verbs are the first or second most frequent categories both in L1 and L2 writings compared to others like nouns, adverbs, or lexical verbs. Especially, *can* is found very commonly in both L1 and L2 writings, entering in top ten most frequent epistemic markers. Similarly, in Algi's (2012) studies, *can* is the most frequently employed modal, *may* being the second. In short, while those items - *do*, *about/around*, and *can* - might drag down the total accuracy score, it is indispensable to include them in the analysis since they play a weighty role in providing overall distributions. Furthermore, Attarn (2014) and Farrokhi and Emami (2008) both discovered that in academic papers the uses of *can* as hedges greatly excelled those as boosters in general. Considering the fact that the argumentative writing texts were collected from college academic settings, this claim could be also true to this corpus. Based on this premise, instead of removing the modal verb *can* from my research, the sample sentences which includes the meaning of possibility, were all coded 1 to be predictable.

Now we will have a look at some of the items that had 100 % of predictability in the accuracy test. *Think*, for example, was found to be dominantly functioning as a hedge by both NS and NNS. Some studies (e.g. Kang, 2017; Kim, 2009) claimed that Korean learners might not have proper understanding of a verb *think* as its hedging role and this deficiency led to low frequencies of using *think* as a hedge. Despite their concerns, Korean students not only perceived *think* more towards a hedging device in Study 1, but also were able to employ *think* for the expression of uncertainty and politeness in delivering their arguments. 10 out of 10 non-native speakers' samples with *think* was used with the first person singular subject. Given the fact that NNSs' second most common hedge was *think*, EFL students seem to readily use prefabricated chunks like 'I think' to mark their opinions with mitigated voice. Moreover, their voice in such phrases is even less confident than that of native speakers. While native speakers often added strength to their subjective thoughts by pairing the phrase *I think* with intensifying adverbs, such as *always* or *certainly*, Koreans used the phrase simply by itself. The test environment might have affected their favor over using ready-to-use phrases without detailed elaboration. Even so, the findings above altogether lead us to assume that the phrase *I think* is one of the main opinion marking hedging tools for English learners.

Another major item to look at is *would*. The modal *would* was categorized as a booster in Farrokhi and Emami's (2008) list, but it completely appeared as a hedge from the both language groups' samples that were collected for the accuracy test. See some examples below.

- l. The radical theorists *would* agree that...
- m. Because she felt it *would* lead to a loss of our...
- n. There is no doubt an integrated market *would* have multiple benefits.
- o. Normal students *would* listen to their teachers.
- p. We *would* be able to make mature Internet culture.
- q. There *would* be no online theft and no bad culture.

Sentences (l) to (n) are from native speakers' texts and the other three are from non-native speakers' texts. Though both groups show stable predictability for *would*, the non-native speakers' frequencies are much lower than the native speakers'. The frequencies of *would* occupies 22% out of the total number of native speakers' hedges while the Korean speakers' remains very low at 4%. Korean learners' inactive uses of the modal *would* can be caused by semantic complexity, which plays an important factor in the acquisition and the production of epistemic devices. Similar to the modal verb *can*, *would* is complex in that it has a few different meanings (Coats, 1983; Perkins, 1983; Palmer, 2014). While the various meanings can be taught, it may often require readers to have intuitions to discern the right meaning. This generally bewilders English learners as a second language across the culture (Hinkel, 2002). This may frustrate learners from choosing such modals for hedging and use other items instead. This phenomenon is well proved by the results from the ten most common items list. In contrast to native speakers who used three modal

verbs - *would*, *can*, and *could* - the most, only one modal verb took higher rank in non-native speakers' corpus.

### **4.3 Limitations and Further Discussion**

#### **4.3.1 Topic dependency**

Based off of results from the corpus analysis and the follow-up accuracy test, we may consider some other effects on them. To give one example, assigned topics for the writing test can vary the uses of certain devices than others. Algi (2012) found that some particular topics drew more employments of modal *can* with ability function among EFL students. One of the assigned topics for the argumentative writings in the YELC corpus was 'Why is college education necessary?'. Such a topic is likely to increase the number of *can* in their writings because the writers would want to describe skills or abilities college education enables students to acquire. Therefore, the topic choices may partially account for the especially big ratio of *can* (1st rank - 29%) in the non-native speakers' data (compared with native speakers' 12%), let alone other considerations mentioned above: the complexity of the modality and the influence of L1 background.

It is difficult to completely exclude the topics that have potential to affect their linguistic choices. It is hard for sure to rule out writers' cultural backgrounds and values from intervening their writing. However, if the written texts can be collected from the same testing environment with several different topics given, it can help avoid such

variables. Therefore, it will be valuable to conduct more advanced experimental studies starting from the collection of corpus data from scratch with such conditions controlled, not allowing any topics tilted towards one nation.

