

A Corpus-Based Analysis on The Use of Personal Pronouns in Chinese Students' Academic Spoken English

Jueyu Lu

The Education University of Hong Kong, Hong Kong (China)

Keywords: Corpus-based analysis, Personal pronoun use, Chinese student, academic spoken English, higher education.

Abstract: As English has become the dominant language of instruction in higher education, it is becoming increasingly important in countries where it has traditionally been taught as a foreign language or a second language, such as China. Studying Chinese students' academic Spoken English performance and comparing it to the general performance of academic English to identify differences can help to better target Chinese students to approach the purpose of academic English and convey arguments better. This study focuses on Chinese students' spoken academic English, with an emphasis on the use of the pronouns "I" and "you". Through corpus-based analysis, the characteristics of Chinese students' use of these two pronouns were identified and compared with the features of academic English speaking in general to identify areas where Chinese students' pronoun use can be improved.

1. Introduction

English language plays an important role in the internationalization of higher education worldwide, which can be considered the lingua franca of international academic (Jenkins, 2014). Without English language, learners would not be able to function effectively in an international educational environment. As an option for "presenting a paper" at an academic conference or lecture, academic speaking is an important part of academic English (Chafe, 1986). Research on spoken academic English began a few years ago with studies on the subjectivity of vocabulary, the frequency of new words (Nesi, 2002), and the use of pronouns (Gómez, 2006). However, these studies have focused on the general characteristics of academic English speaking, which has no targeted research on English as foreign language (EFL) students. Therefore, this study focuses on Chinese students' spoken academic English and explores the characteristics of their pronoun use by conducting a corpus-based analysis.

2. Literature review

2.1. Pronoun use in academic spoken

Pronouns in spoken language serve to provide a point of reference for the listener to understand a verbal event. Personal pronouns are very common in spoken English, mainly in conversation (Biber & Quirk, 2012). As a result, some specific studies on the use of personal pronouns in spoken academic English have been conducted. For example, Rounds (1987) defines the referential and discourse contexts of the personal pronouns "we", "I" and "you" (see *Table 1*).

Table 1: Reference and discourse contexts of personal pronouns (Rounds, 1987).

Referent	Pronoun	Discourse contexts
Teacher	I, we	Reporting previous remarks; announcing future actions
Students	You, we	Referring to student responsibility; admonishments
Students and teacher	I, you, we	Working with specific examples; announcing future actions
Mathematicians and teacher	I, we	Naming; defining
Anyone who does calculus	I, you, we	Working with mathematical procedures

As shown in *Table 1*, “I” and “you” can refer to students, teachers, mathematicians, and anyone who does calculus. In addition to referent objects, previous studies have focused on the frequency use of pronouns. Morell (2001) indicated that the use of “you” was higher in interactive lectures, while the use of “I” was higher in non-interactive lectures.

This provided the search focus for this study, in terms of the frequency of pronoun use among Chinese students in interactive speaking versus monologic speaking.

2.2. Pronouns as discourse markers

However, “I” and “you” are not only used as personal pronouns in spoken language, but also as discourse markers. For example, the phrases “I mean” and “you know” can be regarded as discourse markers according to Erman’s (2001) definition. Discourse markers focus on the coherence of the text (Schourup, 1999). Studies have shown that discourse markers have an impact on foreign college students’ comprehension of college classes, and that macro markers help students better recall textual material (Chaudron & Richards, 1986). Therefore, this study also focuses on the use of Chinese students’ discourse markers “I mean” and “you know”. The role of these two discourse markers in connecting the text and interacting with the audience was investigated.

2.3. Relevance to Chinese students

No one is a native speaker of academic language (Bourdieu et al., 2014). And most users of academic English are not native English speakers. As Jenkins (2014) stated, in many cases, users of academic English are non-native English speakers who use English in an English as Lingua Franca (ELF) setting. And most of the people they communicate with are also non-native English speakers. This involves many Chinese students, as the number of Chinese students has increased in recent years. As non-native speakers of English and learners of academic English, the performance of Chinese students in academic English is worth studying. As of now, however, research has focused on academic English users as a whole, with no focus on Chinese academic English users. Therefore, this study will focus on Chinese students’ spoken academic English, mainly examining their use of pronouns. This will provide Chinese students with a direction for improvement and a model for learning to improve their academic speaking. Based on the above, this study will address the following two hypotheses:

(1) Chinese students’ interactive speaking uses more personal pronouns “I” and “you” than monologue speaking, especially the latter.

