The Comparison of Lexical Hedge Uses in Academic Articles between Thai College Students and International Writers

Kochamon Thongsoy[1]
Prach Aowsakorn[2]
Received date 24 December 2021
Revised date 4 April 2022
Published date 21 April 2022

Abstract

This comparison study investigates the numbers and frequencies of lexical hedging types between two corpora: the first corpus was the academic articles written by Thai college students majoring in English, and the other was the published articles from "Science Direct", the database widely accepted by scholars in linguistic fields. The use of hedges and the frequencies in the corpora were identified, quantified and compared based on Hyland's (1998) and Schmied (2008) classification of hedges. The results show that all types of lexical hedges and modal auxiliaries were found most frequently in both corpora. However, the student writers used lexical hedges more often in their texts when compared with the professional researchers who published their articles in international journals even though the choices of words used are less than those found in published articles. The underlying reasons why the students used those two types most frequently can be partly from lessons in previous courses in which quantifiers and modal auxiliaries were emphasized.

Key words: hedges, hedging devices, lexical hedges

[1] Lecturer, School of Liberal Arts, Mae Fah Luang University, kochamon.tho@mfu.ac.th

[2] Lecturer, School of Liberal Arts, Mae Fah Luang University, prach.aow@mfu.ac.th

Introduction

The term *hedge* was first introduced by Lakoff (1972) to refer to words or phrases, such as *sort of* and *kind of*, that "make things more or less fuzzy" (p. 471). Later on, the use of hedges has gained more attention as it is related to other theories of language uses. For example, in Brown and Levinsons' (1987) politeness theory, hedges are used as a politeness strategy when a speaker has to perform a face-threatening act (FTA) in order to minimize the threat that might damage either the speaker's or the listener's face; in addition, they are also used to soften criticism (Reikkinen, 2009). In addition, some writers used hedges to create interaction with the readers because they "indicate the writer's decision to withhold a complete commitment to a proposition, allowing information to be presented as an opinion rather than accredited fact" (Hyland, 2005b, p. 178), their texts could rather be considered as persuasive than assertive to the readers.

However, the definitions of hedges in previous studies are somewhat different in details. Hence, in this study, the definition of hedges by Hyland (1996) is used as the operational definition which identifies lexical hedge uses as "any linguistic means used to indicate either a) a lack of complete commitment to the truth of a proposition or b) a desire not to express that commitment categorically" (p. 3).

Under this definition, hedges can appear in various forms. To illustrate, writers can use strategic hedges, expressions to soften disagreements two opposite ideas—with expressions like "one cannot know that …" or they can just use lexical hedges, words to reduce the assertive tone or too strong claims such as modal verbs like *may* and *might*, epistemic verbs like *indicate* and *appear*, epistemic adjectives such as *possible*, and epistemic adverbs such as *apparently* (Hyland, 1998). In this study, the focus was on only "lexical hedges" which are more explicit and can be identified by general corpus analysis software.

In an academic context which the accuracy of the information is also highly valued, writers have to express their ideas or information as accurately as they can in order to make their ideas more credible—not easy to argue. For this reason, they sometimes add hedge words or use some hedging strategies in order to make their claims more flexible—adding more chances to be right.

Objective of the Study

The objective of this study is to investigate the use of hedges in academic articles written by Thai undergraduate students majoring in English. In order to see if the students use hedges effectively in the academic writing genre, hedges found in their articles will be compared to those found in research articles published in international academic journals, which are considered of acceptable quality by scholars in the field. It is possible that the result can help us understand

the problems in student writing and shed light on the students' needs and what teachers can do to improve the writing skill.

Scopes of the Study

This study focused on only types and frequencies of lexical hedges used in academic articles written by 3rd year English majors at Mae Fah Luang University and those found in research articles published in international journals. Other types of hedges such as strategic hedges, epistemic verbs, epistemic adjectives and, epistemic adverbs which are not emphasized in the writing lessons of the course are not included.

Review of Related Literature

As the use of hedges has gained interest from many researchers around the world, a number of studies focusing on the use of hedges in many aspects and in many languages have been published nationally and internationally. Recently, in order to compare the use of lexical hedges, corpus analysis has also been used by many researchers. This can possibly be because it "makes it possible to examine trends across many examples of student writing in ways traditional reading cannot" (Aull, Bandarage, & Miller, 2017, p. 32). Corpus analysis can be used to investigate many linguistic features, and not only in one corpus, but also between two corpora or more. Among them are comparative studies between the uses of hedges in texts written for different types of journals, texts in different academic fields, texts written by English native speakers and ESL/EFL learners and even texts written in English and other languages.