#### **4.3.2 The impact of teaching**

Korean learners showed repeated use of certain words or phrases, with little alteration. The opinion marking phrase *I think* is a great example. This might have come from their English lessons in the classroom. Middle school English in South Korea covers this phrase as a good device to start their opinion statement. In fact, the impact of teaching on learners' writing has been claimed by many theorists. As one instance, Hu and Li (2015) found that the items that were explicitly taught appeared more frequently in the participants' writings. Of course, the testing environment with the time limit may have narrowed down their options and led to repetitive uses of *I think* among the second language learners, but their memory loaded from the class could also be another great contribution for that.

At this point, we may well think why this is not true to modality if they are also taught at schools. As recognized above, conditions and degrees of modality are not always straightforward (Hinkel, 1995). Therefore, learning to express and interpret epistemic modality might not be an easy task for learners of English as a second language. Holmes (1982) highlights three sources of potential difficulty. One is the

problem of establishing the precise degree of certainty expressed by particular linguistic forms; another is the problem of selecting the right item among the wide range of linguistic devices available for signaling this aspect of meaning; and last but not least is the interaction of different types of meaning in various contexts. Whereas those ready-made chunks, such as *I think* or *I believe*, can be relatively easy to access and apply, modality has more steps to process until the writers choose an appropriate one for the context. Apparently, the divergences in NS and NNS modal verb uses may depend on more complex factors than can be explained by a lack of teaching. Nevertheless, the way of presenting modality in the classrooms cannot be neglected either. Some have pointed out the deficiency in teaching pragmatics of modal verbs. Hinkel (1995), for instance, claims that the teaching of modality has largely focused on forms rather than sociopragmatic meanings and implications. This also applies to South Korea where English class is predominantly focused on vocabulary, grammar, and reading.

Accordingly, there seems to be a further need for addressing the relationship between teaching and application of modality. Since the epistemic modality is found to be one of the most common ways to either tone down or tone up one's argument, this is not a trivial item in order to elicit proper comprehension and production of it. While many scholars have emphasized the importance of teaching L2 modality (e.g. Hinkel, 1995; Holmes, 1982; Hu & Li, 2015), it cannot be overstated that more experimental studies and profound practical analysis of teaching and its output should be investigated. To what extent is modality presented in school curricula and how far does it lead to

learners' actual production? Do EFL learners display any differences between comprehension and production of epistemic modality? What are the effective ways to enhance second language learners' intuitions towards the use of modality in different contexts? These could be only a few example research questions that need to be addressed. In addition, specific analysis of teaching epistemic modality in Korean classroom settings could also be helpful.

## **Chapter 5. Conclusion**

The results of Study1 established the baseline of how native speakers perceive hedges and boosters and provided comparative data of Korean English learners that they perceive hedges and boosters very similarly with native speakers of English. The main finding from this study is that native and non-native speakers alike perceive hedges and boosters as conveying different levels of certainty. Specifically, sentences with hedges are perceived as less certain by both groups, and sentences with boosters are perceived as more certain for both groups. These results are compatible with the observations and definitions of hedges and boosters in the theoretical linguistics literature. While they both displayed the same patterns in general, native speakers felt much less certainty towards the hedge items than non-native speakers did and the latter group reacted more strongly towards the booster items than the former group. There were only a few exceptions that were deviated from the regular patterns: native speakers gave a higher mark than the

average to a hedge *quite* and non-native speakers gave higher marks to two cognate hedges *apparent* and *apparently*. In terms of boosters, while NNSs showed more consistency over all items in their responses, NSs marked auxiliary boosters lower than the average, such as *must*, *won't*, *do*, and *can't*. From these instances, native speakers were found to have more difficulties in evaluating context-dependent items. A degree adverb, *quite*, can connote different nuances depending on the context, which non-native speakers often fail to notice. Epistemic modal verbs are known to hold different pragmatic meanings which are difficult to discern unless you have had much exposure to them. This factor reversely operated as an advantage to NNSs by not needing to be confused while NSs had so many options to think through. Meanwhile, Korean learners showed stronger perceptions of *apparent* and *apparently* which were considered hedges by linguists and native speakers. On one hand, the Korean language which does not have the exact same word can be blamed for this. On the other hand, dictionaries most of which first define them as *obvious* and *obviously* and Korean students' who would stop learning after the first definition can all be reasons. Overall, native speakers were more sensitive to socio-pragmatic meanings of lexical items, so they would show more divergence in their interpretations of statements with poly-pragmatic rhetorical devices. In contrast, non-native speakers showed constant values towards most items, regardless of their contextual environments, as they are less constituted with the contextual knowledge. That is, their perceptions show a little more agreement to linguists' definitions of hedges and boosters, unless they have different meanings by context.