(2) Chinese students use the discourse markers “I mean” and “you know” less frequently.

3. Methodology

To test the above two hypotheses, this study explored the characteristics of Chinese students’ academic English speaking through a comprehensive research method. A documentary analysis approach, corpus-based analysis and data visualization were used. The specific purpose of this study was to discover Chinese students’ academic English-speaking characteristics by using the “onion model” proposed by Korthagen (2004). The name of the model comes from its shape. There are circles in each circle from the center to the outside, looking like an onion cut in half in the middle. It describes

the different levels at which reflection may occur. This model of reflection is related to professional identity. The layers of the onion model include: competence, beliefs, identity, and mission, as shown in *Figure 1*.

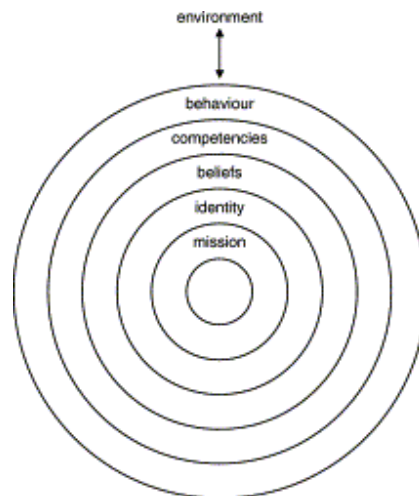


Figure 1: The onion: a model of levels of change (Korthagen, 2004).

In this study, specific performances (behaviors) of Chinese students in spoken academic English were collected through documentary analysis method and corpus-based analysis. These behaviors are then specifically analyzed through data visualization to derive the focus (competence) of Chinese students in the process of language acquisition.

3.1. Data collection

Relevant background information was systematically collected through papers related to academic speaking in English. As a qualitative analysis, documents can provide important information such as voice, interpretation, and meaning that can fully assist in understanding the problem at hand, while also posing questions and defining research directions for subsequent studies (Stage et al., 2013). To complete the document collection, a search was conducted using the keyword “English academic speaking”, and 12 relevant articles were read, of which 4 were identified as core articles. These 4 core articles show that previous research on academic English speaking has mainly focused on the use of personal pronouns, especially the use of “I” and “you”, and explored the correlation between pronoun use and interaction. To further explore pronouns and their role, “personal pronouns” and “pragmatic markers” were used as keywords in another search. 7 relevant documents were read, of which 4 core documents provided direction for the study, which investigates the characteristics of “I” and “you” in Chinese students’ academic speaking. Although documents are non-interactive and non-responsive as a data source, the document analysis method is largely adapted to this study as a research method that can be used to obtain richer resources at low time and money costs.

After determining the direction of the study, a corpus-based analysis was conducted using Michigan Corpus of Academic Spoken English (MICASE) by The University of Michigan English Language Institute (2019) to identify the referential and discourse functions of the pronouns “I” and “you” in Chinese students’ academic speech. As a method to study linguistic phenomena in depth, the corpus helps to analyze research through computers with real communicative contexts (Hasko, 2020). For this purpose, an advanced search was conducted in MICASE to create a corpus of academic spoken corpus of Chinese students. By qualifying speaker attributes’ academic position/role, native speaker status and first language, 7 lectures were selected to form a new corpus. Among them, academic position/role was restricted to junior graduate students, junior undergraduate, senior graduate students and senior undergraduate. Native speaker status was limited to non-native speakers, and first language was limited to Cantonese and Mandarin.

The selected lectures include:

(1) Graduate Online Search and Database Lecture. Mostly interactive. 18039 words. September 30, 1999.

- (2) Computer Science Office Hours. Highly interactive. 19044 words. November 13, 1998.
- (3) Behavior Theory Management Lecture. Mostly interactive. 12698 words. October 4, 1999.
- (4) Chemistry Discussion Section Student Presentations. Highly interactive. 6575 words. June 16, 1998.
- (5) Graduate Physics Lecture. Mostly interactive. 13008 words. 13008 words. March 17, 1998.
- (6) Graduate Macroeconomics Lecture. Mostly monologic. 7865 words. January 19, 1999.
- (7) Principles in Sociology Lecture. Mostly interactive. 10246 words. September 9, 1999.

