DIFFERENCES IN HEDGING IN L1 AND L2 ENGLISH ESSAYS ACROSS TWO GENRES

By

Jennifer Brooke

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Teaching English to Speakers of Other Languages—Master of Arts

2016

ABSTRACT

DIFFERENCES IN HEDGING IN L1 AND L2 ENGLISH ESSAYS ACROSS TWO GENRES

By

Jennifer Brooke

The ability to hedge, or qualify commitment to a claim, is an important aspect of academic writing because it allows writers to position themselves to their audience. Research indicates that L2 English writers struggle to hedge effectively, with studies such as Hyland and Milton (1997) and Hinkel (2005) demonstrating that they use less sophisticated hedges and a more limited range of hedges than L1 English writers do. This corpus study is composed of two parts. First, a methodological study was conducted with three expert raters examining the use of linguistic items traditionally considered hedges in sentential context. Two measures of rater's judgments are reported in relation to raw frequency of each item. The second part contrasts patterns of hedging across genre (timed versus untimed) and English nativeness (L1 versus L2 English writers). Results of the first section indicate significant differences in judged versus raw frequencies. Results of the second section indicate significant differences for some hedging devices between genres and between native speakers (NSs) and non-native speakers (NNSs). Implications are given for data collection, pedagogy, and assessment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This thesis is divided into two related studies: a methodological study involving frequency of linguistic items called *hedges*, and an investigation into differential hedging patterns under the conditions of genre (timed/untimed essays) and English nativeness (L1/L2 English, used interchangeably with NS/NNS in this study). A definition of hedging and a review of hedging research will provide a helpful backdrop for these two studies.

The term *hedging* describes the qualification of a proposition through words and phrases, such as *seem* and *it may be that*, which can also serve to convey doubt or uncertainty. A hedge's basic function is to mitigate a writer's commitment to a claim. The term was coined by G. Lakoff (1973), who defined it as a word "whose job it is to make things fuzzier or less fuzzy" (p. 471); clearly this definition leaves much to be desired, as *boosters* (words that intensify a writer's commitment to a claim, normally considered a kind of "opposite" of hedges) would fall into the latter category as words that can "make things…less fuzzy." Boosters along with hedges are also included Hyland (1994) calls "epistemic modality," which is an umbrella term for a writer's display of confidence (or lack thereof) in the truth of his or her statements. Lakoff's definition is problematic because it conflates these two into a single construct, whereas Crompton (1997) defined a hedge as "an item of language which a speaker uses to explicitly qualify his/her lack of commitment to the truth of a proposition he/she utters" (p. 281). For this reason, Crompton's definition is the one used in the current study.

Hedging is ubiquitous in both academic and everyday language. As Skelton (1988) notes, "it is by the hedging system of a language that a user distinguishes between what s/he says and what s/he thinks about what s/he says" (p. 38). Though hedging can be found in both spoken and written language, this study focuses on written argumentative essays. Persuasive or argumentative genres should elicit more hedging than other genres, as making claims and stating opinions are inherent in formulating an argument. Hedging is one technique proficient writers are able to use to situate these claims in a discourse community (in the case of discipline-specific academic writing) or acknowledge their readers' possible differences in opinion. Failure to hedge when making strong claims may cause readers to perceive the

writer as unreasonable or as presenting untenable arguments. Thus, hedges have a role to play in the interaction between the writer and the reader.

To this end, hedges are important linguistic items for second language writing researchers to study because they are markers of a writers' audience awareness and their individual *voice*. Voice is gaining wider recognition as an important feature of first language (L1) and second language (L2) writing. It has found a place on many writing rubrics, although as Zhao (2012) states, voice seems to be rather "impressionistically assessed in practice" (p. 202). Her study represents an attempt to reliably measure voice in written argumentative texts. She draws on Hyland's (2008) interactional model of voice, which characterizes hedging as part of *stance*. Through voice and stance, the ability to hedge effectively is crucial for writers to be characterized proficient and mature, and to have their claims accepted by their readers.

There is some evidence as well to believe that hedging or voice strength in general can indicate overall essay quality. For example, Zhao and Llosa (2008) found a significant relationship between voice and quality of writing in L1 English essays. Most recently, Yoon (2016) has developed a computerized model of voice that showed a significant positive correlation between hedges and overall writing quality in L2 English timed persuasive essays. Although this is only a correlation and not cause and effect, this indicates that hedging can contribute to a high-quality essay. Not every study has found a significant relationship between voice and essay quality (e.g. Helms-Park & Stapleton, 2003), but it is clear that hedging has a role to play in expression of a writer's voice and perhaps also in the wider quality of the text as a whole.

Unfortunately, hedging in any language is much more difficult to learn for NNS writers than for NS writers. It follows logically that ESL students struggle to hedge effectively. Grammatical, pragmatic, and context-appropriate use of hedging devices may be implicitly or explicitly learned by L1 English writers, but L2 English writers may need more explicit instruction. Zhao and Llosa (2008) concluded that voice may play different roles in L1 and L2 academic writing, based on their replication of Helms-Park and Stapleton (2003). They found a correlation between voice and quality for L1 English argumentative

essays, whereas Helms-Park and Stapleton (2003) found no significant relationship between the two for L2 English essays. The question remains of how hedging specifically might function in L1 versus L2 English writing.

Examining hedging patterns in argumentative texts written by NSs and NNSs of English, as the current study does, extends this line of inquiry. In addition, it should be noted that most of the studies reviewed so far have used timed writing in their explorations of voice, quality, and hedging frequency. It is also possible that hedging may prove to play a different role or be expressed to a different extent in different written genres (i.e. untimed writing versus timed writing). This will be another area of investigation for the current study.

Interest in hedging saw its heyday in the 1990s, attractive partly because of its overlap with many other concepts of applied linguistics and communication sciences. Under the auspices of discourse analysis, discipline-specific variation in hedging has been one area of research (e.g. Hyland, 2000; Salager-Meyer, 1994). Researchers studying speech acts and pragmatics have also found hedging to be a useful area of investigation, as hedges are crucial in performing speech acts such as requesting and disagreeing. Lexical hedges such as *appear* and *suggest* can be thought of as "speech act verbs" because they are performative rather than simply descriptive. In other words, hedges allow the writer or speaker to "act upon" a truth by mitigating his or her commitment to it. In this sense, they can function as pragmatic features because of their ability to mitigate face-threatening acts (as discussed in Meyer, 1997).

Lee and Park (2011) define face-threatening acts as any act that "violate[s] or fail[s] to satisfy positive or negative face concerns" (p. 129). The concept of negative face, first developed by Brown and Levinson (1978), refers to the individuals' desire for personal autonomy. In the framework of writing, face and face-threatening acts relate to the negotiated relationship between the reader and writer. When a writer makes a claim that a reader may disagree with, he or she threatens the reader's desire for personal autonomy, or negative face. Hyland (1994) describes hedging as one way in which writers can address the negative face of their readers by expressing respect for the readers' right to hold alternate opinions about the truth of any given proposition.

4

The construct of hedging is useful to researchers from many different fields and perspectives; however, one challenge is in the difficulty of defining hedging and classifying linguistic items that may function as hedges. As previously mentioned, early definitions of hedging were less than helpful, and now definitions seem to be left to researchers' intuitions. Most importantly for the field, no definitive list of hedging devices has been established. Various researchers have developed their own lists of hedges (e.g. Salager-Meyer, 1994; Hinkel, 2005); some words overlap but few taxonomies are identical, limiting the field's ability to replicate or properly interpret results. Crompton (1997) summarizes the issue nicely: "Hedging cannot, unfortunately, be pinned down and labeled as a closed set of lexical items" (p. 281). This presents a serious problem.

A further issue is how to group items into categories when drawing conclusions about the frequency of different types of hedges in a text or set of texts. There seems to be some agreement that hedges should be classified functionally (Hyland, 2000; Crompton, 1997), but countless classifications have been put forward. Prince et. al (1982) distinguished between *shields* (e.g. *suspect*), wherein the speaker/writer herself is hedged, and *approximators* (e.g. *sort of*), wherein the proposition itself is hedged. The terminology itself is commonsense; however, in practice the difference between a writer's protection of herself and modification of a proposition can be indistinguishable. Another conceptualization comes from Skelton (1983), who makes a distinction between *qualification/comment* and *proposition,* where the former is evaluative and the latter is factual in nature. The difference in his classification is that the same word can fall into either category depending on the context. While the spirit of examining items in context is in line with the direction of the current study, the subjectivity of this classification makes it difficult to implement in practice.

Another classification system to note is that of Hinkel (2005), whose study inspired the current research. Her classifications are a mix of function and part of speech: 1) epistemic hedges (e.g. *potentially, probably*), which refer to the limitations of the writer's knowledge; 2) lexical hedges (e.g. *many, several*), which are similar to epistemic hedges but cannot modify phrases; 3) possibility hedges (e.g. *perhaps, hopefully*), which can also include probability; 4) downtoners (e.g. *at all, a bit*), which

function to delimit meaning and emotive implication of nouns, verbs, and adjectives; 5) assertive pronouns (e.g. *anybody, somebody*), which she argues modify noun phrases; and 6) adverbs of frequency (e.g. *daily, frequently*), whose vagueness makes them "ubiquitously function as hedges" (p. 39). While she supports her classifications from previous research, some of the groupings seem arbitrary. For example, *probably* is a classified as an "epistemic hedge" even though Hinkel states that probability is included in the category of "possibility."

Finally, Hyland (1994) classifies his hedging taxonomy by part of speech. The six classifications include modal verbs (e.g. *can, may, might*), lexical verbs (*seem, suggest, believe*), modal adverbs (*often, occasionally, a bit*), modal adjectives (*few, hardly, just*), modal nouns (*possibility, assumption, estimate*), and assertive pronouns (*any, some, something*), with modal adverbs covering the largest portion of the taxonomy (40 of 130 items). This is in part because adverbs represent a "catch-all" classification that includes items that do not seem to fit into other categories (e.g. *a bit*, *not a*).