Unlike their comprehension, however, Korean English learners displayed a distinctive pattern in production of hedges and boosters from native speakers. Although both language groups showed in their argumentative writings more numbers of hedges than boosters, the total frequencies of hedges and boosters per 1,000 words were much higher in non-native speakers' corpus, regardless of proficiencies. Nevertheless, there existed some discrepancies among different non-native speakers' L2 proficiency groups. More fluent NNS English writers exhibited a closer pattern to NS writers. More proficient students used similar part of speech more frequently to native speakers and diversified the types of hedges and boosters. Even so, the total number of types fell short in NNSs' corpus compared to NSs'. This implies that non-native speakers attempt to use more rhetorical devices in stating their arguments, but they tend to employ the same items repeatedly. To some readers this could give the impression that the writers overused certain items, and this affects their general evaluation of the text quality. Their repetitive uses of certain items were outstanding in using an opinion marking phrase, *I think*. It was the second most common hedge among non-native speakers and they kept its bare form most of the time. They preferred such prefabricated chunks over others in hedging, and this was opposite to native speakers' tendency. L2 speakers seem to find it hard to grasp the nuances of different pragmatic meanings of epistemic markers and, thus, often choose such ready-to-use chunks for safe writing.

I believe the two studies have much methodological contributions. Study 1 is valuable as it is the first behavioral data, to my knowledge, which allows us to establish

the baseline on how native speakers perceive hedges and boosters. This lays a logical foundation for us to compare comprehension of non-native speakers with. Study 2 is also a groundbreaking attempt in that it first showed the possibility that automated corpus analysis methods can be used for the analysis of hedges and boosters. The clear advantage of these methods over manual analysis is that it can be scaled up and produce statistical values over millions of words. Although there are limitations in utilizing programming for data analysis yet, this study surely made a good leap towards further research of computational linguistics on this topic.

Study1 and Study2 have theoretical benefits as well. The two studies shed light on comprehension and production of hedges and boosters for native and non-native speakers of English. The results of Study 1 show that both groups perceive hedges and boosters in a similar manner. In Study 2, the frequency with which native and non-native speakers use hedges and boosters is different. Repetitions of certain lexical items made non-native speakers' texts to be highly saturated with hedges and boosters to outnumber the native speakers' frequency. Finally, this study can be a good addition to research and practice in Korean education settings. South Korea has their indigenous English education policy and curriculum and the attention English get in this society is immense. By comparing Korean English learners with native speakers, the two studies give much implications on how Korean English learners comprehend and produce hedges and boosters and in what manners rhetorical devices should be dealt with in English lessons. With detailed classifications of students by their proficiency levels and experiences of living abroad,

such other factors were first considered in analyzing Koreans' perception and production of hedges and boosters.

## References

- Abbuhl, R. (2006). Hedging and boosting in advanced-level L2 legal writing: The effect of instruction and feedback. In H. Byrnes, H. Weger-Guntharp & K. A. Sprang (Eds.), *Educating for advanced foreign language capacities: Constructs, curriculum, instruction, assessment* (pp. 152-164). Washington, D.C.: Georgetown University Press.
- Afflerbach, P. & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behaviour*, 16 (4): 307-322.
- Afshar, H. S., Asakereh, A., & Rahimi, M. (2014). The impact of discipline and being native/non-native on the use of hedging devices. *Procedia-Social and Behavioral Sciences*, 136, 260-264.
- Akbik, A., Blythe, D., & Vollgraf, R. (2018, August). Contextual string embeddings for sequence labeling. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 1638-1649).

- Algi, S. (2012). Hedges and boosters in L1 and L2 argumentative paragraphs: Implications for teaching L2 academic writing. *Unpublished master's thesis*. Middle East Technical University, Turkey.
- Allison, D. (1995). Assertions and alternatives: Helping ESL undergraduates extend their choices in academic writing. *Journal of Second Language Writing*, 4, 1-15.  
[http://dx.doi.org/10.1016/1060-3743\(95\)90020-9](http://dx.doi.org/10.1016/1060-3743(95)90020-9)
- Attarn, A. (2014). Study of metadiscourse in ESP articles: A comparison of English articles written by Iranian and English native speakers. *International Journal of Learning, Teaching and Educational Research*, 5(1), 63-71.
- Biber, D. (2010). Longman student grammar of spoken and written English. Pearson Education India.
- Biber, D. and Finegan, E. (1989) 'Styles of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect', *Text* 9(1): 93–124.
- Bloor, M., & Bloor, T. (1991). Cultural expectations and socio-pragmatic failure in academic writing. In P. Adams, B. Heaton, & P. Howarth (Eds.), *Socio-cultural issues in English for academic purposes* (pp. 1-12). Basingstoke: Modern English Publications/British Council.
- Bolinger, D. (1972) Degree Words. *The Hague and Paris: Mouton*.