1. Hedges in Different Text Types in the Same Field

Even in the same genre and the same field of study, academic writing in the medical field in this case, a previous study revealed that the number of hedges the writers use can also be affected by types of journals and their target audience. To illustrate, Schmied (2008) analyzed lexical hedges in two types of academic writing in medical fields, specialized academic texts from international Anglo-American journals and popular academic texts from a popular science magazine named *The New Scientist*. He claimed that these two types of texts, although in the same fields, have different target readers and hedges found in popular medical texts were almost three times more frequent when compared with those found in specialized academic texts.

2 Hedges in Different Academic Fields

In addition to studies of texts within the same field, previous studies also revealed that texts from different topics contain different types of hedging devices. This means even in academic contexts, different fields can also affect the number and the types of hedging device. To illustrate, Hyland (2005a) claimed that more than 70% of all hedges from 56 academic articles in different fields were found in the humanities/social science papers. This means the use of hedges in these

two fields was two times more frequent when compared with those found in academic articles in physics and engineering. He also claimed that "these findings reflect the fact that research articles express the different epistemological and social assumptions of disciplinary communities (p. 106).

3. Hedges in English Texts Written by Native Speakers and Non-native Speakers

First language and language proficiency of authors has also been considered as another factor that affects the number and the types of hedges in writing. This can be the reason why a number of researchers have been interested in this issue as well. For example, Vassileva (2001) examined Bulgarian authors' use of hedges in their L1 and L2 (English) scientific writing in comparison with native English speakers and found that the participants used hedges differently. This study revealed that Bulgarian writers seem to be more committed when writing in English because they used less hedges. The researcher claimed that it was possibly because the participants were not familiar with hedges when wiring in L2, but they were proficient enough to use hedges when writing in their first language.

The findings from Vassileva (2001) are also similar to those found in Atai and Sadr's (2006) study in which they investigated the use of hedging strategies in academic writing of English and Persian Native Speakers in English applied linguistics research articles. In this study, they found that English native speakers used a more variety of hedges to express their degree of commitment.

In the same way, He, Jiangqin, and Feng (2010) also investigated hedges used by native speakers of English and Chinese English learners in abstracts of academic papers and found that although they shared some similarities, the differences regarding the uses of hedges in both corpora are that the choices of hedges used by the native speakers were about two times of those used by Chinese English learners and that the Chinese writers seldom used nouns and adjectives as hedges.

4. Hedges in Texts in Different Languages

Different languages can also affect the use of hedges in texts as well. There have also been some studies that aimed at comparing texts in the same field in different languages. For example, Hu and Cao (2011) compared hedges in abstracts of applied linguistic articles in English- and Chinese-medium journals and found that abstracts published in English contained more hedges than those published in Chinese. Thus, from this study, it revealed a possibility that English language proficiency of the writer may not be the only reason why L2 students included more or less lexical hedges in their writing. Some other underlying reasons can be the writing lessons in class, i.e., whether or not the teachers focused on this feature before giving the assignment, or even differences between the cultures of those students themselves.

Because the results from previous studies seem not consistent and may not be generalized to Thai contexts, there is definitely a need for more studies. The purposes of this

study are to identify types and frequencies of hedging devices used in successful academic articles (those published in international journals) and those written by Thai college students and to compare the types and the frequency of lexical hedges in academic articles written by the students with articles published in academic journals, which are considered of acceptable quality.

In other words, the present study aims at answering these two research questions:

1) How many types of lexical hedges are used in academic articles written in the 3rd year English majors at Mae Fah Luang University and how often do they use each type?

2) What are the similarities and differences between the types and the frequency of lexical hedges used by the 3rd year English majors at MFU and those found in research articles published in international journals?

Research Methods

1. Data and Data Collection

The data in this study were collected from twenty academic articles written by 3rd year students in the Bachelor of Arts in English programme at Mae Fah Luang university. The students were registered to the course titled "Academic and Professional Writing" which was a mandatory course for 3rd year students in the programme. The task, which was an individual task, in the course was to write an approximately 5-page academic article. All of the articles written by the students are related to topics in English language teaching and learning. The data were from the students' first drafts for the assignment to ensure that they were written by the students themselves and they were not affected by the teacher's and their peers' comments.

