Different Usage of Hedged Confidence Language in Four Different Academic Disciplinary Fields

Author's Statement

Numbers are everywhere - when we count money, read time, and understand information. They allow us to measure, compare and solve problems, becoming an indispensable existence of our lives.

This significance of numerical data came even closer to me as I wrote this paper for 76-107: Writing about Data. Dr. Yoon provided us frameworks with diverse contexts to read, interpret data, and analyze through discussion with classmates. With these training sessions, the objective of this assignment was to collect my own data, represent it, and analyze it to create a persuasive argument.

As a student who studies engineering, I will definitely face countless frameworks full of numerical data, which makes this experience even more valuable to me. Through this assignment, I was able to learn how to collect and represent data efficiently and understand more deeply about the power of data in communication with readers.

I want readers to consider how unique features of each academic discipline might affect the language used in its papers. As a reader, you will see specialized vocabulary that is indicated (i.e. tagged) through footnotes. These taggings are one of the assignment requirements, inducing me to carefully choose pieces of expert vocabulary and use them, and therefore increase the readability of the paper.

Jaehyuk

Abstract

Hedged confidence language is a language that signals moderate certainty and openness to other points of view. This language is particularly important when writing a research paper as it leaves room for writers' arguments to be refuted by the readers, which allows writers to maintain the credibility and effectiveness of their papers. Although there has been research about usage of hedged confidence language in diverse disciplines, research about the comparison within the same disciplinary fields is under-researched. Therefore, to explore this, the Michigan Corpus of Upper-Level Student Papers (MICUSP) was used and analyzed, and the frequencies of hedged confidence language in diverse papers were measured by the tool called Docuscope. The results demonstrate that, among four disciplinary fields (i.e. Engineering, Natural Science, Social Science, Humanities), Natural Science uses hedged confidence language most frequently, whereas[1] this language was the least preferred in Engineering. More interestingly, within these Natural Science and Engineering fields, it was found that biology and civil and environmental engineering have the highest usage of hedged confidence language compared to other Natural Science and Engineering disciplines respectively. These different levels of hedged confidence language even within the same disciplinary field suggests the different nature of disciplines specifically in terms of how they convey the claim. In the end, these findings could provide educational information to both writing instructors and students in terms of how much they need to use hedged confidence language when writing a research paper in specific disciplines.

Keywords: comparison, Natural Science, Engineering, nature, educational information

[1] CGA Vocabulary: Metadiscourse (General)

Introduction

Argumentation is an important component of all research papers across all disciplines, and argumentation becomes more successful when the writers properly modulate the level of confidence in their language. Conversely, being overconfident or uncertain about one's claims is problematic for a writer as doing either may affect the perceived credibility and the effectiveness of their work. For example, the sentence "children living in poverty do poorly in school" might make readers intuitively distrust the claim because it is impossible to collect data from every child in the world, and there are still children who are successful in school despite the poverty. This sentence is therefore modified by adding the verb 'tend' to acknowledge that the following phenomenon is simply a tendency.

What this example sentence indicates is that maintaining the right level of confidence in writing is crucial to conveying a clear message, which can be performed through using hedged confidence language. However, entire fields and individual disciplines within those fields are likely to use varying degrees of hedged confidence language due to their different disciplinary natures. While most of the earlier research has been focused on hedged confidence language in a single discipline, it is also important to examine the general variation in hedged confidence language across the different disciplinary fields and individual disciplines so that writers can convey a clear message with adequate clarity. Hence, we can establish[2] the following question that will be answered throughout this report: How do the different disciplinary fields and disciplines within the same field compare with each other in terms of the use of hedged confidence language?

[2] CGA Vocabulary: Metadiscourse (First-Person)

Methods

Data Collection

For this study, the Michigan Corpus of Upper-level Student Papers (MICUSP) was used as the source of the main data. MICUSP is a corpus of 829 papers written by University of Michigan students in their final undergraduate year or in their first three years of graduate education. The papers are from 16 disciplines across four academic divisions: Humanities and Arts, Social Science, Biological and Health Sciences, and Physical Sciences.