Clearly, no classification is watertight, and different conceptualizations of hedging classifications among these three researchers provide evidence that we should use caution in interpreting results of previous research on hedging in L1 and L2 writing. Face-value frequency (i.e. reported frequency as returned by hand counting items or by a text retrieval program) is deceiving because different studies use different lists of items when citing the frequency of hedges in texts. Furthermore, the current study goes a step further by challenging the idea that lexical items traditionally considered hedges function as hedges in all or even most contexts.

These methodological issues, including developing a taxonomy of items that can be used as hedges and examining the way in which writers use these items, are the focus of the first study. This is a necessary preceding step to the second study regarding patterns of hedging use in NS/NNS and timed/untimed argumentative essays, which is the focus of the second study.

CHAPTER ONE

Literature on Hedging Research Methodology

Careful methodology is crucial to the field of second language acquisition and second language writing. This section describes 1) the usefulness of corpus research in exploring hedging, 2) the process used to develop a list of items traditionally considered hedges, and 3) the process of determining how these items function in context.

Corpora

A corpus approach becomes extremely useful when the goal is to extract occurrences of a linguistic feature in a large amount of texts. Corpus linguistics has come to the forefront of research and practice in the field of second language acquisition. Free online corpora are becoming ubiquitous, such as the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). Researchers have begun to develop lists such as the Academic Word List (Coxhead, 2000) and PHRASal Expressions List (Martinez & Schmitt, 2012) using a corpus approach. An increasing number of textbooks have also begun to use corpus-informed vocabulary lists, motivated by the numerous advantages of examining real-world language in large quantities. From a pedagogical perspective, corpus analysis provides evidence of what language really is, not as it "ought" or is imagined to be. These are advantages for researchers, too, as corpora represent a wealth of information not only about word or phrase frequency, but also empirical questions about discourse structure (Biber et al., 2007), author self-mention (Hyland, 2001), and a variety of linguistic features such as hedging. Corpora also have practical benefits: for one, large pools of data allow for more informed results with less researcher fatigue in the data extraction phase of a study.

Taxonomies

Developing a taxonomy of hedging devices is no easy task. However, previous research has provided evidence that some types of hedges are more commonly used than others. Using a similar framework of part-of-speech classification referenced earlier in Hyland (1994), Hyland and Milton (1997) found modal verbs and adverbial hedges to be used more commonly than other items in their corpora of timed argumentative essays written by Hong Kong and British high school students. Adjectives and nouns

were the least used group of hedges. In a corpus of scientific research articles, Hyland (1996) found lexical verbs to be the most common group, followed closely by adverbials, adjectives, and then modal verbs (though he notes that these results vary somewhat from the trend of general academic English because of the discipline). These previous findings indicate that some parts of speech can perform a hedging function better than others.

Several other factors are taken into consideration when developing a taxonomy. Researchers should consider the differences between spoken and written modalities to make decisions about what to include in hedging taxonomies, as we already know modality is important from the literature (e.g. Akinnaso, 1982; Biber, 1986). For example, *like* is often used as a hedge in informal spoken English, as in *We could, like, talk about it later.* But writers are unlikely to use *like* in this way. They commonly use it as something other than a hedge, as in *I like to take the bus because it's convenient.* In a related sense, the genre of the writing one plans to examine also plays a role. Hedges are used in written genres from text messages to published research, but register dictates the appropriateness of which to use where. To illustrate, *if you get my drift* is a multi-word hedge that might be used between friends sending instant messages, but the probability of it appearing in an academic abstract is low. This line of argumentation points to the development of different kinds of hedging taxonomies for different modalities and genres; the "one-size-fits-all" approach of a single list of items for any text simply does not work. To that end, this study seeks to develop a list of items commonly found in argumentative essays, versus other genres and modalities.

Function in Context

It has already been acknowledged that many linguistic items traditionally considered hedges are polysemous in nature. This makes calculating frequency of hedges in a corpus of texts extremely difficult. As mentioned in the introduction, face-value frequency of items considered hedges in a set of written texts is somewhat deceiving. In reality, the context of any particular lexical item is crucial to determining whether that item is functioning as a hedge or not. Many items traditionally considered hedges have more than one function, even modal verbs, which are usually considered hedges almost all the time (Hyland,

9

1996, 2000). For example, *could* can express ability (e.g. *He could see over her shoulder*), or qualify commitment (e.g. *You could be right*). Another hedging item with multiple grammatical functions is *rather.* It functions as hedge in *He was rather short* but not in *I'd like tea rather than coffee.* As Crompton (1997) points out, the polysemy of a hedging device such as *rather* means context must play a role in determining whether or not it is a hedge. Simply counting its raw instances in a set of texts glosses over this crucial difference in usage.

This becomes even more important when considering second language (L2) writers whose vocabulary is in the process of development, and who may have limited knowledge about the grammatical features, register, or subtle connotations of a word. Their possible misuse of lexical items classified as hedging devices adds even more difficulty to determination of hedging. Writers and readers may have differential ideas about the meaning and use of a particular hedging device. In sum, context is essential to any line of research that aims to explore writers' intentions. While a handful of studies have done the hard work of examining sentential context of hedging devices for appropriate use (e.g. Aull & Lancaster, 2014), many lack this element (e.g. Hinkel, 2005) or use it in a limited way (e.g. Hyland & Milton, 1997). A brief review of these three studies is given here.

Aull and Lancaster (2014) examined linguistic stance markers in a corpus of 4,000 argumentative essays by L1 and L2 English speakers who were first-year undergraduates, upper-level undergraduates, and published academics in the United States. This list included *approximative* hedges[1] (those "through which the writers intimate the extent or degree to which a proposition is true" (p. 160). They reviewed concordance lines for each of these words to "verify that each instance was working in the target functional capacity" (p. 159) and removed items that were not. We can infer that the authors made principled judgments about removing items, though this is not the focus of their study and they do not elaborate on the process.

---

[1]The list included *apparent(ly), approximately, essentially, essentially, evidently, generally, in general, in many cases, in many ways, in most cases, primarily, largely, mostly, often, relatively, roughly, somewhat, usually,* and *sometimes.*

Hyland and Milton (1997) contrasted frequency of epistemic items in two corpora of comparable school examination essays, written by high school students in Hong Kong and Great Britain. 50 sentences with each of the target items were randomly selected from each corpus and examined, but the authors give no indicated of occurrences being discarded, only identified for epistemic function (i.e. semantic classes of certainty, probability, possibility, usuality, and approximation). The results of epistemic function classification for each item were extrapolated to the rest of the corpus. This decision might have been motivated by the large corpus size of around one million words.

Hinkel (2005) investigated hedging and boosting in 745 essays written by NS and NNS of English for university placement tests in the United States. Hedging devices were counted by hand, but no judgments of usage were made. The author's focus was on range and sophistication of hedging devices and differential usage by different L1 background writers (e.g. Vietnamese versus Arabic), rather than how the words were used.

All three of these studies used a corpus approach, but treated their results in different ways. Aull and Lancaster (2014) and Hyland and Milton (1997) both examined items in context, the former to ensure appropriate function of items and the latter to classify items into semantic classes. Hinkel (2005), on the other hand, relied on raw frequencies from which she drew conclusions about range, sophistication, and frequency of hedges (and boosters) in a variety of L1 writers' essays. None of these studies examined all items in sentential context and discarded those not functioning as hedges (e.g. *rather* as a conjunction rather than as a modal adverb that hedges a proposition).

A final word should be said about the use of human raters in a corpus study. Although human raters have long been employed to assign scores to written texts, no study to my knowledge uses human judgments to classify the function of linguistic items. Pure frequency counts have limits to their usefulness, as we have already established. Grant & Ginther (2000), whose study used a linguistic feature analysis tagging program, examined hedges along with many other linguistic features in writing. They note that "due to the nature of L2 texts themselves, human interaction with the texts will remain a critical part of the analysis process" (p. 143). In a similar vein, Simpson-Vlach and Schmitt (2010) used not only

computer-generated frequency but also rater judgments of chunk fixedness, cohesion, and teaching worth as criteria for inclusion in their formulaic phrase list.

This study seeks empirical evidence of the importance of sentential context in determining hedging frequency through the following research question:

RQ1: What is the ratio of raw frequency to judged frequency of hedging devices in each corpus and for each classification?

Method

Corpora

Because of its theoretical and practical advantages, a corpus approach is used in this study. Four corpora were used. The first two (collectively the timed writing corpus) are composed of 111 NNS and 44 NS essays collected by Yoon and Polio (2014),[2] for a total of 49,274 words. NNS writers each wrote three essays on argumentative topics: 1) laptop use in class, 2) the need for a simpler or more complicated procedure of getting a visa, and 3) on- and off-campus housing options. NS writers each wrote one essay on the first topic. NNS participants were students enrolled in an intensive English program at a large Midwestern university, and NS participants were undergraduate students enrolled in language methodology classes at the same university. This corpus was selected because of the argumentative genre and accessibility to the researcher.

The second two corpora (collectively the untimed writing corpus) are taken from a subset of the Michigan Corpus of Upper-level Student Papers (MICUSP). MICUSP is a public-access collection of texts by NS and NNS senior undergraduates or graduate students in 16 disciplines. These texts are either accepted and ungraded (such as research proposals) or received an A in a university class, indicating they have been judged to be high-quality writing. The current study is limited to argumentative essays, which comprise 22% of the total MICUSP corpus (186 essays). Argumentative essays were tagged in MICUSP based on the following features: 1) thesis-driven, 2) evidence from outside sources included as support,

---

[2] Yoon & Polio (2014) looked at linguistic complexity in two genres, narrative and argumentative. Only the argumentative corpus they collected is used in this study.

and 3) possible generation of new idea for the field. To be classified as argumentative, the essay's rhetorical purpose was "construct[ing] a coherent argument and support[ing] it with evidence/examples" (Römer & O'Donnell, 2011, p. 170. See the full paper for a review of how essays in MICUSP were classified.)