For the published articles to be compared with, twenty articles published in academic journals were downloaded from the database "Science Direct," which was the database that the teacher suggested the students use for finding sources to be cited in their articles. The keywords used for searching was "English language teaching" and the results were limited to only those with full-texts available published within ten years back then (between 2008-2018). The data were from the first twenty articles shown as the searching results. With these conditions, these articles might have been the real sources that the students read and deemed their model papers.

It should be noted that the first languages of writers of the published articles in this study are not taken into account. Since all articles were published in international refereed journals, they were used as those with acceptable quality regardless of the first languages of the writers. It can even be assumed that all of them were revised and edited, possibly by more than one person. In addition, tracing the identities of people who were involved in editing the language used in these articles is impossible.

All the articles, excluding reference lists and appendix, from both the students and online database were converted into text files in order to be compatible with the concordance software. In this study, the concordance program used was AntConc version 3.5.7, a free program that can

count words in .txt files and sort in different ways. The concordance program showed that there are 65513 words in the first corpus, the students' articles, and 183910 words in the second corpus, the research articles published in international academic journals available from the database. The reason why the students' articles are much shorter can be from the requirement and instructions of the assignment. In the course, it was emphasized that the students should not exceed 6 pages, but the length of the articles published in journals can vary.

In addition to the difference regarding the length of the two corpora, it should also be noted that the numbers of word types found in both corpora are also remarkably different. To be more specific, the number of word types found in the first corpus is just 4406 while the second corpus contains 10349, which is more than two times of those found in the first corpus. This is absolutely the result from the considerably low vocabulary knowledge of the students.

2. Data Analysis and Statistical Techniques

After converting the files to .txt, the concordance programs were used in order to search for all lexical hedges. Each lexical hedge has to be identified word-by-word by browsing the word list shown by the program.

The present study focuses on only lexical hedges, which are surface features of hedging, because of their salient features. The types of hedges used in this study are based on the types classified by Hyland (1998) and the one modified by Schmied (2008) shown in the Figure 1, which are 1) modal auxiliaries, 2) epistemic lexical verbs, 3) epistemic adjectives, 4) epistemic adverbs, 5) epistemic nouns, and 6), numerical hedges.

Figure 1: Lexical hedges with some examples (Schmied, 2008)

Modal Auxiliaries	Epistemic Lexical Verbs		Epistemic Adjectives, Adverbs and Nouns			Hedging Numerical Data
	Epistemic judgement verbs	Epistemic evidential verbs	Epistemic adjectives	Epistemic adverbs	Epistemic nouns	
must/need	*propose*	*show*	*likely*	*probably*	*probability*	*about*
can/could	*suggest*	*appear*	*possible*	*apparently*	*possibility*	*approximately*
will/ would	*believe*	*seem*	*most*	*possibly*	*assumption*	*some*
shall/should	*speculate*	*tend*	*significant*	*perhaps*		*around*
may/might	*think*	*look like*	*clear*	*often*		
	indicate		*certain*	*usually*		

Even though the concordance program can help to count lexical hedges used in the articles, it should be noted that some of the words listed by the program can be used in other functions, not as lexical hedges. For example, the modal *must* can be used as both epistemic

and deontic modals. In a sentence like "The students must submit their work on time," *must* is used as a deontic modal regarding the freedom to act—not as a lexical hedge to show probability as in a sentence like "The students must be familiar with the format because they have done assignments in that format for many times." Similar to modals, words in other parts of speech can also be used as both lexical hedges and in other functions. For example, the word *think* in *The researchers think that ...* can be counted as a lexical hedge, but when it is used in *The teacher tries to make the student think …*, *think* here is not a lexical hedge.

Thus, although the concordance program can help to identify words, each word must be checked "by hand" again to ensure that it was used as a lexical device before categorizing them based on their parts of speech.

To ensure reliability, a second rater, who is also an English lecturer at Mae Fah Luang University, also analyzed 20% of the data. The percentage of agreement was 98% and Cohen's Kappa, the inter-rater reliability was 0.96, which could be interpreted as almost perfect agreement. It can be said that the way that the data were analyzed and categorized was consistent and reliable. Thus, only one researcher analyzed the rest of the data alone.

Finally, the number of each type of lexical hedges in both corpora was calculated into the number of that type in 1000 words before comparing, to ensure that the length of each article cannot affect the number of hedges found. This is because the students' articles were just five to six pages long as they were limited by the requirement of the assignment while the length of published articles were longer. This will be totally unfair if compared just by the tokens. Then, the Pearson's chi-square test of independence, which can be used when the data are frequencies and the sample sizes (total of hedges in both corpora) are not equal, was used in order to see if there is an association between the two variables, the different groups of writers, or not.