Moreover, specific languages such as ‘tend’ or ‘may’ used in papers can be searched through MICUSP, and their frequencies in particular fields can be discovered as well. Since MICUSP enables both quantitative and qualitative analysis on disciplinary aspects of student papers, it is regarded as a powerful tool for this study and therefore was selected as a database.

Measurement

Docuscope. Docuscope began as an education tool in 1998 at Carnegie Mellon University. It has been developed over 20 years, and now Docuscope has grown into a suite of various tools, from Docuscope Corpus Analysis to MyScribe and OnTopic. In this study, we will specifically focus on Docuscope Corpus Analysis, which has two core elements: its own dictionary and analysis and visualization software. Through these elements, Docuscope analyzes words and phrases into rhetorical perspective and classifies them into key categories.

Docuscope Categories. Each category is defined distinctly. Every category in Docuscope contains specific examples of words and phrases such as ‘unlikely’ for the *Low Confidence* language category. Docuscope counts the frequency of these phrases per 100 words of the paper,

allowing users to perform quantitative analysis[3] of specific linguistic features in papers of different disciplinary fields and disciplines.

Hedged Confidence. For this study, the *Hedged Confidence* language category is particularly selected to analyze. Hedged confidence language is a language that conveys tempered assurance and openness to alternative viewpoints. For example, in the sentence "dragonfly nymphs may not be particularly active organisms", a modal verb "may" is considered to be hedged confidence language as it gives moderate certainty and prepares for the possibility of the argument being not true. Several other examples of words that qualify as hedged confidence language are 'can', 'might', 'suggest', and 'possible'[4].

The reason why this category was particularly chosen is because, except for a few obvious disciplines and disciplinary fields, many can find it difficult to predict which field might prefer or not prefer hedged confidence language. This difficulty might come from the fact that we use hedged confidence language in so many different contexts. Since the difficulty in prediction of results tends to captivate readers, the *Hedged Confidence* language category was an interesting and reasonable choice for this research.

Text Selection

Since the present study was interested in how hedged confidence language varies throughout four disciplinary fields first, the 16 disciplines represented in the MICUSP were grouped into four fields: Engineering, Humanities, Natural Science, and Social Science. This broad categorization enabled us to find out which disciplinary field seems compelling to examine further.

---

[3] CGA Vocabulary: Metadiscourse (Methods, Results and Discussion)

[4] Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

Disciplinary Fields	**Individual Disciplines**
Engineering	Civil and Environmental Engineering (CEE), Industrial and Operational Engineering (IOE), and Mechanical Engineering (ME)
Humanities	English, History and Classical Studies, and Philosophy
Natural Science	Biology, Natural Resources, and Physics
Social Science	Economics, Education, Linguistics, Political Science, Psychology, Nursing, and Sociology

Method of Analysis

Based on data processed by Docuscope, three vertical bar graphs - which show normalized mean counts of hedged confidence language per 100 words - were created. These graphs were analyzed by comparing the values to know the differences in central tendency and ultimately coming up with a possible reason that explains such distinctions.

To be more specific, the graphs were used to categorize disciplinary fields or disciplines into *most common*, *common*, and *least common*. Then, if the difference between *most common* and *least common* groups seems significant when considering the central tendency value of the *common* group, we interpreted the reason why such differences are formed. During this process, we studied the essence of each disciplinary field and discipline and what aspects make them different from others. For instance, when interpreting the field of Engineering, we differentiated Engineering from Natural Science by considering Natural Science as a study to understand the real world while Engineering as a study of applying those scientific concepts to solve problems in the real world.

Results

Hedged Confidence language Usage given Individual Discipline in four Disciplinary Fields

Hedged confidence language is used most frequently in the field of Natural Science among four academic disciplinary fields.