- Campbell, I. (2007) Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine* 26, 3661-3675.
- Coates, J. (1983). *The Semantics of modal auxiliaries*. Beckenham:Croom Helm.
- Coates, J. (1987). Epistemic modality and spoken discourse. *Transactions of the philological society*, 85(1), 110-131.
- Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. In D.R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literature, language, and learning: The nature and consequences of reading and writing* (pp. 105-123). Cambridge: Cambridge University Press.
- Chafe, W. (1986). Evidentiality in English conversation and academic writing. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 261-272). Norwood, NJ: Ablex.
- Chafe, W. & Danielewicz, J. (1987). Properties of spoken and written language. In R. Horowitz & F. Samuels (Eds.), *Comprehending oral and written language* (pp. 83-114). New York: academic Press.
- Choi, Y.-H., & Ko, M.-S. (2005). Hedging in EFL academic writing of Korean postgraduates. *English Teaching*, 60(1), 3-27.

- Crismore, A. (1989) *Talking with Readers: Metadiscourse as Rhetorical Act*. New York: Peter Lang.
- Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, 40(1), 95-113.
- Farrokhi, F., & Emami, S. (2008). Hedges and boosters in academic writing: native vs. non-native research articles in applied linguistics and engineering. *Journal of English Language Pedagogy and Practice*, 1(2), 62-98.
- Fraser, B. (1975). Hedged performatives. In *Speech acts* (pp. 187-210). Brill.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. In A. Cowie (Ed.), *Phraseology* (pp. 145-160). Oxford: Oxford University Press.
- Halliday, M.A.K. (1994) *An Introduction to Functional Grammar* (2nd edn). London: Edward Arnold.
- Hayes, J.R. & Flower, L.S. (1983). Uncovering cognitive processes in writing: an introduction to protocol analysis. In P. Mosenthal, L. Tamar, & S. Walmsey (eds). *Research in Writing*. New York: Longman. 206-220

Hermeren, L. (1978). On modality in English: A study of the semantics of the modals.

Lund: CWK Gleerup.

Hinkel, E. (1995). The Use of Modal Verbs as a Reflection of Cultural Values. *TESOL*

*Quarterly* 29, 325–343.

Hinkel, E. (2002). *Second language writers' text*. Mahwah, NJ: Lawrence Erlbaum

Associates.

Hinkel, E. (2005). Analysis of second language text and what can be learned from them.

In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning*. New Jersey: Lawrence Erlbaum Associate, 615-628.

Hinkel, E. (2009). The effects of essay topics on modal verb uses in L1 and L2 academic

writing. *Journal of Pragmatics*, 41(4), 667-683.

Holmes, J. (1982). Expressing doubt and certainty in English. *RELC journal*, 13(2), 9-28.

Holmes, J. (1984). Hedging your bets and sitting on the fence: Some evidence for hedges

as support structures. *Te Reo*, 27 (1), 47-62.

Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9 (1), 21-

44.

Holmes, J. (1995). *Women, men and politeness*. London: Longman.

- Hu, C., & Li, X. (2015). Epistemic Modality in the Argumentative Essays of Chinese EFL Learners. *English Language Teaching*, 8(6), 20-31.
- Hu, Z., Brown, D., & Brown, L. (1982). Some linguistic differences in the written English of Chinese and Australian students. *Language Learning and Communication*, 1, 39-49.
- Huebler, A. (1983). Understatements and hedges in English. Amsterdam: John Benjamins.
- Huh, M. H., & Lee, I. (2016). On the Use of Metadiscourse in EFL Undergraduate Student Writing. *English Teaching*, 71(3), 99-120.
- Hunston, S. and Thompson, G. (eds) (2000) Evaluation in Text. Oxford: Oxford University Press.
- Hwang, S. J. (1987). *Discourse features of Korean narration*. Arlington, TX: The Summer Institute of Linguistics and The University of Texas.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. ENGLISH FOR SPECIFIC PURPOSES-ELMSFORD AND OXFORD-, 13, 239-239.
- Hyland, K. (1996). Talking to the academy: Forms of hedging in science research articles. *Written Communication*, 13(2), 251-281.
- Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam/Philadelphia.

- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied linguistics*, 20(3), 341-367.
- Hyland, K. (2000). Hedges, boosters and lexical invisibility: Noticing modifiers in academic texts. *Language Awareness*, 9(4), 179-197.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, 7(2), 173-192.
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of second language writing*, 6(2), 183-205.
- Hyland, K. and Tse, P. (2004) 'Hooking the Reader: A Corpus Study of Evaluative That in Abstracts', *English for Specific Purposes*.
- Kang, J. (2017). Epistemic stance construction in L1 and L2 undergraduate argumentative writing.
- Kim, C. K. (2009). A corpus-based comparison of metadiscourse in argumentative texts of a British newspaper and Korean university students' English argumentative texts. *Discourse and Cognition*, 16(3), 65–88.
- Kobayashi, Y. (2016). Investigating metadiscourse markers in Asian Englishes: A corpus-based approach. *Language in Focus*, 2(1), 19-35.