Only 39 of the 186 argumentative essays in MICUSP were written by NNSs. I chose to balance the corpus by number and discipline, selecting 39 NS essays by matching the disciplines represented in the NNS corpus. The disciplines and number of essays in each can be found in Table 1. This resulted in a total aggregate corpus word count of 227,764 words. The disciplines of English and Sociology are heavily represented in this corpus. Information about each of the four corpora is summarized in Table 2.

Table 1

Number of MICUSP Essays Included in Analysis by Discipline

Discipline	NS	NNS
Biology	1	1
Education	1	1
English	10	8
History	1	1
Natural Resources & Environment	3	3
Nursing	2	4
Philosophy	4	4
Political Science	2	2
Psychology	5	5
Sociology	10	10
Total	39	39

Table 2

Corpus Characteristics

	Timed Corpus		Untimed Corpus		Aggregate NS/NNS Wordcount
	No. of essays	Word-count	No. of essays	Word-count	
NS Corpus	111	34,163	39	98,844	133,007
NNS Corpus	45	15,111	39	128,920	144,031
		Aggregate Timed/Untimed Wordcount		Total Corpora Wordcount	
		49,274	227,764	**554,076**	

Taxonomy

Hinkel (2005) developed a hedging taxonomy that I expanded to include items considered hedges by other researchers, with additions from Hyland (1994, 1996), Salager-Meyer (1994), Skelton (1998), and myself. The taxonomy used in this study was much more expansive than previous research, and every effort was made to make the list as exhaustive as possible. See the appendix for the full taxonomy of hedging devices used in this study. Its limits must be acknowledged, as it is impossible to include every item that could possibly be considered hedging.

This study used human raters to judge if an item in the hedging taxonomy was functioning as a hedge in each instance it occurred in the text. I consider this to be a crucial element of my study. Human raters can evaluate the way a linguistic item is being used, while concordancing software cannot. It is logical, therefore, that fewer items would be included in a judged frequency count, where "judged frequency" means the number of instances for each item that raters concluded were functioning as hedges. For example, a hedge such as *perhaps* may appear 100 times in a set of texts, giving it a raw frequency of 100. However, three human raters may judge only 80 instances of *perhaps* as hedges based on the way the writer is using it in context. Therefore, *perhaps* has an 80% rate of acceptance. A 20% difference between raw frequency and judged frequency indicates that *perhaps* functions as a hedge most of the time. But a

lower rate of acceptance would indicate the worthwhileness of examining the use of *perhaps* in context. Looking at each instance of each item in sentential context is extremely time-consuming. Therefore, to provide empirical evidence for the necessity of this process, it is important to demonstrate how different raw frequency of a hedge would be from judged frequency.

To facilitate comparison, items in the taxonomy were classified according to part of speech, following Hyland (1994). In light of the range of classification systems used in previous research, it seemed simplest to group hedges in this way. Classifying items within the taxonomy based on part of speech more clearly reveals patterns of hedging frequency and use than individual-item analysis would.

Procedure

Document files for essays in each corpus were converted to text files and misspellings were corrected by hand. AntConc, a free corpus-analysis software developed by Anthony (2014), was used to tag hedging devices from the taxonomy. This tagging software was also used by Aull and Lancaster (2014), a similar study that also extracted frequency counts from a corpus. Raw frequency for each item was recorded. Following and exceeding the precedent set by Hyland & Milton (1997), who examined 50 randomly selected sentential contexts for each of their hedging items, in the present study every sentential context was examined by three raters to determine if the item used in context was hedging or not.

Screenshots were taken of each item's sentential contexts (partial in most cases) and uploaded to an Internet database in order to make the data accessible to raters. Three native-English-speaking raters (one of which was the author) were recruited. All had experience with hedging and were provided with Crompton's (1997) definition of hedging: "an item of language which a speaker uses to explicitly qualify his/her lack of commitment to the truth of a proposition he/she utters" (p. 281). No explicit training was provided besides this definition and the procedure of recording their ratings. Over the course of 4-6 weeks, each rater examined each sentential context and gave a binary yes or no judgment of whether the item was an instance of hedging.

This dichotomous rating method may seem like a blunt instrument for describing a nuanced construct, but a fuller picture of the difficulty raters encountered in judging items dichotomously is

15

captured by assessing two measures of frequency: one which reports items all three raters agreed upon (conservative) and one which reports items two or more raters agreed upon (less conservative). These two measures were taken because of the moderate correlation between the three raters. In order to compensate, a less conservative measure (i.e. frequency of two or more raters agreeing that a given item is functioning as a hedge) provides a more forgiving perspective of judged frequency without ubiquitously accepting all items in the taxonomy as hedges in all instances.

Correlations were calculated between each rater for each sub-corpus. Raters correlated with one another from .334-.692 across the four sub-corpora. These moderate correlations underscore the fact that it is much more difficult than it seems to determine whether an item technically classified as a hedge is actually functioning in a hedging capacity or not. Rater reliability for each of the four sub-corpora was as follows: Timed NNS, α = .82; Timed NS, α = .85; Untimed NNS α = .67; Untimed NS α = .69. It is interesting to note that rater correlation and reliability was higher in the timed corpora than the untimed. After all the items had been rated, raters met to discuss their criteria for rating types and individual tokens for the Untimed NS corpus. Their discussion was recorded and excerpts transcribed to provide qualitative support for the ratings given in this study.

To answer the first research question, the total occurrences of hedging devices were calculated in the three measures of frequency (raw, conservative, and less conservative). The ratio of the less conservative measure of judged frequency to raw frequency for each classification of hedges was also calculated as a percentage.

Results

The number of hedging devices per 1,000 words for each of the three measures of frequency in each corpus is shown in Table 3 and Figure 1. The judged frequency counts are significantly lower than the raw frequency counts, as expected. The highest rate of hedging was found in the timed NNS corpus in each measure of frequency.

Table 3

Occurrences of Hedging per 1,000 Words in Each Corpus

Frequency Measure	Timed NS	Untimed NS	Timed NNS	Untimed NNS
Raw Frequency	15.81	44.39	91.26	24.81
Less Conservative (2+ Raters)	5.59	13.03	23.16	8.45
Conservative (3 raters)	4.04	7.14	14.96	4.65

Figure 1

Differences in Three Hedging Frequency Measures
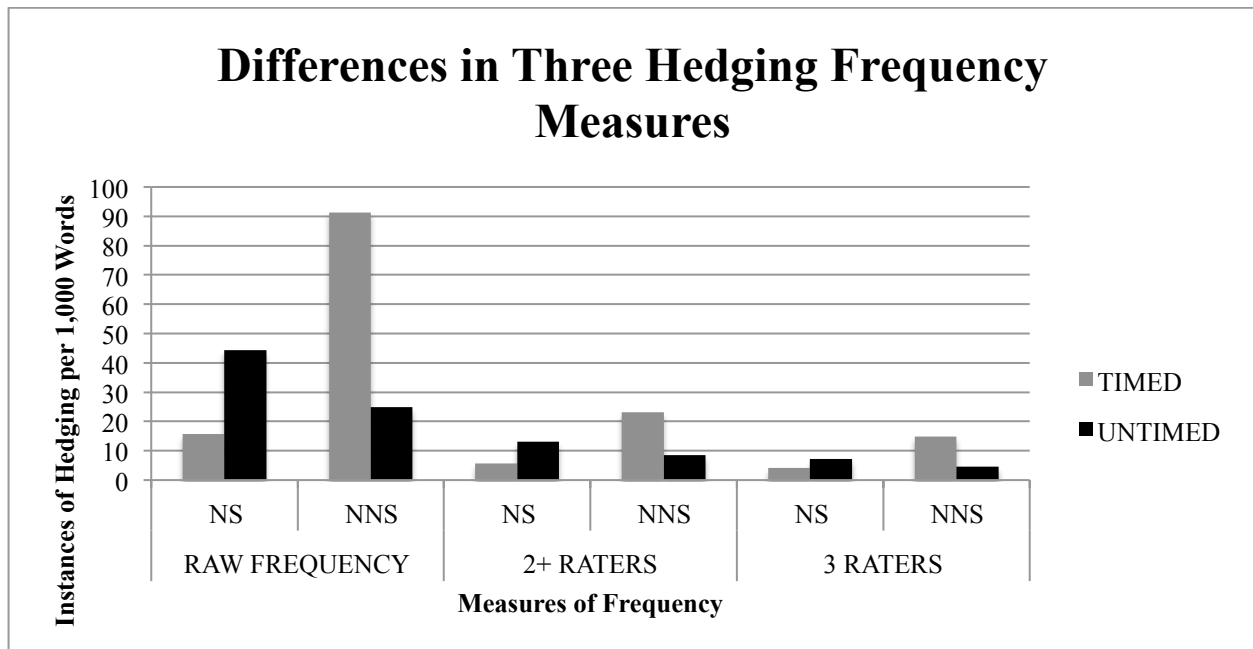
The percentage ratios of raw frequency to the two measures of judged frequencies broken down by hedging classification are shown in Table 4. The measure of frequency in the left-hand column includes only instances of each item that were judged to be hedging by all three raters (conservative). This measure has a significantly lower percentage than that given in the right-hand column, which includes instances of each item that were judged to be hedging by two or more raters (less conservative). The percentages in each column indicate the amount of items that were judged to be hedging by each measure as compared to the raw number of items appearing in the four corpora. For example, only 1.69% of items

17

classified as modal adjectives that appeared in the corpora were judged by all three raters to be functioning as hedges. This percentage increased to 10.84% of modal adjectives for the less conservative measure.