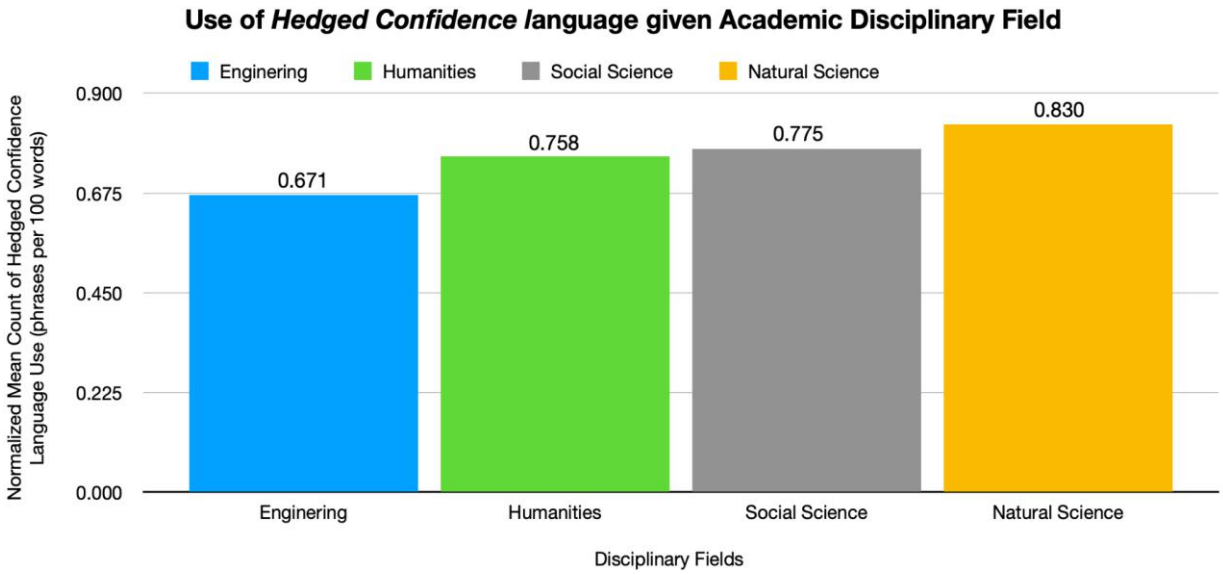

Figure 1: Normalized mean count (phrases per 100 words) of hedged confidence language use in four disciplinary fields: Engineering, Humanities, Social Science, and Natural Science. Natural Science has the highest usage of hedged confidence language.

Figure 1 demonstrates that hedged confidence language is most frequently used in Natural Science with the largest difference of 0.159 phrases per 100 words from Engineering; while the mean count of Engineering is 0.671 phrases per 100 words, that of Natural Science is 0.830 phrases per 100 words. On the other hand, Humanities and Social Science surprisingly have a similar usage of hedged confidence language, only with a difference of 0.017 phrases per 100 words. These findings show that hedged confidence language usage can be grouped into three stages: *most common, common, and least common*. Natural Science is in the *most common* stage, Engineering is in the least *common* stage, and the remaining Humanities and Social Sciences are in the *common* stage. This clear division may be explained by regarding the correlation between what individual fields tend to focus on and how much

uncertainties are contained in their studies.

Hedged Confidence language Usage given Individual Discipline in Field of Natural Science

Since Natural Science papers unexpectedly include the largest amount of hedged confidence language according to the above finding, Natural Science is selected as a subject to be examined further. Diving into Natural Science more specifically shows that biology uses more hedged confidence language compared to the other disciplines.