- Kwachka, P. & Basham, C. (1990). Literacy acts and cultural artifacts: on extensions of English modals. *Journal of Pragmatics* 14, 413–429.
- Lakoff, G. (1972). Hedges: A study of in meaning criteria and the logic of fuzzy concepts. *ChicagoLinguistic Society Papers*, 8, pp. 183-228.
- Lee, S. (2006). A corpus-based analysis of Korean EFL learner's use of amplifier collocations. *ENGLISH TEACHING (영어교육)*, 61(1), 3-17.
- Leech, G. (2003). Modality on the move: The English modal auxiliaries 1961–1992. In: Facchinetti, Roberta, Krug, Manfred, Palmer, Frank (Eds.), Modality in Contemporary English. *Mouton de Gruyter, Berlin and New York*. 223–240.
- Lennon, P. (1989). Introspection and intentionality in advanced second language acquisition. *Language Learning*. 39 (3): 375-96.
- Lewin, B. A. (2005). Hedging: an exploratory study of authors' and readers' identification of 'toning down' in scientific texts. *Journal of English for Academic Purposes*, 4(2), 163-178.
- LOCNESS. (2004). *Louvain corpus of native English essays*. Retrieved February, 27, 2019 from the World Wide Web  
<https://www.lernercorpusassociation.org/resources/tools/locness-corpus/>

- Lorenz, G. (1998). Overstatement in advanced learners' writing: Stylish aspects of adjective intensification. In S. Granger (Ed.), *Learner English on computer* (pp.53-66). London: Longman.
- Low, G. (1996). Intensifiers and hedges in Questionnaire items and the lexical invisibility hypothesis. *Applied Linguistics*. 17 (1): 1-37.
- Mirzapour, F., & Mahand, M. R. (2012). Hedges and boosters in native and non-native library and information and computer science research articles. 3L: Language, Linguistics, *Literature*®, 18(2).
- Muller, G. (1985) *The American College Handbook of Contemporary English*. New York:Harper & Row.
- Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied linguistics*, 10(1), 1-35.
- Olinghouse, N., & Wilson, G. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, 26(1), 45-65.
- Palmer, F. R. (1979). *Modality and the English Modals*. London: Longman
- Palmer, F. R. (2014). *Modality and the English modals*. Routledge.

- Park, M.-R. (1990). Conflict avoidance in social interaction. In H. Hoji (Ed.), *Japanese and cc.* John Benjamin Publishing Company.
- Quirk, R., Greenbaum, S., Leech, G., & Starvik, J. (1985). *A comprehensive grammar of English language.* London and New York: Longman.
- Richardson JTE (2011) The analysis of 2 x 2 contingency tables - Yet again. *Statistics in Medicine* 30:890.
- Robberecht, P., & Van Peteghem, M. (1982). A functional model for the description of modality. *In Fifth International Conference on Contrastive Projects*, Jyväskylä, Finland. Scarcella, R., & Brunak, J. (1981). On speaking politely in a second language. *International Journal of the Sociology of Language*, 1981(27), 59-76.
- Šeškauskienė, I. (2008). Hedging in ESL: A case study of Lithuanian learners. *Kalbu studijos*, (13), 71-76.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27 (4), 657-677.
- Skelton, J. (1988a). The care and maintenance of hedges. *ELT journal*, 42(1), 37-43.
- Skelton, J. (1988b). Comments in academic articles. *Applied linguistics in society.* London: CILTIBAAL.

- Smagorinsky, P. (1994). Think aloud protocol analysis: beyond the black box. In P. Smagorinsky (ed) *Speaking about Writing: reflections on research methodology. Thousand Oaks, CA: Sage*, 3-19.
- Stevenson, A. (Ed.). (2010). *Oxford dictionary of English*. Oxford University Press, USA.
- Stratman, J. & Hamp-Lyons, L. (1994). Reactivity in concurrent think aloud protocols: Issues for research. In P. Smagorinsky (ed). 89-114.
- Swain, M. (2005). The output hypothesis: Theory and research. In *Handbook of research in second language teaching and learning* (pp. 495-508). Routledge.
- Talbot, D. (1992). A preliminary investigation into the recognition of expository text structure by Chinese tertiary students in think aloud interviews. *Perspectives*. 4 (2): 41-61.
- Tan, H., & Eng, W. B. (2014). Metadiscourse Use in the Persuasive Writing of Malaysian Undergraduate Students. *English Language Teaching*, 7(7), 26-39.
- Thompson, G., & Thetela, P. (1995). The sound of one hand clapping: The management of interaction in written discourse. *Text-Interdisciplinary Journal for the Study of Discourse*, 15(1), 103-128.
- Van der Hoek, W., & Meyer, J. J. (1997). A complete epistemic logic for multiple agents. In *Epistemic Logic and the Theory of Games and Decisions*. 35-68.