Table 4

Percentage of Judged Frequency of Hedging Devices by Group

Group	Conservative Measure (3 raters)	Less Conservative Measure (2 or more raters)
Assertive Pronouns	0.00%	2.08%
Modal Adjectives	1.69%	10.84%
Modal Nouns	4.83%	62.95%
Modal Adverbs	17.80%	32.55%
Lexical Verbs	51.65%	72.82%
Modal Verbs	57.54%	76.32%

In the conservative measure of judged frequency only two categories were judged to be hedging more than 25% of the time: lexical verbs, at 51.65%, and modal verbs, at 57.54%. At the less conservative measure of judged frequency two additional categories met this criterion: modal adverbs, at 32.55%, and modal nouns, at 62.95%. Lexical verbs jumped to 72.82% and modal verbs to 76.32% judged frequency in the less conservative measure. Neither assertive pronouns nor modal adjectives were rated above the 25% threshold for either measure.

Discussion

It is clear from Figure 1 that the differences between the raw frequency counts and judged frequency counts in each corpus are quite pronounced, especially for the timed NNS corpus.

The percentage of items that were judged by raters to be functioning as hedges in context was surprisingly low and extremely varied; for the conservative measure, judgments range from 0.00% (assertive pronouns) to 57.54% (modal verbs), and for the less conservative measure, from 2.08% (assertive pronouns) to 76.32% (modal verbs). It is interesting to note that in both measures of frequency, the same two groups (assertive pronouns and modal verbs) were rated the lowest and the highest. This

indicates rater agreement that assertive pronouns rarely, if ever, function as hedges, while modal verbs and lexical verbs function as hedges a fair proportion of the time.

In terms of the conservative measure, it is somewhat startling that so few instances of items from the taxonomy were considered hedges by all three raters. It is clear that the assertive pronoun group, at 0.00%, should be discarded from consideration as hedges. This category included *any, anybody, anyone, anything, some, somebody, someone,* and *something*. Raters found these words to be vague, indicating that the writer was not completely convinced of something, but not necessarily mitigating the truth of a proposition. The second lowest category, modal adjectives (e.g. *clear, apparent*), does marginally better at 1.69%. Modal nouns (e.g. *assumption, possibility*) and modal adverbs (e.g. *seldom, roughly*) fare only slightly better, at 4.83% and 17.80%. Taken together, these groups averaged a conservative judged frequency of 6.08%, which means that only one out of roughly 17 occurrences in the corpus was deemed hedging. At 51.65% and 57.54% respectively, lexical and modal verbs are the exception to this trend. Their high percentage of acceptance by all three raters (more than five out of every 10 instances) points to the fact that they may be the most clear-cut of all the hedging devices in the taxonomy. In contrast to the other categories, these percentages indicate that verbs such as *seem, tend, may,* and *might* have a consistently strong hedging function as compared to other parts of speech.

The less conservative measure paints a slightly brighter picture; only assertive pronouns and lexical adjectives fail to meet the 25% threshold. Lexical verbs increase to 72.82% and modal verbs to 76.32% with this measure. Modal adverbs turn out a stronger performance as well, at 32.55%. Modal nouns jump to from the conservative measure of 17.80% to 62.95%, indicating that two raters considered most modal nouns hedges, while the third did not. These results are in line with Hyland and Milton's (1997) previous findings that lexical verbs and modal verbs were the most frequently used category of hedges.

This highlights the importance of this less conservative measure for the implications of this study, as it accounts for differential rater severity and prevents one rater from overly skewing the results. The conservative measure, which only includes instances of items that all three raters considered to function

as hedges, is too susceptible to individual rater variation. For this reason, the less conservative measure of judged frequency will be used in Part 2 of this study.

As the low percentages reported in Table 1 are quite striking, the time taken to painstakingly examine each item in context seems to have been well worthwhile. These results point to the methodological flaw inherent in simply calculating raw frequencies, as this approach discounts the crucial factor of how the item is used. While rating, each rater kept notes about their criteria for making decisions about whether particular items were hedging. After completing their ratings, the three raters met to discuss the items together. Their comments and difficulties with the process of rating are summarized in the following section.

Quotations, Citations, and Paraphrases

I instructed raters to exclude any hedging device that appeared in a quotation or citation with the rationale that any hedge in these contexts was borrowed rather than being the writer's own word choice. Quotations and citations appeared only in the untimed corpus (MICUSP), as one of the criteria for inclusion in the argumentative essay section of MICUSP is that the essay refers to outside sources. That being said, the issue of paraphrase was unanticipated before rating began and seemed to be a grey area for the raters. At times it was possible to ascertain that a sentence was a paraphrase of a quote used in a different essay in the same discipline; in this case, should a hedging device within the paraphrase borrowed from a quotation be considered the writer's own words? The raters gave the writers the benefit of the doubt in these cases, judging the item as usual. However, the inability to determine the author's intent makes these cases very difficult to resolve.

Hedges and Boosters

In the wider L1 and L2 writing field, hedges are normally grouped with other linguistic features such as *boosters* or *intensifers*, which serve to mark certainty (words such as *very* and *always*). These two linguistic features are often studied in conjunction, as they are considered a kind of opposite. Hedges serve to mitigate a writer's commitment to a claim, while boosters serve to intensify it. That being said, rater comments in the present study indicate that hedges and boosters may not be so easy to distinguish.

Items such as *just, indeed, merely, much, may,* and *at least* were drawn from Hinkel (2005) and were included as hedges in the taxonomy used in this study. However, the raters agreed that these items almost always functioned as boosters.

Let us take the example of *just*. One issue is the polysemy of the word itself; instances where *just* means *fair* must be excluded from its judged frequency. The polypragmatic nature of *just* presents another hurdle. While differences in meaning are relatively easy to exclude, determining whether *just* is functioning as a hedge within sentential context proves much more difficult. To illustrate this, a screenshot of the first five instances of *just* in the Untimed NNS corpus as returned by AntConc is given in Figure 2.

Figure 2

Screenshot of Five Instances of *just* in the Untimed NNS Corpus

Hit	KWIC
1	oval of an invasive causes ecosystem disturbance, just like its introduction, and hasty action can
2	disturbances would leave the native species with just the adverse costs of thickened shells and
3	's biodiversity, but their removal can be just as hazardous. Even with the right goal,
4	a species is admirable. If people were just as aware of the value in their
5	life and the limited resources here. I just don't know what to do to

In Hits 1, 3, and 4, *just* allows the writer to compare two things with *like* or *as,* while in Hit 2 *just* seems to function similarly to *only*. In Hit 5 *just* seems to be closest to a hedge; however, based on the definition of hedging used in this study, "an item of language which a speaker uses to explicitly qualify his/her lack of commitment to the truth of a proposition he/she utters" (Crompton, 1997, p. 281), raters determined that *just* in this context was not functioning as a hedge. Although the writer is expressing his or her attitude toward a proposition, he or she does not seem to be qualifying commitment in any way. In fact, raters wondered if in a sense, the writer was attempting to intensify his or her lack of knowledge. As is clear from this discussion, some linguistic items considered hedges by some researchers might actually be functioning as intensifiers or boosters.

Factual Statements

Raters were generally skeptical of items from the taxonomy that were functioning to convey factual information or state a quantity. This especially applied to what Hinkel (2005) calls adverbs of

frequency, such as *annually, daily, monthly, yearly*, etc. For example, whether *some* conveys an inexact or

unknown amount or a writer's reluctance to name an exact amount is debatable. A screenshot of seven

instances of *some* in the Timed NS corpus as returned by AntConc is given in Figure 3 to illustrate this

point.

Figure 3

Screenshot of Seven Instances of *some* in the Timed NS Corpus

```
8              cannot learn the way you would like.   Some teachers may argue that students get distract    Jordan Hensl
9       s will promote increased engagement and learning.  Some students may end up off-task, but            Louise_Arg.t
10          provided with the same device. For example,   some laptops are much more up to date             Pumbaa_Arg.t
11         it more heavily outside of the classroom.   Some teachers stray away from this, while others      Regina Philc
12            the students. Sure, there will always be   some students who are so wrapped up in              Shalayia Yak
13      abilities can enhance the learning experience for   some students and they should be given that      Shalayia Yak
14                or in the classroom at all.        In   some cases, there are positive ways that a         Simba_Arg.tx
```

Again returning to Crompton's (1997) definition of hedging, the raters wondered if *some* truly

functioned to mitigate commitment to the truth of a proposition in these cases. *Some* in Hit 10, for

instance, seems to be more factual in nature than an expression of the writer's uncertainty. Two of three

raters judged *some* to be hedging in Hits 8 and 9, and one rater judged Hit 13 to be hedging. The raters

determined that *some* in the other instances in Figure X was not functioning as a hedge, but rather as part

of a factual statement.

Believe was another item where the consideration of factual statements came into play. When the

writer seemed to be mitigating her or his own claim, the raters considered this hedging. However, when

the writer was referring either to another party's belief, or to a religious or political belief, the raters did

not consider this hedging. A screenshot of five instances of *believe* in the Untimed NNS corpus as

returned by AntConc is given in Figure 4.

Figure 4

Screenshot of Three Instances of *believe* in the Untimed NNS Corpus

```
43              , there is no human nature as utilitarians   believe, or to be more precise, there is
44                    . I wonder why on earth he can   believe in a 'good' human nature.  In contrast,
45       without regard to Bourgeois civil society.    I   believe that modern bureaucracy is a key to
```
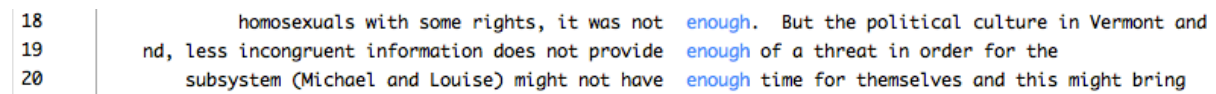
Hit 45 was considered a hedge by all three raters because in this case *believe* functions to modify

the writer's expression of his or her own opinion. However, in Hit 43 *believe* functions as in a factual

way, as part of a set of beliefs held by utilitarians; in Hit 44, raters judged *believe* to be an expression of opinion of parties other than the writer. Thus, both were excluded from the judged frequency count.