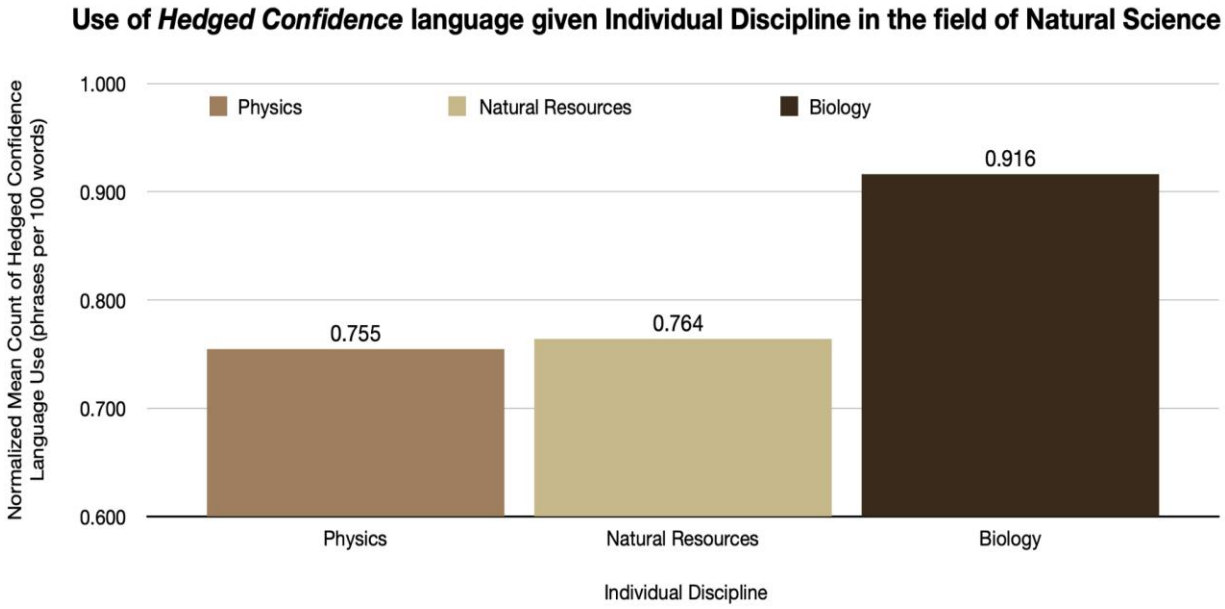

Figure 2: Normalized mean count (phrases per 100 words) of hedged confidence language use[5] in individual disciplines in the field of Natural Science: physics, natural resources, and biology. Hedged confidence language use is the highest in biology.

Figure 2 shows that hedged confidence language is most commonly used in biology while the other two disciplines relatively have a similar amount of usage; the mean count of biology is 0.916 phrases per 100 words, while physics and natural resources have the mean counts of 0.755 and 0.764 phrases per 100 words respectively. Although the mean difference of 0.157 phrases per words between biology and the other two disciplines may be trivial, it is still important to note that biology has higher frequencies than the other two Natural Science

[5] CGA Vocabulary: Metadiscourse (Expert Vocabulary)

disciplines and is significant to understand why such a phenomenon happens. This variant may be justified through considering how much empirical evidence composes biology and other two disciplines.

Hedged Confidence language Usage given Individual Discipline in Field of Engineering

The field that has the largest difference with Natural Science is none other than Engineering according to Figure 1. This is important because Natural Science and Engineering share many similarities considering the fact that they both belong to the same STEM field. Despite such similarities, Engineering interestingly having a largest difference in hedged confidence language leads Engineering to be investigated more.