These seven transcriptions were integrated into a new corpus to target the academic speaking characteristics of Chinese students, and this new corpus was read comprehensively and entered into AntConc for further research. Although the corpus may not be adequately populated with data with time, it can support this study to access, highlight, and systematically explore the linguistic phenomena of Chinese students’ academic speaking.

3.2. Data analysis

In this study, the data retrieved from AntConc were counted and visualized by Statistical Product Service Solutions (SPSS), which made the presentation of the data more intuitive. Although it may not be communicated effectively due to the complexity of the presented graphics, this study reasonably circumvents this problem by selecting graphics suitable for the display of relevant data for visual representation.

First, AntConc retrieved the number of occurrences of “I” and “you” in each spoken segment. And the normalized frequencies were calculated. The data were then visualized in SPSS to represent the characteristics of the data and to support the validation of hypothesis (1). Then the relevant frequencies were compared with the general frequency data for the use of “I” and “you” in MICASE by Gómez (2006). This allows to draw differences between the pronoun use of Chinese students and the use of pronouns in general academic speaking.

To test hypothesis (2), the number of occurrences of “I mean” and “you know” in each spoken segment was retrieved by AntConc. Also, their frequency as discourse markers in “I” and “you” was calculated and visualized. Then, similar to the test hypothesis (1), the resulting features are compared with the features of discourse marking in MICASE summarized by Gómez (2006) to draw conclusions.

4. Results and discussion

To introduce a qualitative study, the number of occurrences of “I” and “you” in the corpus of Chinese students’ academic spoken was counted by AntConc and their frequency (per 1,000 words) was calculated. *Table 2* shows the results:

Table 2: Number and frequency of the pronouns “I” and “you” in each spoken.

	I		You	
	Number	Frequency	Number	Frequency
Spoken (1)	719	39.86	559	30.99
Spoken (2)	591	31.03	708	37.18
Spoken (3)	247	19.45	414	32.60
Spoken (4)	124	18.86	181	27.53
Spoken (5)	213	16.37	418	32.13
Spoken (6)	182	23.14	65	8.26
Spoken (7)	183	17.86	295	28.79

4.1 Frequency

The data for “Frequency” in *Table 2* were visualized to obtain *Figure 2*:

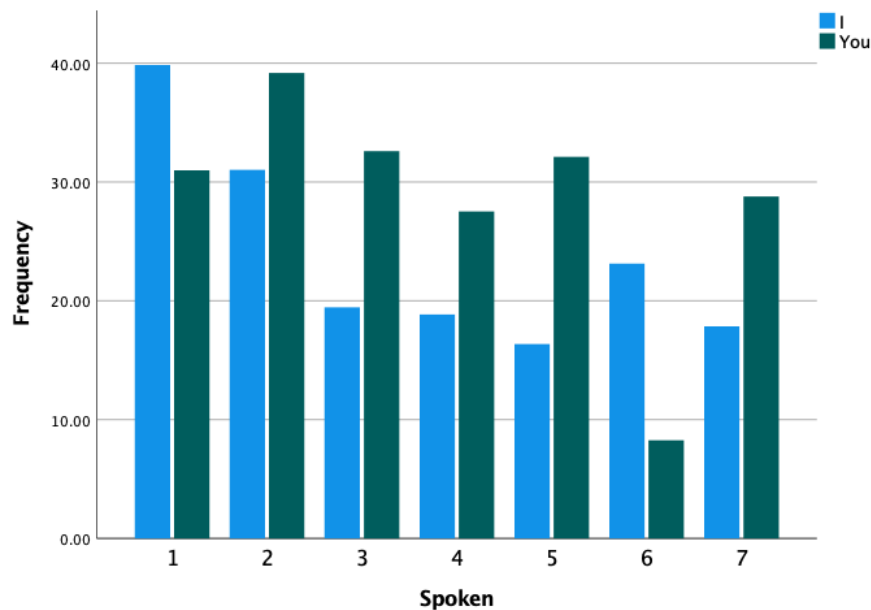


Figure 2: Frequency of the pronouns “I” and “you” in each spoken.