Enough will provide one more example of this issue with an added element of complexity: a negative statement (i.e. a sentence with *not*) tended to make an item more like a hedge than it would be in a positive statement (i.e. a sentence with no negation). A screenshot of three instances of *enough* in the Untimed NNS corpus as returned by AntConc is given in Figure 5.

Figure 5

Screenshot of Three Instances of *enough* in the Untimed NNS Corpus

```
18            homosexuals with some rights, it was not  enough.  But the political culture in Vermont and
19      nd, less incongruent information does not provide  enough of a threat in order for the
20          subsystem (Michael and Louise) might not have  enough time for themselves and this might bring
```

Hits 18 and 19 were determined to be hedging by two out of three raters, who expressed that they felt the negative context was crucial in the writer's qualification of commitment. However, though Hit 20 is also a negative context, *not having enough time* seems to be more factual (i.e. less of the writer's opinion and more of a concrete measurement) than *not having enough rights* (Hit 18) or *not providing enough of a threat* (Hit 19).

Returning to the discussion of Skelton's (1988) *proposition* (more factual in nature) and *qualification/comment* (more evaluative in nature), raters seemed to be in favor of classifying the latter as hedges more than the former. This is likely because Skelton's conceptualization of a hedge differed from Lakoff's definition used in the current study. In addition, Skelton was discussing hedges in the contexts of any genre of writing, not just argumentative essays, which is outside the scope of this paper.

Sentential Context

All three raters expressed the need for even more context than the sentence (or partial sentence, in most cases). The reason is that whether an item is truly functioning as a hedge depends on how the writer uses it in his or her overall argument. One example is given in Figure 6, a screenshot of one instance of *most* in the Timed NNS corpus as returned by AntConc.

Figure 6

Screenshot of One Instance of *most* in the Timed NNS Corpus

```
21      important details.  Lastly, |most of students who use laptops will not just
```

The topic of this argumentative essay was whether students should be able to use laptops in class, though it is difficult to determine which side the writer took as his or her stance based on this sentence fragment. In this case, is the writer attempting to diminish a counterargument or intensify his or her own claim? It is impossible tell without reading the entire essay.

Most was determined by the raters to function as a hedge only 2.22% of the time across all four corpora (less conservative measure). From the perspective of the raters, it depends on whether the writer is downtoning to *most* from *all* or is upgrading to *most* from *none.*

For example, consider the following versions of the sentence fragment in Hit 21:

All students who use laptops will not…

No students who use laptops will…

Students who use laptops will not…

Compared to these, *most [of] students who use laptops* seems to be a hedged proposition. However, consider two more versions of this sentence fragment:

Few students who use laptops…

A handful of students who use laptops…

From this perspective, *most [of] students who use laptops* seems to be an intensified proposition. This was the case in many instances of hedging devices across all four corpora. Raters would have benefited from being able to read the entire text of the essay to determine how writers were using hedging devices in the context of their argument. However, this approach is even more time-consuming that examining context at the sentence level, and diminishes the advantages of looking at many instances of a single item across corpora at the same time.

Summary of Rater Comments

To summarize, several interesting observations came from the rater's comments and discussion:

24

1) What is considered the writer's own words is subjective (i.e. paraphrase).

2) The distinction between hedge and booster is not clear-cut (e.g. *just*).

3) The same linguistic item can be used to convey a mitigated opinion or a fact (e.g. *believe*).

4) Sentential context is better than no context, but still does not provide enough information.

These four observations provide insight for future investigations into hedging or other related linguistic features, such as boosters. It should be meaningful that the raters in this study, who are all experts in second language acquisition, found relatively few items to be clear-cut hedges in all instances. The moderate correlations of rater agreement point to the vagueness of the hedging construct, as well as the difficulty of interpreting a given definition of it in sentential context. As Hyland (1996) states, "neither a purely formal treatment [of hedges] nor a detailed contextual analysis will always determine an unequivocal pragmatic function" (p. 479). The raters in the present study certainly found this to be true. Despite this, it is clear that the judged frequencies produced in this study provide a more accurate picture of hedging in these four corpora than a raw frequency count would. While this study has presented evidence for the weakness of the latter approach, the challenge remains of how to blend the strengths of a corpus approach and provide sufficient context for determination of hedging.

Limitations and Future Directions

One limitation of the current study is that classifying hedging devices grammatically is not without its problems; words such as *likely* can function as either adjectives or adverbs, and at times adjectives and adverbs were returned in the same search (e.g. *probabl** returned both *probable*, an adjective, and *probably*, an adverb). However, these limited interferences did not unduly influence the analysis. Secondly, as has been noted, the taxonomy of hedging devices, while extensive, is not an exhaustive list. In particular, raters commented that hedging at the phrase level was much more clear than at the individual word level. For example, the phrase *it appears to be the case* is clearly functioning as a hedge; whereas when breaking down the analysis at the word level there would be two separate hedges, *appear* (lexical) and *the case*. The question of what level of analysis at which frequency should be measured remains open.

25

Furthermore, more investigation is needed into the way lexical items in different phrases interact with each other in a sentence to qualify a writer's commitment to a claim. For example, *might* in Hit 20 of Figure 5 was considered a hedge by raters in its own right, though its relationship to *enough* a few words later in the sentence is unclear. This draws attention to the need for examining surrounding context of the item in question. For example, there is a vast difference between *more than enough* and *not enough*, which would be masked by a raw frequency count of *enough* in a corpus. Future studies should further explore the concept of multi-word hedging units. One step in the right direction has recently been undertaken by Yoon (2016), who developed an exhaustive list of phrasal markers of voice with grammatical constraints coded in the programming language Python. This kind of analytical approach is complementary to the human rater approach taken by the current study.

Despite these limitations, the current study has important implications for future research on hedging and other linguistic features like it (e.g. boosters). The way in which an item is used is just as (if not more) important than the number of times it occurs in a text. Polysemy and polypragmaticism are inherent features of many hedges, just as they are inherent features of language itself. Researchers need to conduct their investigations into these features with extreme care and a more rigorous methodology in order to make claims about their patterns of frequency in texts by writers who are NSs or NNSs of a language, in different disciplines, or at different proficiency levels.

CHAPTER TWO

Literature on L2 English Writers and Hedging

It is no secret that ESL students struggle to hedge effectively. Allison (1997) found his students' writing often lacked the nuance of making claims in an approachable way, and anecdotally many ESL writing teachers would support this. From an empirical perspective, it is interesting to note that research has not found a great difference in the number of hedging devices employed by NSs and NNSs in a corpus of essays. Rather, the difference has been found in the type of hedges used and their range. In Part 1 of this study, I looked at the methods several studies used in investigating hedging in NNS writing. Now, I turn to their findings.

Hinkel (2005) found that NNSs used slightly fewer hedging devices than NSs overall, though the difference was minimal. Her main conclusions were that NNSs tended to have a very limited number of hedging devices that they employed frequently, and these seemed to be less sophisticated than those used by the NSs. She defined sophistication in terms of whether the word or phrase was more prevalent in spoken versus written discourse (e.g. *almost* versus *fairly*)[3]. From this framework, NNSs used less formal and less sophisticated hedges, while NSs used more formal and more sophisticated hedges. Finally, this study found a difference in the classification of hedges used by the two groups; NNSs used more significantly more epistemic hedges (which refer to limitations of the writer's knowledge) and but fewer lexical hedges (which perform a similar function to epistemic hedges, but cannot modify an entire proposition) than NSs.

Hyland and Milton (1997) developed a list of items expressing epistemic modality, which includes both hedges and boosters. They found near equivalence in raw frequency of epistemic modifiers used in both the L1 and L2 English essays, about one device every 55 words. The differences between NSs and NNSs emerged in the range and type of devices used. The 10 most frequently used words from their taxonomy accounted for 75% of the total occurrences of words from the list in L2 essays. Furthermore, out of 75 words on their list, 30 appear 10 times or less, and nine do not appear at all in the

---

[3] As examples of informal/less sophisticated hedges, Hinkel (2005) gives *as at all, almost, at least, basically, (a) few, enough, hardly, just, (a) little, only, simply,* and *quite.* As examples of formal/more sophisticated hedges, she gives *fairly, mildly, partly, partially, scarcely,* and *virtually.*

L2 corpus. Though these figures conflate boosters and hedges, they point to the limited range of words

with which L2 English writers modify their commitment to a claim (either intensifying or hedging it).

In terms of grammatical distribution, results of Hyland and Milton (1997) indicate that NNSs

tended to rely more on modal verbs, while NSs used more adverbial hedges. The researchers categorized

their epistemic modifiers into classifications of *certainty, probability, possibility, usuality,* and

approximation (we will exclude *certainty* and *usuality* from our discussion as these refer to boosters).

While *possibility* and *approximation* hedges had approximately the same rate of occurrence between NSs

and NNSs, NSs used 73% more hedges expressing *probability* than NNSs.

Based on the findings of these two studies, NNSs do not differ dramatically when it comes to the

raw number of hedging devices they employ in their writing. Instead, the difference is in the type and

limited number of hedges they use. As mentioned before, the ability to hedge is an important one for

academic writers to develop. However, more hedging is not necessarily better than less hedging. Context,

audience, purpose, and genre are all factors in whether hedging is appropriate and what kind of hedge is

best suited. Developing this ability and sense of where and how to hedge is difficult for NNSs of English

for several reasons.

For one, the polysemy of hedging devices makes developing this skill especially problematic.

Though a word or phrase may be highly frequent in the input, its various usages may be difficult to

distinguish for a NNS. Exacerbating the problem, English for Academic Purposes (EAP) textbook

coverage of hedging is appallingly scant. Hyland's (1994) review of 22 post-beginning to advanced EAP

textbooks showed little discussion devoted to hedging devices, with the exception of modal verbs.