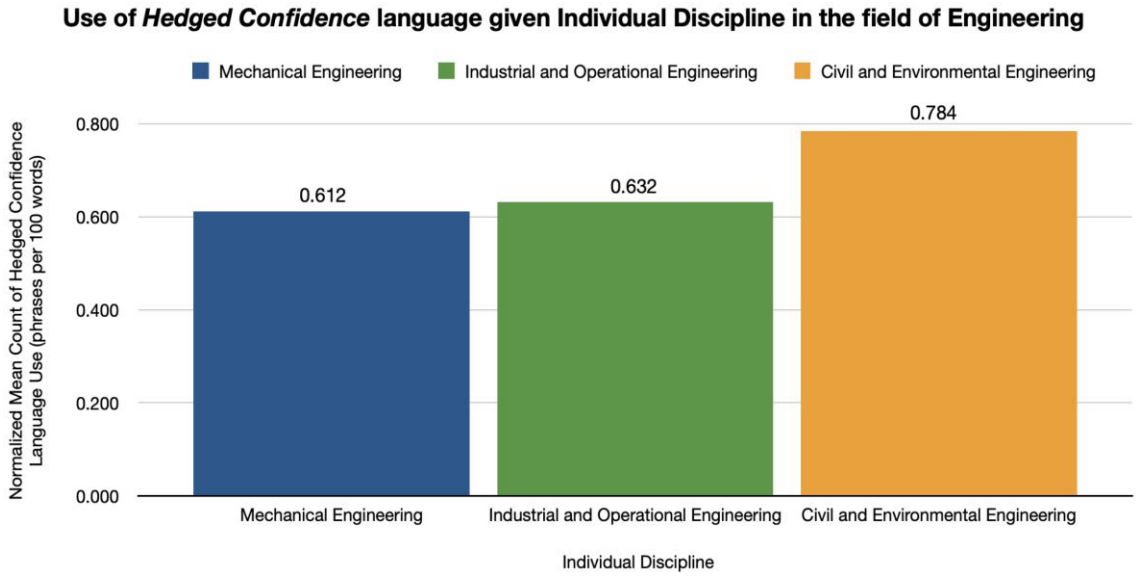

Figure 3: Normalized mean count (phrases per 100 words) of hedged confidence language use in individual disciplines in the field of Engineering: mechanical engineering (ME), industrial and operational engineering (IOE), and civil and environmental engineering (CEE). CEE has the highest use of hedged confidence language.

Figure 3 suggests that hedged confidence language is most frequently used in CEE and least used in ME; while the mean count of CEE is 0.784 phrases per 100 words, that of ME is 0.612 phrases per 100 words. Compared to the central tendency values of CEE and ME, the difference of 0.020 phrases per 100 words between IOE and ME is relatively trivial. The graph

conclusively shows similar use of hedged confidence language in IOE and ME and anomalously high use of hedged confidence language in CEE at the same time. The mean difference of CEE and other two Engineering disciplines is 0.162 phrases per 100 words. This difference can be analyzed through reflecting how much social science perspectives comprise each Engineering discipline.

Discussion

Use of Hedged Confidence language in Four Disciplinary Fields

The results show that Natural Science has the highest usage of hedged confidence language and Engineering has the lowest usage. This phenomenon might be due to the difference in what these two fields tend to concentrate on. To be more specific, Natural Science is a study of the physical world where exceptions and uncertainties often occur. This feature tends to make it difficult for scientists to provide a definitive answer through their research and their findings. Since there is nothing completely "right" or "wrong" - which is why Natural Science develops everyday as new concepts are discovered - writers in this field try to be cautious as much as possible in their language choices. For instance, in a biology paper discussing subcellular protein localization, the student writer states that "Given adhesion function of HWP1, the result of WOLFPSORT makes more sense. And it is consistent with the result of TargetP. However, we cannot rule out the possibility that HWP1 may localize to nucleus."[6] Although the scientific findings in this example are concrete, which can be seen from the first two sentences, the student writer uses the word "possibility" to acknowledge that there might be other situations affecting the result. Such a language with moderate certainty allows room for potential flaw in their reports and therefore makes readers consider the arguments more like phenomenal hypotheses

[6] Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

and less like factual reality.