As shown in the figure, except for Spoken (1) and Spoken (6), “you” is used more frequently than “I” in the rest of spoken. This is not consistent with Gómez’s (2006) suggestion that “I” is used more frequently than “you” in MICASE. Therefore, the frequency of “I” and “you” in Chinese students’ academic English was further calculated and the results are as follows (*Table 3*):

Table 3: Frequency of “I” and “you” in whole corpus.

I		You	
Number	Frequency	Number	Frequency
2259	25.82	2640	30.18

The results show that “you” is used more frequently than “I” in Chinese students’ academic speaking. However, considering that there are 6 interactive spoken but only one monologue spoken in the corpus, there is a bias due to the uneven distribution of the sample, so the corpus is classified as interactive and monologue and compared again. The frequency of the data in *Table 2* was calculated for the categories “interactive” and “monologue”, and the results are as follows (*Table 4*):

Table 4: Number and frequency of the pronouns “I” and “you” in interactive spoken and monologic spoken.

	I		You		Total	
	Number	Frequency	Number	Frequency	Number	Frequency
Interactive	2077	26.09	2575	32.35	4652	58.44
Monologic	182	23.14	65	8.26	247	31.40
Whole		49.23		40.61		

The frequencies of “I” and “you” in the interactive speech were averaged and summed with the frequencies in the monologue speech. The results showed that the frequency of “I” is 49.23 and the frequency of “you” is 40.61. “I” is more frequent than “you”. This result is in line with Gómez’s (2006) summary of the frequency of both in MICASE. This shows that Chinese students’ academic speaking is generally consistent with the characteristics of academic speaking.

In order to further investigate the relationship between the use of the pronouns “I” and “you” and the form of the speech (interactive/monologic), the “Frequency” data in *Table 3* were visualized as follows (*Figure 3*):

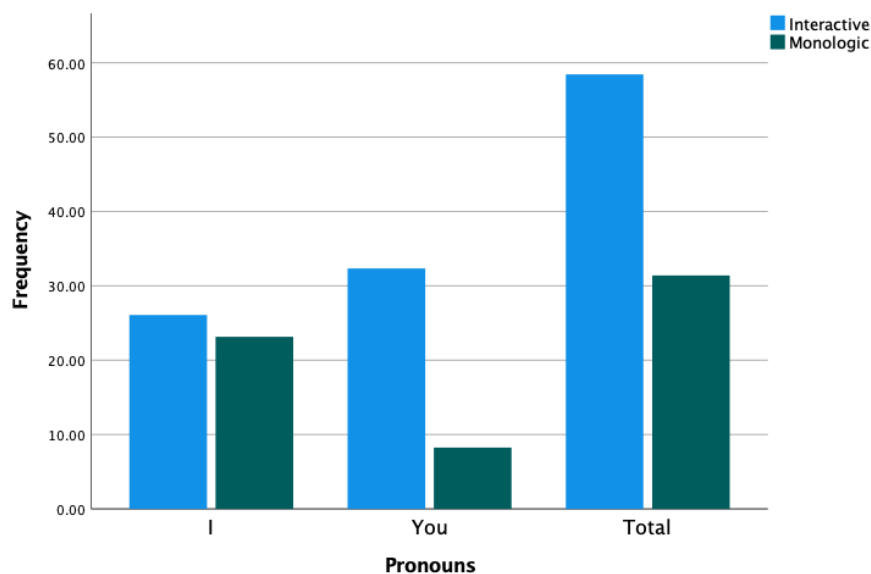


Figure 3: Frequency of the pronouns “I” and “you” in interactive spoken and monologic spoken.

As can be seen from the figure, the frequency of pronoun use in interactive spoken English is higher than that in monologic spoken English, whether it is the use of “I”, the use of “you” or the sum of the two. In particular, the use of “you” in interactive spoken language is nearly four times more frequent than its use in monologic spoken language. This shows that the use of pronouns is more common in interactive speech than in monologue, and the increase in the use of “you” is especially obvious. This confirms that Chinese students’ academic English speaking is consistent with the general characteristics of academic English speaking: more interaction is presumed due to the involvement of other speakers, and the more interaction there is, the more pronouns “I” and “you” are used, especially the latter, which is in accordance with the general characteristics of academic speaking in MICASE. This also verifies hypothesis (1).