Though his analysis is nearly 20 years old, a review of EAP textbooks used at the Michigan State

University English Language Center conducted in 2015 revealed similar results, with modal verbs being

the most commonly covered, and almost no multi-word hedging devices introduced until a very advanced

level[4]. The fact that hedging is such an important aspect of academic writing yet receives so little attention from textbook developers calls for more research to fill the gap and provide evidence that ESL students are in need of training to develop their ability to hedge.

Besides the issues of polysemy and lack of textbook coverage, Hyland (2000) proposes two additional difficulties ESL students encounter in becoming proficient at hedging in English. First, variance in linguistic form of hedges. For example, *will* and *would* are forms of the same modal verb, but one has a hedging function and one does not. Negative transfer from the students' L1 presents another difficulty. This could be at the word-level or at the level of academic discourse in terms of hedging norms. In sum, hedging is a proficiency ESL students likely need explicit instruction for, and one that will increase in conjunction with students' overall language proficiency.

Literature on Genre: The Five-Paragraph Essay and Authentic Writing Tasks

Generally, corpus studies that examine features such as hedging in the argumentative writing of NNSs have used texts produced under time constraints (e.g. Grant & Ginther, 2000; Hinkel, 2005; Hyland & Milton, 1997). The very nature of empirical research makes timed writing an attractive data collection method because it offers a great deal of control over multiple variables (such as time on task). Participants in timed writing studies are normally instructed to write an essay about a given prompt with no access to dictionaries, outside sources, or help from others. Controlling for these variables allows researchers to draw more reliable conclusions about features of interest in the written product.

It is important to consider what genre timed writing elicits. In most cases, a timed essay will be structured according to the five-paragraph essay format: an introductory paragraph, three body paragraphs, and a concluding paragraph. As a contrived genre, the five-paragraph essay is argued to be useful for teaching students organization, argument structure, and coherence, often serving as a scaffold towards more authentic writing tasks. The five-paragraph essay format has come under attack as a formulaic structure that students struggle to leave behind as they advance into more authentic writing

---

[4] *Academic Writing for Graduate Students* by Swales and Feak was the only textbook provided as a resource to EAP teachers at the MSU ELC that contained any in-depth coverage of hedging, and it is not used as a classroom textbook.

tasks (i.e. longer, untimed essays). This genre has recently generated lively discussion among ESL professionals through blogs and online platforms (rather than through peer-reviewed journals). Blog post titles such as *Let's Bury the 5-Paragraph Essay: Long Live Authentic Writing* (Sztabnik, n.d.), *In Defense of the 5-Paragraph Essay* (Sheppard, 2016)*,* and *Why We Still Won't Teach the 5-Paragraph Essay* (Caplan & de Oliveira, 2016) demonstrate the extent to which this is a widely debated issue among writing teachers (of both L1 and L2 English students).

When preparing their students to take a timed writing exam (e.g. placement tests for Intensive English Programs or high-stakes assessments such as the TOEFL), ESL instructors often teach their students specific strategies. Efficiently dividing a short amount of time among planning, writing, and revising is a crucial skill for success in a timed writing context. In contrast, for untimed writing, instruction focuses on the recursive steps of the writing process, which includes receiving feedback and producing multiple drafts. Despite this, ESL students taught writing in a process-based (untimed) pedagogy are often assessed in a product-based (timed) way. Process-based writing pedagogy and assessment has increased in popularity, as Porto (2001) and Walker and Pérez Ríu (2008) point out. Despite the fact that some research indicates little statistical difference in score between timed and untimed essays (Caudery, 1990), many stakeholders in the ESL writing profession maintain a negative perspective on the dissonance between process-based pedagogy and product-based assessment.

Though a full discussion of the merits and drawbacks of the five-paragraph essay is beyond the scope of this paper, the point is that timed writing more often than not elicits a five-paragraph essay structure, and that this genre differs from untimed authentic writing tasks in at least two crucial ways. The first difference of note is in topic. The prompts given in timed writing tasks are designed to be somewhat generic, because it is important that each examinee have enough background knowledge on the topic to write without outside sources. The topics of untimed essays (i.e. authentic writing tasks), on the other hand, are rarely generic; they require more extensive background knowledge, greater depth of discussion, and often topic-specific vocabulary. Writer investment in his or her claims is an issue here. A generic

31

five-paragraph essay topic, coupled with the inability to consult outside sources, is less likely to generate the same kind of attachment to claims that authentic writing tasks have the potential to.

The second major difference is in audience. As has already been discussed, hedging is considered part of Hyland's (2008) interactional model of voice. In short, hedging is one way writers can express awareness of their audience and interact with their readers. Hedging is only necessary in the context of readers, and more information about his or her readers gives a writer a better sense of how much to hedge his or her claims. Authentic writing is purposeful in terms of being written for a specific, anticipated audience. Timed writing creates an artificial, anonymous audience of raters. It is easy to see how writers may produce different hedging patterns when writing with these two different audiences in mind, even on the same topic.

Thus, I propose that timed writing tasks and authentic (i.e. untimed) writing tasks may be conceptualized as two different genres that make different demands upon their writers. If this is true, then *time* presents an important feature to consider when researching linguistic features in L1 or L2 writing. Collecting writing data under time conditions could mask patterns that would otherwise appear in authentic writing tasks. Despite the convenience and sometimes necessity of timed data collection in writing research, the possibility must be confronted that writers may perform differentially in terms of linguistic features under timed and untimed writing conditions. To examine the possibility of a timed writing effect on hedging, this study compares hedging in timed and untimed writing, from both NSs and NNSs. Using the less conservative judged frequency measure from Part 1 of this study, Part 2 explores the following research questions:

RQ2: What are the differences in judged frequency of hedging between timed and untimed essays?

RQ3: What are the differences in judged frequency of hedging between NS and NNS essays?

Method

Materials

In order to answer research questions 2 and 3, the data was aggregated into a timed/untimed writing corpus, and a NS/NNS writing corpus. A large number of items were eliminated from the corpora according to the criteria described in the next section.

Procedure

Several criteria were set to eliminate low-frequency items from the analysis and ensure a balanced comparison between conditions: 1) Any item that did not appear in all four corpora, 2) Any item that appeared less than 10 times in any combined corpus, and 3) Any item that received less than 25% percentage of rater acceptance. Some of these items had a high proportion of positive rater judgment (i.e. two or more raters agreed that the item was hedging in the contexts in which it occurred in the corpora) but were not frequent enough to warrant inclusion in the analysis. These criteria eliminated a large number of items, leaving only items raters considered hedges at a high rate. The items included in the analysis can be seen in Table 5.

Table 5

Items from the Taxonomy Meeting Criteria for Inclusion in LL Analysis

Timed/Untimed Aggregate Corpus	NS/NNS Aggregate Corpus
believe	*appear*
seem	*believe*
few	*seem*
little	*tend*
often	*little*
possible	*often*
potentially	*possible*
probable(ly)	*almost*
sometimes	*likely*
could	*potentially*
may	*sometimes*
might	*could*
would	*may*
	might
	would

To answer the research questions, the number of hedges per 1,000 words in each corpus was calculated with the less conservative judged frequency measure described in Part 1. To further explore differences at the item-level, the corpora were aggregated based on genre (timed/untimed) and nativeness (NS/NNS). Log-likelihood ratios were calculated for each item in the combined corpora that met the criteria. Log-likelihood statistics are useful for comparing frequencies of linguistic items between two corpora and ascertaining statistically significant differences in frequency (Oakes, 1998).

Results

Overall Frequency

The number of hedges per 1,000 words in each corpus as measured by two or more raters in agreement is shown in Figure 7. NNSs in the timed corpus used the most hedges overall, approximately 23 per 1,000 words. NSs in the untimed corpus follow, using approximately 13 per 1,000 words. NNSs in the untimed corpus and NSs in the timed corpus used the fewest hedges overall, approximately nine and six per 1,000 words respectively.

Figure 7

Hedging Occurrences in Four Corpora
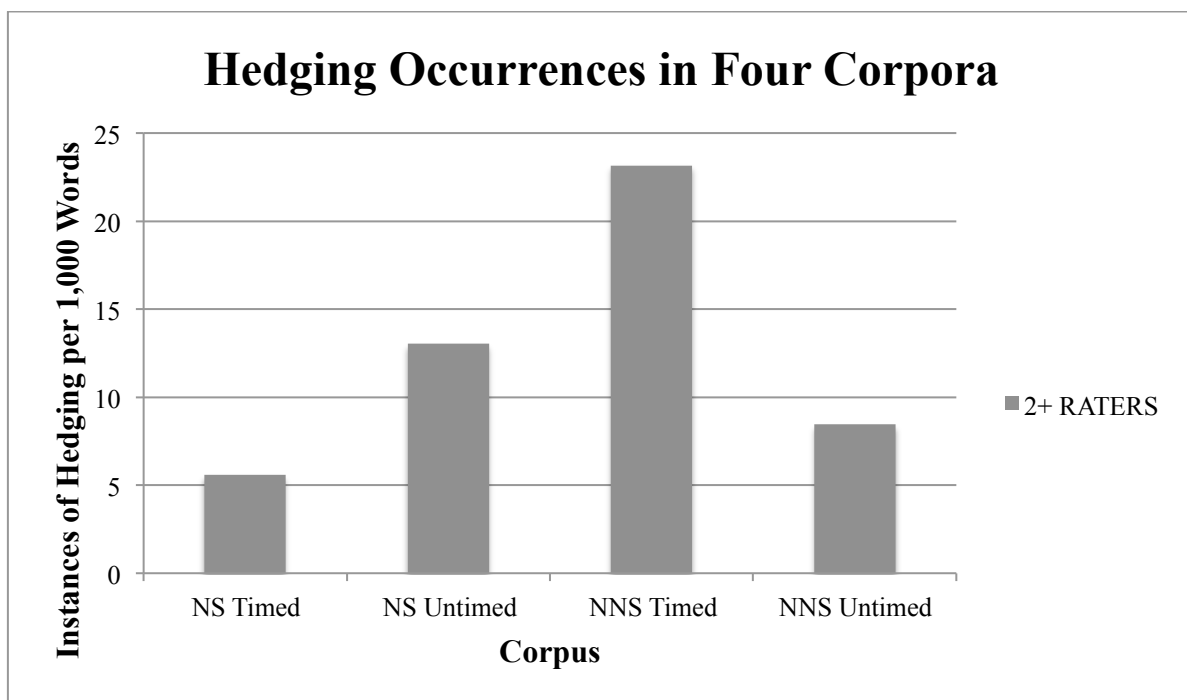
Item-Level Analysis

Log-likelihood ratios conducted on individual items the aggregate timed/untimed corpora indicated many significant differences in hedging between these conditions. Results are shown in Table 6.