On the other hand, Engineering is the study of applying scientific principles to improve our lifestyles and increase our convenience, which can be seen from papers in MICUSP like: "Design Modifications to Reduce Vibration on Electric Vehicle Drive Shaft"[7] and "Soil Improvement: Designing with Tensar Geogrids"[8]. The words "design" and "improvement" indicate that writers in this field tend to focus on specific real-world problems and come up with solutions based on the most efficient designs. Compared to Natural Science where scientists tend to consider all the possibilities with no definitive answer, Engineering primarily focuses on increasing the efficiency of the items. This relatively specific nature of Engineering, in which specific, quantifiable answers are sought for known problems, may cause the relative rarity of hedged confidence language in engineering alongside Natural Science. This altogether suggests that depending on the nature of the fields in terms of what they specifically study, there can be varying degrees of hedged confidence language usage in different disciplinary fields.

Use of Hedged Confidence language given Individual Discipline in field of Natural Science

Studying Natural Science more deeply shows that biology prefers hedged confidence language the most among the constituent disciplines. This could be because biology draws more on empirical data compared to the other scientific disciplines. Biology is a study of living organisms. This means that to develop scientific concepts in biology, biologists necessarily interact physically with living organisms. As a result, developing biological knowledge tends to involve empirical methods such as experimentation, experience, and observation where uncertainties tend to be inevitable and abundant.

For example, in a biology paper discussing the role of leptin in cardiovascular disease,

[7] Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.
[8] Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

the student writer writes that "After adjustment for age and BMI, researchers found that elevated leptin levels correlated significantly with increased blood pressure in obese individuals. These results suggested a possible role of leptin in hypertension."[9] Although a certain correlation between leptin levels and blood pressure in obese individuals is discovered, the student writer uses the word "possible" to acknowledge the potential uncertainty in their observations and therefore protect the arguments from being easily dismissed by the readers.

In contrast, physics and natural resources are the studies of matter and interactions between humans and environments respectively. Although physics uses empirical data to develop knowledge, it is inherently more theoretical. The word "theory" means a speculative hypothesis that relies on mathematical concepts many times and is often not proved experimentally. That is, compare to biology, physics has more knowledge based on mathematical principles rather than experiments. Since mathematical principles tend to be distinct (e.g. geometric principles), the greater prevalence of mathematics in physics relative to biology may induce student writers to use less hedged confidence language. Furthermore, despite the fact that natural resources is also a study of living organisms, it is an interdisciplinary field where many different areas - such as economics and social sciences - are often focused on concurrently, making it a less empirical discipline than biology.

Hence, biology tends to be more empirical than the other two Natural Science disciplines. This difference may cause greater uncertainty to be treated by student writers in biology and therefore lead them to be more cautious about their arguments by using more hedged confidence language. These interpretations clearly suggest that different amounts of empirical aspects in the discipline may lead to the various degrees of different hedged confidence language usage.

[9] Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

Use of Hedged Confidence language given Individual Discipline in field of Engineering

Similar to the above circumstance in Natural Science, civil and environmental engineering (CEE) has the highest mean frequency of hedged confidence language compared to other two Engineering disciplines. This may be attributable to the difference in the amount of the social science element involved in each engineering discipline.

CEE is the interdisciplinary field that spans engineering, social sciences, and policy to solve sustainability and social challenges.[10] That is, when developing knowledge in CEE, it is likely to be necessary to consider a social science perspective such as the management of noise production from infrastructure construction. Since social sciences focus on the interaction between humans and society, variables tend to be difficult to control and therefore results are likely to be unpredictable. This causes Social Sciences to intrinsically contain high amounts of uncertainty as a field, which can be clearly seen from Figure 1 as the usage of hedged confidence language in Social Science ranked second highest among the fields tested. As a result, a high proportion of social science elements in CEE tends to lead the writers in this field to be more cautious in their language choices and to thus communicate the kind of uncertainty included in human behavior.