Findings and Discussions

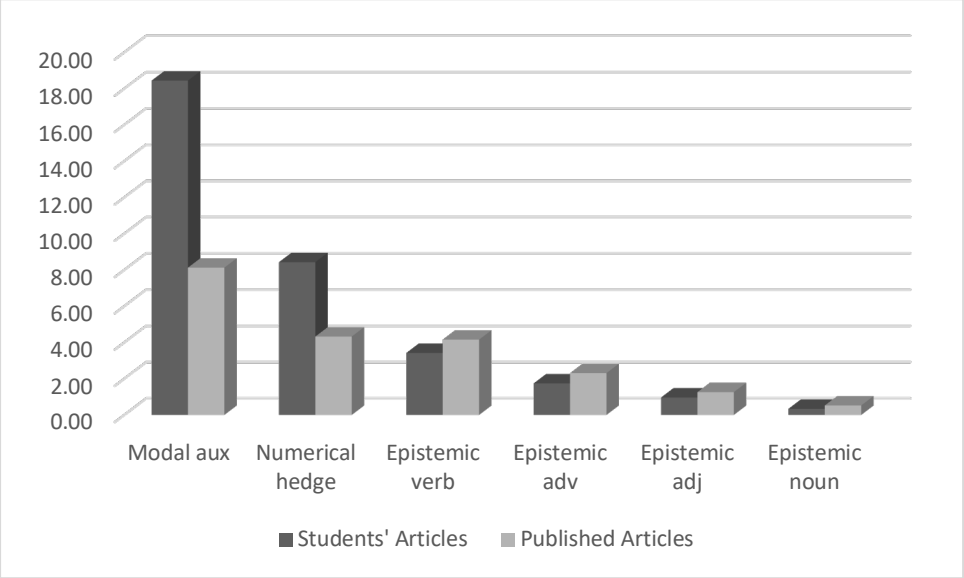
The data revealed that the students employed all types of lexical hedges, modal auxiliaries, epistemic lexical verbs, epistemic adjectives, epistemic adverbs, epistemic nouns, and numerical hedges, as shown in the Table 1.

Table 1: The overall number of hedges found in each corpus

Types of Hedges	Students' articles F (per 1000 words)		Published articles F (per 1000 words)	
1. Modal auxiliaries	1208	(18.44)	1494	(8.12)
2. Epistemic verbs	224	(3.42)	764	(4.15)
3. Epistemic adjectives	63	(0.96)	231	(1.26)
4. Epistemic adverbs	114	(1.74)	424	(2.31)
5. Epistemic nouns	22	(0.34)	96	(0.52)
6. Numerical hedges	552	(8.43)	795	(4.32)
Total	**2183**	**(33.32)**	**3804**	**(20.68)**

Among all types, the modals were found most frequently in the students' articles with the amount of 1208 and the types of lexical hedges that were used the least are epistemic nouns. This is the same as lexical hedges in the other corpus, the one with published articles, as shown in Figure 2. This can be said that the students were able to use all types of hedges, just like professional researchers. However, they may not have been confident enough to make strong claims, so they included more hedges in their papers.

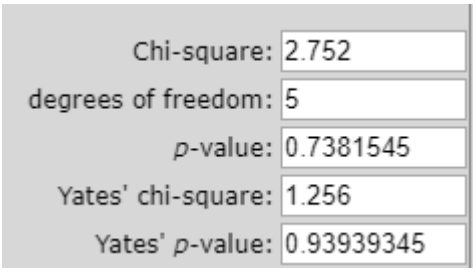
Figure 2: Types of hedges in both corpora per 1000 words



After all the lexical hedges were identified and categorized, Chi-square test of independence was used in order to see if the number of hedges found in each group is statistically independent or not. The hypothesis for the test is that the distribution of the outcome

is independent of the groups. However, since the p value is greater than the predetermined alpha level of significance (0.05), the hypothesis that the outcomes of the two set are independent has to be rejected and an association between the two variables can be assumed. The result of the Chi-square test of independence is shown in the Figure 3 below.

Figure 3: Chi-square test of independence result



Like in students' articles, the data revealed that modal auxiliaries are found most frequently in the published articles as well with the tokens of 1494 out of 183910 words in the corpus, making it 8.12 words in every 1000 words. In this category, the modal *can* was found most often in both corpora. In fact, its negative form *cannot*, counted separately by the concordance program, was also among top ten most frequently found hedges in the students' writing, as shown in the Table 2.