Table 6

Log-likelihood Ratios of Judged Frequency in Timed and Untimed Corpora

| Item | Corpus | | LL |
	Timed (49,274 words)	Untimed (227,764 words)	
believe	**36**	28	47.57*
seem	11	**227**	37.78*
few	**6**	9	4.06*
little	4	32	1.23
often	17	103	1.14
possible	4	**26**	10.67*
potentially	**24**	26	23.83*
probably	**10**	17	5.60*
sometimes	**36**	27	48.86*
could	**80**	263	6.69*
may	**143**	293	56.87*
might	35	151	0.13
would	**40**	122	4.84*

Note. * p < .05. Bold numbers indicate higher frequency.

10 of 13 items that met the criteria for inclusion in the log-likelihood ratio analysis showed significant differences between the timed and untimed conditions: *believe, seem, few, possible, potentially, probable(ly), sometimes, could, might,* and *would*. Significant log-likelihood values for these aggregate corpora are somewhat higher than for the NS/NNS corpora, ranging from 4.06 to 48.86.

Log-likelihood ratios conducted on the aggregate NS/NNS corpora indicated some significant differences in hedging between these conditions, as shown in Table 7.

Table 7

Log-likelihood Ratios of Judged Frequency in NS and NNS Corpora

	Corpus		
Item	NS (144,031 words)	NNS (133,007 words)	LL
appear	7	**17**	5.13*
believe	26	38	3.32
seem	134	104	1.78
tend	26	25	0.02
few	**12**	3	5.09*
little	16	20	0.82
often	61	59	0.06
possible	37	39	0.33
almost	14	23	2.99
likely	42	43	0.23
potentially	9	**41**	24.80*
sometimes	29	34	0.90
could	118	**161**	10.52*
may	217	219	0.86
might	99	87	0.11
would	95	67	2.89

Note. * p < .05. Bold numbers indicate higher frequency.

A significant difference was found for four hedges: *appear, few, potentially,* and *could*. Although the other hedges in Table 6 met the criteria for inclusion, including appearing in all four corpora, appearing 10 or more times in each aggregate corpus, and being rated as hedges at a rate of 25% or higher, no significant differences were found. Significant log-likelihood values ranged from 5.09 to 24.80.

Discussion

A judged frequency count of hedges per 1,000 words showed that writers use different patterns of hedging based on both genre (i.e. time constraints) and nativeness. NNSs hedge in the timed condition

more than those writing authentic writing tasks or NSs writing either genre. One possible explanation for these results relates to the writers' ability to consult outside sources in the authentic writing task, which was not possible in the timed writing task. It is possible that the former group of writers felt less compelled to hedge their propositions because they were actually more convinced of their argument themselves and were able to support it with outside evidence. In contrast, writers in the timed corpus relied solely on their own knowledge and had no opportunity to support their argument with quotations or references. This may have caused them to hedge more because they were less committed to their own claims.

However, we see the opposite pattern for NSs, who hedged more overall in the untimed corpus than the timed corpus. An effect of topic is a possible explanation, one that was offered to explain results of hedging in Grant and Ginther (2000). While the NNSs wrote on three different topics for the timed writing task, the NSs wrote on only one, the use of laptops in class. Perhaps the writers had strong opinions on this topic and did not feel the need to hedge their claims. At the item level, the results of the log-likelihood ratios demonstrated significant differences in both sets of aggregate corpora between conditions, although differences were more pronounced in the timed/untimed corpora.

Timed and Untimed Essays

The hypothesis that time constraints impact hedging frequency is confirmed. One of the most interesting findings of this study is that writers in the timed corpus hedged more than those in the untimed corpus. Keeping in mind that the frequencies reported in this section are the less conservative measure of judged frequency found in Part 1, not raw frequencies, it seems that writers under timed conditions relied more on modal verbs *could, may,* and *would*, the lexical verb *believe*, modal adjective *few*, and modal adverbs *potentially, probably,* and *sometimes*. Writers in the untimed corpus used *seem* and *possible* more often. Some individual words present an interesting platform for discussion.

Believe. Believe is an interesting item to look at because it was used differently by writers in the two genres. With the less conservative measure of frequency, *believe* was used more in the untimed corpus. This is in line with the hypothesis that inability to consult outside sources caused writers under

37

time constraints to hedge more. Writers in the timed corpus used it with a limited function (i.e. only as a hedge).

In contrast, writers in the untimed corpus used *believe* with a greater range of functions, some of which were not judged to be hedging. These writers used *believe* more in the factual sense described in the Discussion of Part 1. Raters did not consider *belve* to be a hedge when it described religious or political beliefs, or the belief of another person who was not the writer. So although in terms of raw frequency, *believe* occurred significantly more often in the untimed corpus (144 times), it was judged to be hedging in the less conservative judged frequency measure only 19.44% of the time. *Believe* occurred just 39 times in the timed corpus, but was judged to be hedging in the less conservative judged frequency measure 92.31% of the time. It is clear from these results that the use of the item is indeed extremely important in considering patterns of hedging frequency.

Modal verbs. A second observation is that writers in the timed condition relied much more on the modal verbs *could, may*, and *would.* As with *believe,* the raw frequency of these three items was actually higher in the untimed corpus than the timed corpus, while the judged frequency showed the opposite result. Again, how the word is used is the consideration here. This pattern points to the fact that writers in the untimed condition used these three modal verbs with a greater range of functions. *May* was sometimes used in reference to the month of the year (though it was most often excluded because it was part of quotations or citations), as seen in Figure 8.

Figure 8

May Referring to Month of the Year in Two Instances in the Aggregate Untimed Corpus

```
33              ).  Much of the academic discussion following the  May 2004 FDA decision to deny OTC status for
34                        . that its product be available OTC in  May 2004 (Steinbrook, 2327).  The director of the
```

Could was excluded by raters when they ascertained it to be used in a sense of expressing ability, as demonstrated in Figure 9.

Figure 9

Could Expressing Ability in Two Instances in the Aggregate Untimed Corpus

```
31                    grade class, I began reading; maybe I │could figure out why.  After a closer reading,
32              things that happened in her country, but │could not openly express her feelings about what
```

Would was excluded when it was used in the perfect conditional construction (i.e. *would have +*

past participle) because in these instances the raters judged it was not functioning as a hedge. Only

23.69% of occurrences in the untimed corpora were judged by the less conservative measure to be hedges,

as compared to 68.97% of occurrences in the timed corpora. Examples of *would* used in the perfect

conditional can be seen in Figure 10.

Figure 10

Would in the Perfect Conditional in Two Instances in the Aggregate Untimed Corpus

```
13                     of Carlos. However, a truly dedicated teacher │would have pushed himself as hard as he
14                            the end either. If he did, he │would have been much more confident that he
```

Whereas the writers in the untimed corpus used these three modal verbs less as hedges than the

writers in the timed corpus did, these results indicate that the former have much greater control over a

range of functions of these words than the latter do.

Native Speakers and Non-Native Speakers

Only four words had statistically different frequencies in the NS or NNS corpus. NSs used the

lexical verb *appear,* modal adverb *potentially*, and modal verb *could* more than NNS, but used the modal

adjective *few* slightly less frequently than the NNSs. There does not seem to be a principled explanation

to account for these four differences, as each word comes from a different classification. Previous

research that found that NSs and NNSs tend to rely on different grammatical categories or classifications

of hedges in their writing was not supported by these results.

Lack of comparability between the current study and previous research is not completely

unsurprising, however, based on the evidence of the varied and discrepant approaches taken to defining,

classifying, and calculating frequency of hedges in NS and NNS writing. This study had extremely

stringent criteria for words in a given corpus to be counted as hedges and to be compared to hedges in

other corpora. More differences between NS and NNS hedging may emerge if the threshold for inclusion in the analysis were lowered. However, I maintain that the criteria set in this study were reasonable. It is invalid to compare frequency of a word in two corpora if the word occurs 9 or fewer times in one corpus, because of there is too large of a chance that rater fatigue or simply small sample size could influence the judged frequency of that word. Furthermore, this study ensured that items included in the analysis were functioning as hedges at least 25% of the time across the four corpora. Though excluding items below this threshold may have masked some of the more pronounced differences between the aggregate corpora, it seems more than reasonable to exclude items that function as something other than hedges up to 75% of the time.

Römer (2009) has explored divergence between NS and NNS writing in a different sense. She compared the phraseology of NS and NNS academic writing using a corpus approach, concluding that differences in n-grams and p-frames (n-grams that differ by one word in identical position) were much more pronounced between novice and expert academic writers than between NSs and NNSs. She argues that NSs of English must learn academic writing the same way NNSs must, and in this sense a NS of English and a NNS of English in the same year in university are on equal footing.

Though Römer takes a different approach to comparing NS and NNS writing, her conclusions seem to be borne out in the current study as well. More differences were found between corpora when proficiency was a moderating variable (as a function of the NNS students in the timed corpus versus the NNS students in the untimed corpus) than when English nativeness was a moderating variable. Perhaps this conclusion points us, as a field, to investigate other intervening variables besides nativeness, such as proficiency and time constraints on writing.