- Vande Kopple, W. J. (1997). *Refining and Applying Views of Metadiscourse*.
- Varttala, T. (2001). *Hedging in scientifically oriented discourse: Exploring variation according to discipline and intended audience*. Electronic doctoral dissertation. Acta Electronica Universitatis Tamperensis 138 (<http://acta.uta.fi/pdf/951-44-5195-3.pdf>).
- Ventola, E. (1997). Modalization: Probability-an exploration into its role in academic writing. *Culture and styles of academic discourse*, 157-179.
- Webster, M. (2002). *New Encyclopedic Dictionary*.
- Yarber, R. (1985) *Writing for College*. Glenview, IL: Scott Foresman.
- YELC. (2011) *Yonsei English Learner Corpus*. Retrieved April 21, 2019 from Yonsei Information Lab.

## Appendix 1. Survey Questions

## A. Questionnaire entries

<b>Hedges</b>	<b>Target sentence</b>	<b>Neutral sentence</b>
usually	High school students usually study hard.	High school students study hard.
seem	They seem to grow faster.	They grow faster.
think(head)	I think that computer has made life easier.	Computer has made life easier.
would	Smoking would be harmful to children.	Smoking is harmful to children.
likely	The economy is likely to recover slowly after the long recession.	The economy will recover slowly after the long recession.
can't	This can't be true.	This isn't true.
often	Cancer often occurs without symptoms.	Cancer occurs without symptoms.
might	The lawyer might be able to win the case.	The lawyer can win the case.
could	The second method could cause more problems.	The second method causes more problems.
may	Explicit instruction may help accelerate learning.	Explicit instruction helps accelerate learning.
maybe	Maybe he is a lawyer.	He is a lawyer.
suggest	I suggest that regular check-ups help prevent worse symptoms.	Regular check-ups help prevent worse symptoms.
suppose	I suppose that practice is the best way to accomplish something.	Practice is the best way to accomplish something.
propose	I propose that teenagers prefer online class.	Teenagers prefer online class.
assume	I assume that globalization made traveling easier.	Globalization made traveling easier.
about	The box weighs about 20lb.	The box weighs 20lb.
around	The package will be delivered around noon.	The package will be delivered at noon.
probably	Being a vegetarian is probably a hard task.	Being a vegetarian is a hard task.
perhaps	Perhaps John will come over for dinner.	John will come over for dinner.
quite	The results are quite different.	The results are different.

sometimes	Its value sometimes exceeds the limit.	Its value exceeds the limit.
nearly	Nearly all packaged foods contain high level of sugar.	All packaged foods contain high level of sugar.
possibly	There is possibly no data that proves the idea.	There is no data that proves the idea.
slightly	These two colors are slightly different.	These two colors are different.
potentially	This new oil can potentially drop oil prices.	This new oil can drop oil prices.
unlikely	This secret is unlikely to be made public.	This secret will not be made public.
approximately	The train leaves approximately at 7.	The train leaves at 7.
rarely	My uncle rarely visits us.	My uncle does not visit us.
hardly	I hardly find any problems in this.	I do not find any problems in this.
opinion	In my opinion, stress is the main cause of diseases.	Stress is the main cause of diseases.
attempt	This study will attempt to explain their relationships.	This study will explain their relationships.
hypothetically	Hypothetically, the problem is solved.	The problem is solved.
Believe (head)	I believe that the global warming is serious.	The global warming is serious.
little	There are little more pros than cons to not having a smartphone at school.	There are more pros than cons to not having a smartphone at school.
I think (middle)	The computer, I think, has made life easier.	The computer has made life easier.
I think (end)	The computer has made life easier, I think.	The computer has made life easier.
I feel	I feel that volunteer work is important.	Volunteer work is important.
believe(middle)	The global warming, I believe, is serious.	The global warming is serious.
believe(end)	The global warming is serious, I believe.	The global warming is serious.
rare	It is rare that opportunity knocks twice.	Opportunity does not knock twice.