Table 2: The most frequently used lexical hedges in both corpora

Ranks	Students' articles		Published articles	
Lexical hedges		Tokens Found	Lexical hedges	Tokens Found
1	can	795	can	586
2	many	182	some	256
3	most	166	should	245
4	should	157	may	214
5	some	150	many	206
6	may	96	most	203
7	could	52	show	149
8	show	50	could	131
9	cannot	46	often	128
10	think	42	would	126

When compared with articles in a different field, Schmied's (2008) comparative study of lexical hedges in popular and specialized academic articles in science, the data in this study also reveal the similar results. To be specific, modal auxiliaries were found the most and the frequency

of hedges in this category is more than two times higher than the category ranked the second. However, in his study, the category that was ranked the second was epistemic verb and numerical hedge came the third but the amounts of hedges in these two categories were very close.

However, not all previous studies yielded the same results. For example, in a study that aimed at comparing texts written in English by native speakers and Chinese learners of English conducted by He, Jiangqin, and Feng (2010), the results showed that both English writers and Chinese learners of English employed full verbs the most and modal auxiliaries come second. The differences may be from both the influence of the first language, the differences in cultures, or even the language proficiency of the writers.

Actually, it is quite surprising that, from the data, the student writers employed more lexical hedges than researchers who published their articles in international journals. Among 65513 words in the first corpus, 2183 words are lexical hedges, making it 33.32 words per 1000 words while in the second corpus, published articles, 3804 of 183910 words are lexical hedges, making it 20.68 words per 1000 words.

This finding seems different from previous studies that have been done in similar areas. For example, Samaie, Khosravian, and Boghayeri (2014) studied the introduction section of academic research articles in the field of literature written in Persian and English and found that English writers used 45.26 hedges per 2000 words while Persian writer used only 23.19 lexical hedges per 2000 words. In this study, none of modal auxiliaries was used by Persian writers, possibly from the feature of the language itself.

Similarly, Yang (2013) who examined the use of hedges in three academic writing corpora, which were the English scientific writing corpus, the Chinese-authored English scientific article corpus, and the Chinese scientific article corpus also found that per 1000 words 19.7 hedges were found in English articles written by English speakers, 9.0 were found in English articles written by Chinese writers, and 11.0 were found in Chinese articles.

Even in a different genre and context, Riekkinen (2010) who investigated the use of hedges in spoken language also revealed that English native speakers used more lexical hedges when speaking in terms of both frequency and variety of hedges when compared with non-native speakers of English. In these studies, the researchers claimed that cultures, language proficiency, and linguistic differences played an important role on the types and the frequency of the hedges used.

One of the reasons why lexical hedges were found more often in the student writing can be that sometimes students double-hedged in their sentences. The following samples are from the corpus of students' articles.

(1) It seems likely that … (students' article)

(2) It seems quite obvious that … (students' article)

In the samples above, there are at least two types of lexical hedges in the main clauses of the sentences. The use of two or more lexical hedges in both sentences makes the sentences verbose and redundant. This feature is not normally found in published articles.

When looking at the frequency of lexical hedges used the most by both the students and the professional researchers, as mentioned earlier, one thing that the two corpora share in common is that auxiliary modals are used most often in both corpora—with *can* as the most frequently used.

The reasons why students used modals as lexical hedges most often can be that they are common and quite easy to use. In addition, these students had learned to use them in previous courses in their programs.

It should be noted that that the modal *can* found in the corpora would be categorized as a lexical hedge only when it could also be paraphrased as "it is possible that ___," not just "(the subject) is able to _____."

From both corpora, the contracted forms of modals, such as *can't* or *shouldn't* were rarely found, possibly because of its genre where formal language is needed. Because contracted forms are considered informal, both groups of writers tended not to use them. However, the negative contracted form *can't* was found once from the data and it was surprising because it was from a published article. It should also be noted that the negative form of can '*cannot*' is also ranked as the ninth most frequently found in the corpus of students' writing.

Conclusions and Recommendations

1. Conclusions

The present study investigated the use of lexical hedges in academic articles written by third year students majoring in English at Mae Fah Luang university and found that the students used all types of lexical hedges. The type of hedges that was used the most is modal auxiliaries with the frequency of 18.44 words per 1000 words which is almost two times higher than the type that was second most frequently used, numerical data hedges (8.43 words per 1000 words).