Implications for Pedagogy and Assessment

Hedging, as part of voice, has been shown to contribute to overall writing quality in both L1 (Zhao & Llosa, 2008) and L2 English (Yoon, 2016). However, it is important to note that more hedging is not necessarily always better. Overhedging can be just as problematic as not hedging enough. The amount of hedging appropriate for a piece of writing is dependent on the writer's background knowledge, his or

her audience, the genre, and the academic discipline. Hyland (1996) demonstrated that scientific disciplines, for example, tend to hedge much more than other academic disciplines on average.

From a pedagogical perspective, the role of the ESL instructor is to equip students with linguistic and pragmatic knowledge of hedging devices. Developing expertise in his subject area and intuition about how much to hedge any given proposition is up to the student himself. For teachers wondering how to help their students in developing hedging proficiency, many researchers have offered ideas and pedagogical strategies for teaching students how to hedge effectively.

Skelton (1988) recommends what he calls *sensitization* exercises, which involve asking students to rank a text from 1 to 10 based on the appropriate use of hedges. He is also a proponent of having students rewrite passages (perhaps those used in the aforementioned exercises) to make them more or less doubtful. Hyland (1996) suggests teaching students to use concordancing software to identify hedges in academic writing or in their particular disciplines. He also recommends having students work with texts (both short and long), doing consciousness-raising activities such as the following: 1) distinguish factual verses opinion statements; 2) remove hedges and discuss the effect on the text; 3) rank hedging devices on a scale of certainty; and 4) translate hedges into their L1 and compare the meaning. In terms of teaching students to use hedges productively, Hyland advises teachers to develop students' sense of audience (perhaps by having them rewrite a text for a different audience) and focus on high frequency items. In a similar vein, Holmes (1982) recommends that lower proficiency ESL students focus on mastering a single classification of hedges.

While this last piece of advice is sound for low-level learners, teachers should also strive to challenge learners as they progress in their English proficiency with more sophisticated hedges, a greater variety of hedges, and more multi-word hedges. A concluding thought is that NNS seem to approximate NS writers in their use of modal verb hedges (with the exception of *could*); this is logical, as ESL students receive the most exposure to modal verbs in EAP textbooks (Hyland, 1996), and since modals have other functions besides hedging they likely are highly frequent in a learner's input. ESL students need to encounter and learn to use more sophisticated modals like *suggest* and *seem* and phrases like *it may be*

41

that and *it can be assumed* in order to more closely approximate native speaker usage. Teachers as well as textbook developers should take note of this.

In terms of implications for assessment, the main finding of this study was that essays written in a timed environment used more hedging than untimed essays. Assessors of timed writing tasks should expect this from both NS and NNSs of English.

Limitations and Future Directions

This data can provide some interesting insights for the effect of time constraints writers' hedging frequency. However, it is important to keep in mind that this is cross-sectional data; therefore, we cannot draw conclusions about how the hedging frequency of an individual writer would change under time conditions. In addition, language proficiency may play a role here, as essays written by NNSs in the timed corpus had been in the United States for an average of only 17.81 months, while essays written by NNSs in the untimed corpus are senior undergraduates or graduate students writing high-quality papers. We can assume the NNSs with essays in the untimed corpus are of higher proficiency level and likely have greater linguistic and pragmatic resources than those of a lower proficiency level. Thus genre differences and proficiency differences are conflated. This did not present a problem in the NS/NNS aggregated corpora, however, as less proficient NNS and more proficient NNS writers are grouped together.

In addition, as discussed previously, any hedging device in a quotation or citation was automatically discarded the judged frequency count. These items were included in the raw frequency count, however, which unfortunately acts as a penalty against writers whose essays contained many hedging devices in quotations and citations but not as many outside of them. This may also be one reason the rate of hedging in the untimed corpus is somewhat lower than expected. Finally, topic differences may have contributed to eliciting different patterns of hedging as well. Future research could explore differences in hedging frequencies (raw or judged) between essays by the same set of writers on two different topics.

APPENDIX

Table 8

Taxonomy of Hedging Devices and Classifications

Classification	Item
assertive pronoun	any
	anybody
	anyone
	anything
	some
	somebody
	someone
	something
lexical verb	appear
	assume
	believe
	claim
	doubt
	in my view
	propose
	report
	seem
	speculate
	suggest
	suspect
	tend
modal adjective	about
	apparent
	clear
	enough
	essential
	few
	hardly
	just
	kind of
	like
	little
	many
	more
	most
	much
	often
	possible
	potential
	rare

Table 8 (cont'd)

	several
	sort of
	unlikely
modal adverb	a bit
	a good/great deal
	according to
	actually
	all but
	almost
	annually
	apparently
	approximate(ly)
	as good as
	as well as
	at all
	at least
	barely
	basically
	broad(ly)
	by some/any chance
	clearly
	comparative(ly)
	daily
	dead (+ adj)
	essentially
	evidence/evidently
	fairly
	frequently
	hopefully
	if you catch my meaning
	if you know what I mean (to say)
	if you understand what I mean
	in (the) case (of)
	in a way
	in the least/slightest
	indeed
	indicate
	likely
	maybe
	merely
	mildly
	monthly

Table 8 (cont'd)

	more or less
	nearly
	normal(ly)
	not a
	occasionally
	oftentimes
	only
	partially
	particularly
	partly
	per day/hour/year
	perhaps
	possibly
	potentially
	practically
	presumably
	pretty (+ adj)
	probable(ly)
	quite (+ adj)
	rarely
	rather
	regularly
	relatively
	roughly
	scarcely
	seldom
	simply
	slightly
	somehow
	sometimes
	somewhat
	sporadically
	sufficiently
	surprisingly
	theoretically
	to our knowledge
	truly
	unexpectedly
	usually
	virtually
	weekly
modal noun	assumption

Table 8 (cont'd)

	estimate (n.)
	possibility
	something like
modal verb	could
	may
	might
	would

REFERENCES

REFERENCES

Akinnaso, F. N. (1982). On the differences between spoken and written language. *Language and Speech, 25*(2), 97-125.

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

Aull, L., & Lancaster, Z. (2014). Linguistic markers of stance in early and advanced academic writing: A corpus-based comparison. *Written Communication, 31*(2), 151-183.

Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language, 62*(2), 384-414.

Biber, D., Connor, U., & Upton, T. (2007). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins.

Brown, P., & Levinson, S. (1978). Universals in language usage: Politeness phenomena. In E. N. goody (Ed.), *Questions and Politeness* (pp. 56-289). Cambridge: Cambridge.

Caplan, N., & de Oliveira, L.C. (2016, 12 February). Why we still won't teach the 5-paragraph essay [Web log post]. Retrieved from http://blog.tesol.org/why-we-still-wont-teach-the-5-paragraph-essay/#more-7487

Caudery, T. (1990). The validity of timed essay tests in the assessment of writing skills. *ELT Journal, 44*(2), 122-131.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. *English for Specific Purposes, 16*(4), 271-287.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9*(2), 123-145.

Helms-Park, R., & Stapleton, P. (2003). Questioning the importance of individualized voice in undergraduate L2 argumentative writing: An empirical study with pedagogical implications. *Journal of Second Language Writing, 12*(3), 245-265.

Hinkel, E. (2005). Hedging, inflating, and persuading. *Applied Language Learning, 15*(1&2), 29-53.

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

Hyland, K. (1994.) Hedging in Academic Writing and EAP Textbooks. *English for Academic Purposes*, *13*(3), 239-256.

Hyland, K. (1996). Nurturing hedges in the ESP curriculum. *System, 24*(4), 477-490.

49

Hyland, K. (2000). "It might be suggested that...": Academic hedging and student writing. *Australian Review of Applied Linguistics*, *16,* 83-97.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*(3), 207-226.

Hyland, K. (2008). Disciplinary voices: Interactions in research writing. *English Text Construction*, *1*(1), 5–22.

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Second Language Writing, 6*(2), 183-205.

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2,* 258-508.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299-320.

Meyer, P. G. (1997). Hedging strategies in written academic discourse: Strengthening the argument by weakening the claim. In R. Markkanen & H. Schröder, *Hedging and Discourse: Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (pp. 21-21). New York: de Gruyter.

Porto, M. (2001). Cooperative writing response groups and self-evaluation. *ELT Journal, 55*(1): 38-46.

Römer, U. (2009). English in academia: Does nativeness matter? *Anglistik: International Journal of English Studies, 20*(2), 89-100.

Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): The design, compilation and genre classification of the Michigan corpus of upper-level student papers (MICUSP). *Corpora, 6*(2), 159-177.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes, 13*(2), 149-170.

Sheppard, R. (2016, January 4). In defense of the 5-paragraph essay [Web log post]. Retrieved from http://blog.tesol.org/in-defense-of-the-5-paragraph-essay/

Sztabnik, B. (n.d.). Let's bury the 5-paragraph essay: Long live authentic writing [Web log post]. Retrieved from http://talkswithteachers.com/5paragraphessayvsauthenticwriting/

Walker, R., and Pérez Ríu, C. (2008). Coherence in the assessment of writing skills. *ELT Journal, 62*(1), 18-28.

Yoon, H. J., & Polio, C. (2014). A longitudinal study of linguistic complexity in two genres. Paper presented at the Colloquium on Cross-Linguistic aspects of linguistic complexity in second language research. Vrije Universiteit, Brussels, 19-12-2014.

Yoon, H. J. (2016). Automated assessment of authorial voice in written discourse. Paper presented at the Second Language Studies Symposium. Michigan State University, East Lansing, MI, 27-2-2016.

Zhao, C. G. (2013). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing, 30*(2), 201-230.

Zhao, C. G, & Llosa, L. (2006). Voice in high-stakes L1 academic writing assessment: implications for L2 writing instruction. *Assessing Writing: An International Journal, 13*(3), 153-170.