<b>Boosters</b>	<b>Target sentence</b>	<b>Neutral sentence</b>
must	Mr. Jones must have been sick	Mr. Jones has been sick
clearly	This study clearly shows that weight training can control diabetes.	This study shows that weight training can control diabetes.
rather	People will rather stay home in hot weather.	People will stay home in hot weather.
always	Flower is always a good gift.	Flower is a good gift.
definitely	Mrs. Kim will definitely come.	Mrs. Kim will come.
indeed	They can run very fast indeed.	They can run very fast.
prove	This proves that kids learn faster.	This suggests that kids learn faster.
will	This will turn out alright.	This turns out alright.
won't	It won't be wrong.	It is not wrong.
most	New York is the most crowded city.	New York is a crowded city.
no	There is no evidence.	There is not any evidence.
every	Every ant is found to be diligent.	Ants are found to be diligent.
never	The politician will never say the truth.	The politician will not say the truth.
so	This fact is so true.	This fact is true.
really	People really hope to foresee their future.	People hope to foresee their future.
ever	This will never ever happen again.	This will never happen again.
actually	She could actually speak German.	She could speak German.
certainly	He certainly can speak Korean.	He can speak Korean.
extremely	The reaction was extremely positive.	The reaction was positive.
completely	Their findings were completely identical.	Their findings were identical.
truly	The empire in Mexico was truly large.	The empire in Mexico was large.
surely	He will surely be missed.	He will be missed.
obviously	Jenna obviously likes her work.	Jenna likes her work.
highly	I highly recommend Sarah for the position.	I recommend Sarah for the position.
greatly	The characteristics of the two differ greatly.	The characteristics of the two differ.
apparently	Apparently, he was working for his father's company.	He was working for his father's company.
frequently	The participants were frequently found to multi-task.	The participants were found to multi-task.
totally	They totally match well.	They match well.

absolutely	His thought was absolutely right.	His thought was right.
necessarily	This endeavor necessarily involves some risk.	This endeavor involves some risk.
exactly	These two pieces are exactly the same size.	These two pieces are the same size.
assuredly	The merger will assuredly lead to job layoffs.	The merger will lead to job layoffs.
know	I know that the outcome is not correct.	The outcome is not correct.
fact	It is a fact that water boils at 100 °C.	Water boils at 100 °C.
prove	It is proved that fruit juice has more sugar than its fruit	Fruit juice has more sugar than its fruit.
obvious	It is obvious that the weather affects the mood.	The weather affects the mood.
sure	I am sure that this machine is safe.	This machine is safe.
clear	It is clear that eating less helps digestion.	Eating less helps digestion.
apparent	This is an apparent truth.	This is a truth.
definite	It is definite that there will be a change to the schedule.	There will be a change to the schedule.
evident	It was evident that the doctor felt uncertain.	The doctor felt uncertain.
do	We do want to help you.	We want to help you.

## B. Question Formats

Please read and evaluate the following two statements.

- (1) Computers, I think, have made life easier.
- (2) Computers have made life easier.

In which sentence is the speaker more confident?

- In sentence (1) the speaker is more confident.
- In sentence (1) the speaker is somewhat more confident.
- In sentence (2) the speaker is somewhat more confident.
- In sentence (2) the speaker is more confident.

## Appendix 2. List of Hedges and Boosters in Corpus Analysis

<b>Hedges</b>	<b>POS 1</b>			<b>Boosters</b>	<b>POS 2</b>		
		believe					AD
seem	ADJ	e	VBP	most	ADJ	undoubtedly	V
						Unquestionably	AD
likely	ADJ	feel	VBP	every	ADJ	ly	V
about	ADJ	tend	VBP	obvious	ADJ	Will	AUX
around	ADJ			sure	ADJ	won't	AUX
little	ADJ			clear	ADJ	Should	AUX
hypothetical	ADJ			apparent	ADJ	shouldn't	AUX
rare	ADJ			definite	ADJ	Fact	NN
some	ADJ			evident	ADJ	Prove	VBP
usually	ADV			clearly	ADV	Know	VBP
often	ADV			rather	ADV	Prove	VBP
maybe	ADV			always	ADV	Do	VBP
probably	ADV			definitely	ADV		
perhaps	ADV			indeed	ADV		
quite	ADV			never	ADV		
sometimes	ADV			so	ADV		
nearly	ADV			really	ADV		
possibly	ADV			ever	ADV		
slightly	ADV			actually	ADV		
potentially	ADV			certainly	ADV		
unlikely	ADV			extremely	ADV		
approximate							
ly	ADV			completely	ADV		
rarely	ADV			truly	ADV		
hardly	ADV			surely	ADV		
hypothetically	ADV						
				obviously	ADV		
oftentimes	ADV			highly	ADV		
would	AUX			greatly	ADV		
can	AUX			apparently	ADV		
might	AUX			frequently	ADV		
could	AUX			totally	ADV		
may	AUX			absolutely	ADV		
can	AUX			necessarily	ADV		
opinion	NN			exactly	ADV		

attempt	NN	assuredly	ADV
		indisputabl	
think	VBP	y	ADV
suggest	VBP	inevitably	ADV
suppose	VBP	intensively	ADV
propose	VBP	unarguably	ADV
assume	VBP	undeniably	ADV