When compared with lexical hedges used in published articles in academic journals, modal auxiliaries and numerical data hedges were also the two most frequently used types of hedges. In fact, when put in order of frequency, the types of hedges in both corpora are also ranked the same, which is similar to some previous studies. Statistically, we can assume an association between the two corpora. Thus, it seems the writers' first language and their language proficiency do not affect the types of lexical hedges they employ.

Even though the student writers used lexical hedges a little bit more frequently than the professional writers, it seems this was the case only for modal auxiliaries and numerical hedges, which are not complicated to use. The underlying reasons why the students used those two types

most frequently can be partly from lessons in previous courses in which quantifiers and modal auxiliaries were emphasized as well as the fact that most of lexical hedges in these two types are simple (only one-syllable words) and more common--included in the New General Service List (Browne, Culligan, & Phillips, 2013).

However, one thing that indicates the need of the students is that the choices of words used in published articles are greater than those used by the students. Grammatical errors and some awkward structures were also found in the students' texts. This means that there are needs of the students to learn how to use a wide variety of lexical hedges properly and accurately. In addition, we also need to help them apply all the grammatical rules they have learned appropriately when they write their own texts.

2. Implications and Pedagogical Applications

From this study, teachers of English as a second or foreign language writers can improve or develop their pedagogical tools, such as textbooks, course materials, and even computer software or mobile applications, to help their students master the use of all types of lexical hedges. For example, they can develop materials that emphasize lexical hedges that the students used less when compared with those who published in the academic journals, namely epistemic verbs, adverbs, adjectives, and nouns. In addition, for the types of lexical hedges that the students seemed to overuse, the teachers can also raise the students' awareness by pointing out the frequencies of those words found in papers written by scholars in the field.

3. Direction for Further Research

Since it is unfortunate that this study was based on a limited amount of articles in each corpus. In some aspects, the results are still inconsistent when compared with previous studies. Hence, more research should be done in order to explore more texts, both in the same field and in other fields, and possibly by different groups of writers.

More studies can definitely shed light on this topic and help educators develop their lessons and teaching materials accordingly.

References

Aull, L. L., Bandarage, D., & Miller, M. R. (2017). Generality in student and expert epistemic stance: A corpus analysis of first-year, upper-level, and published academic writing. *Journal of English for Academic Purposes*, *26*, 29-41.

Atai, M., & Sadr, L. (2006). A cross-cultural genre study on hedging devices in discussion section of applied linguistics research articles. In *Proceedings of the conference of Pan-Pacific Association of Applied Linguistic* (pp. 42-57).

Brown, P, & Levinson, S.C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge University Press.

Browne, C., Culligan, B. & Phillips, J. (2013). *New general service list: The most important words for second language learners of English.* The New General Service List. http://www.newgeneralservicelist.org.

He, Y., Jiangqin, X., & Feng, L. Lexical Hedges in English Abstract Writing. In *2010 International Conference on Computational Intelligence and Software Engineering*. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5676749

Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: a comparative study of English- and Chinese-medium journals. *Journal of Pragmatics, 43,* 2795-2809.

Hyland, K. (1996). Talking to the academy: Forms of hedging in science research articles. *Written Communication, 13*(2), 251-281.

Hyland, K. (1998). *Hedging in scientific research articles* (Vol. 54). John Benjamins Publishing.

Hyland, K. (2005a). Prudence, precision, and politeness: Hedges in academic writing. *Quaderns de Filologia-Estudis Lingüístics*, *10*, 99-112.

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

Lakoff, G. (1972). Hedges: A study of meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2*, 458-508.

Riekkinen, N. (2009). Softening criticism: The use of lexical hedges in academic spoken interaction.

Riekkinen, N. (2010). "This is not criticism, but..." Softening criticism: The use of lexical hedges in academic spoken interaction. *Helsinki English Studies (6)*, 75-87.

Samaie, M., Khosravian, F., & Boghayeri, M. (2014). The frequency and types of hedges in research article Introductions by Persian and English native authors. *Procedia-Social and Behavioral Sciences*, *98*, 1678-1685.

Schmied, J. (2008). Hedges in specialised vs. popular academic interaction: A case study of medical text. *Discourse and Interaction*, *1*(2), 85.

Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes, 20*, 83-102.

Yang, Y. (2013). Exploring linguistic and cultural variations in the use of hedges in English and Chinese scientific discourse. *Journal of Pragmatics*, *50*(1), 23-36.