On the other hand, mechanical engineering (ME) is a study that combines physics and mathematics to design and create machines for our convenience. Similarly, industrial and operational engineering (IOE) is a study that analyzes data to create more efficient systems, processes, and machines that improve lives for all people[11]. It may seem that IOE and ME

[10] University, C.M. (n.d.). Welcome to Civil and Environmental Engineering at CMU - Civil and Environmental Engineering - Carnegie Mellon University. [online] www.cmu.edu. Available at: https://www.cmu.edu/cee/ [Accessed 1 Apr. 2024].

[11] University of Michigan. (n.d.). *Industrial & Operations Engineering – Michigan Engineering.* [online] majors.engin.umich.edu. Available at: https://majors.engin.umich.edu/program/industrial-operations-engineering/ [Accessed 1 Apr. 2024].

consist of social science perspectives as much as CEE does, but ME and IOE relatively rely more on scientific and mathematical principles. As physics and mathematics tend to have a certain answer to the problem and does not consider human behaviors as much as CEE does, it may be reasonable for student writers in ME and IOE to use less hedged confidence language in their papers. To sum up, the level of hedged language usage appears to correlate with the extent to which a discipline engages with social science considerations.

Limitations

Our finding that Natural Science has the highest frequency of hedged confidence language stands in contrast to the results of prior research[12] by Masahiro Takimoto in 2015 which claims that "natural science papers were underrepresented in the number of hedges." This discrepancy may occur due to the differences in the sample size and the definition of Natural Science between Takimoto (2015) and the present study. While this study only considers biology, physics, and natural resource as Natural Science, there are four disciplines of physics, electrical engineering, mechanical engineering, and chemistry identified as Natural Science in the prior research.

The above discrepancy between prior research and this study implies that the main limitation of our study is that our sample size is restricted. Since the data used in this study is solely based on the papers of University of Michigan students, it is not reasonable to generalize[13] our current findings. In other words, although the findings of the study are true in the University of Michigan student community, it is not guaranteed to state the findings also apply to the other communities. Therefore, more samples with larger populations should be collected, and random

[12] Takimoto, M. (2015). A corpus-based analysis of hedges and boosters in English Academic Articles. *Indonesian Journal of Applied Linguistics*, 5(1), p.90. doi:10.17509/ijal.v5i1.836.
[13] CGA Vocabulary: Metadiscourse (Reasoning)

sampling should be conducted to increase the generalizability.

Another limitation is that there are subjects that are officially included in the disciplinary fields but not shown in the database. For example, chemistry is one of the official subjects in Natural Science, but it does not appear in the MICUSP database. Since this study considers only a few subjects to represent the mean or other statistical measures of disciplinary fields, comparison between disciplinary fields might be misleading and therefore needs room for improvement. Thus, the data of more disciplines should be collected so that the statistical measures of the disciplinary field can be calculated more accurately to provide a clear comparison between the fields.

Uniqueness of Current Study

Our study focuses on the hedged confidence language use in many different disciplines within the same field for the comparison purposes. This comparison not only enables investigating the discipline of highest and lowest frequency but also allows identifying how significant the difference is. In the end, we hope the results and interpretations of current study can be used for educational purposes. This study would provide both writing instructors and students a brief sense of how much hedged confidence language they might use in a paper for disciplines in Natural Science and Engineering.

References

Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

Takimoto, M. (2015). A corpus-based analysis of hedges and boosters in English Academic Articles. *Indonesian Journal of Applied Linguistics*, *5*(1), p.90. doi:10.17509/ijal.v5i1.836.

University, C.M. (n.d.). *Welcome to Civil and Environmental Engineering at CMU - Civil and Environmental Engineering - Carnegie Mellon University.* [online] www.cmu.edu. Available at: https://www.cmu.edu/cee/ [Accessed 1 Apr. 2024].

University of Michigan. (n.d.). *Industrial & Operations Engineering – Michigan Engineering.* [online] www.majors.engin.umich.edu. Available at: https://www.majors.engin.umich.edu/program/industrial-operations-engineering/ [Accessed 1 Apr. 2024].