## Appendix 3. Statistical Results for Study1

**Tests of Within-Subjects Effects**

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
HedgesAndBoosters	Sphericity Assumed	70.257	1	70.257	269.214	.000	.729
	Greenhouse-Geisser	70.257	1.000	70.257	269.214	.000	.729
	Huynh-Feldt	70.257	1.000	70.257	269.214	.000	.729
	Lower-bound	70.257	1.000	70.257	269.214	.000	.729
HedgesAndBoosters * Language	Sphericity Assumed	.523	1	.523	2.005	.160	.020
	Greenhouse-Geisser	.523	1.000	.523	2.005	.160	.020
	Huynh-Feldt	.523	1.000	.523	2.005	.160	.020
	Lower-bound	.523	1.000	.523	2.005	.160	.020
Error (HedgesAndBoosters)	Sphericity Assumed	26.097	100	.261			
	Greenhouse-Geisser	26.097	100.000	.261			
	Huynh-Feldt	26.097	100.000	.261			
	Lower-bound	26.097	100.000	.261			

**Tests of Within-Subjects Contrasts**

Measure: MEASURE\_1

Source	HedgesAndBoosters	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
HedgesAndBoosters	Linear	70.257	1	70.257	269.214	.000	.729
HedgesAndBoosters * Language	Linear	.523	1	.523	2.005	.160	.020
Error (HedgesAndBoosters)	Linear	26.097	100	.261			

**Levene's Test of Equality of Error Variances<sup>a</sup>**

	F	df1	df2	Sig.
Hedges	12.291	1	100	.001
Boosters	1.209	1	100	.274

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Language  
Within Subjects Design: HedgesAndBoosters

**Tests of Between-Subjects Effects**

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	922.767	1	922.767	3342.955	.000	.971
Language	6.698	1	6.698	24.265	.000	.195
Error	27.603	100	.276			

### Within-Subjects Factors

Measure: MEASURE\_1

HedgesAndBoosters	Dependent Variable
1	Hedges
2	Boosters

### Between-Subjects Factors

		Value Label	N
Language	1	Native	47
	2	Nonnative	55

### Descriptive Statistics

	Language	Mean	Std. Deviation	N
Hedges	Native	1.3122	.32527	47
	Nonnative	1.7773	.44581	55
	Total	1.5630	.45687	102
Boosters	Native	2.5911	.67716	47
	Nonnative	2.8530	.56122	55
	Total	2.7323	.62805	102

## Certificate Of Completion

Envelope Id: 0C218D00229643D4B5518B8939A8E5C5	Status: Completed
Subject: Please DocuSign: Final Thesis_MinsunLee.pdf	
Source Envelope:	
Document Pages: 102	Signatures: 2
Certificate Pages: 1	Initials: 0
AutoNav: Enabled	Envelope Originator:
Envelopeld Stamping: Enabled	Cynthia Losinsky
Time Zone: (UTC-08:00) Pacific Time (US & Canada)	1600 Holloway Ave
	Administration 17A
	San Francisco, CA 94132
	cynthial@sfsu.edu
	IP Address: 130.212.169.64

## Record Tracking

Status: Original	Holder: Cynthia Losinsky	Location: DocuSign
5/12/2020 1:29:44 PM	cynthial@sfsu.edu	

## Signer Events

Anastasia Smirnova  
 smirnov@sfsu.edu  
 San Francisco State University  
 Security Level: Email, Account Authentication (Required)

## Signature

DocuSigned by:  
  
 7DB2FBA96666439...  
 Signature Adoption: Pre-selected Style  
 Using IP Address: 76.102.20.16

## Timestamp

Sent: 5/12/2020 1:35:17 PM  
 Viewed: 5/12/2020 7:59:36 PM  
 Signed: 5/12/2020 8:00:29 PM  
 Freeform Signing

**Electronic Record and Signature Disclosure:**  
 Not Offered via DocuSign

## In Person Signer Events

## Signature

## Timestamp

## Editor Delivery Events

## Status

## Timestamp

## Agent Delivery Events

## Status

## Timestamp

## Intermediary Delivery Events

## Status

## Timestamp

## Certified Delivery Events

## Status

## Timestamp

## Carbon Copy Events

## Status

## Timestamp

## Witness Events

## Signature

## Timestamp

## Notary Events

## Signature

## Timestamp

## Envelope Summary Events

## Status

## Timestamps

Envelope Sent	Hashed/Encrypted	5/12/2020 1:35:17 PM
Certified Delivered	Security Checked	5/12/2020 7:59:36 PM
Signing Complete	Security Checked	5/12/2020 8:00:29 PM
Completed	Security Checked	5/12/2020 8:00:29 PM

## Payment Events

## Status

## Timestamps