4.2. Discourse function

The study then further analyzed the discourse function of “I” and “you” in Chinese students’ academic speaking and found that these two pronouns have two main purposes in speaking: referent and discourse marker. When “I” is used as a referent, it refers to the speaker him or herself, and when it is used as a discourse marker, it takes the form of “I mean”. Similarly, when “you” is used as a referent, it refers to the listener and is mostly in plural form, and when it is used as a discourse marker, it is expressed as “you know”.

The number and frequency of occurrences of “I” (*Table 5*) and “you” (*Table 6*) in the seven spoken segments under two different functions were counted as follows:

Table 5: Number and frequency of the function of “I” in each spoken.

	S (1)	S (2)	S (3)	S (4)	S (5)	S (6)	S (7)	Total
Refer to speaker	697 (96.94%))	490 (82.91%))	221 (89.47%))	116 (93.55%))	208 (97.65%))	180 (98.90%))	144 (78.69%))	2056 (91.01%))
Discourse makers	22 (3.06%))	101 (17.09%))	26 (10.53%))	8 (6.45%))	5 (2.35%))	2 (1.10%))	39 (21.31%))	203 (8.99%))
Total	719	591	247	124	213	182	183	2259

Table 6: Number and frequency of the function of “you” in each spoken.

	S (1)	S (2)	S (3)	S (4)	S (5)	S (6)	S (7)	Total
Refer to hearers	516 (86.14%))	653 (92.23%))	368 (88.89%))	176 (97.24%))	411 (98.33%))	65 (100.00%))	231 (78.31%))	2420 (91.67%))
Discourse markers	43 (13.86%))	55 (7.77%))	46 (11.11%))	5 (2.76%))	7 (1.67%))	0 (0.00%))	64 (21.69%))	220 (8.33%))
Total	599	708	414	181	418	65	295	2640

Table 5 and Table 6 show that “I” and “you” as discourse markers account for 8.99% and 8.33% in the total number of occurrences of the two pronouns, respectively. Compared with Gómez’s (2006) statistics of 6% and 17%, the total frequency of the two pronouns as discourse markers is lower than that of MICASE, although “I” as discourse marker is more frequent in Chinese students’ academic speaking. This indicates that Chinese students use “I mean” and “you know” as discourse markers less commonly than most of the features of academic speaking represented by MICASE.

For further study, a visualization of the “frequency” data in Tables 5 and Table 6 was performed (Figure 4 and Figure 5).

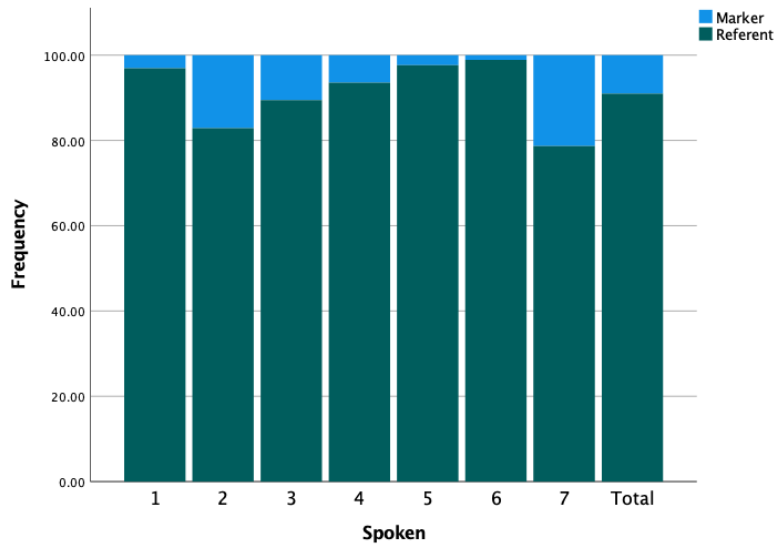


Figure 4: Frequency of the function of “I” in each spoken.

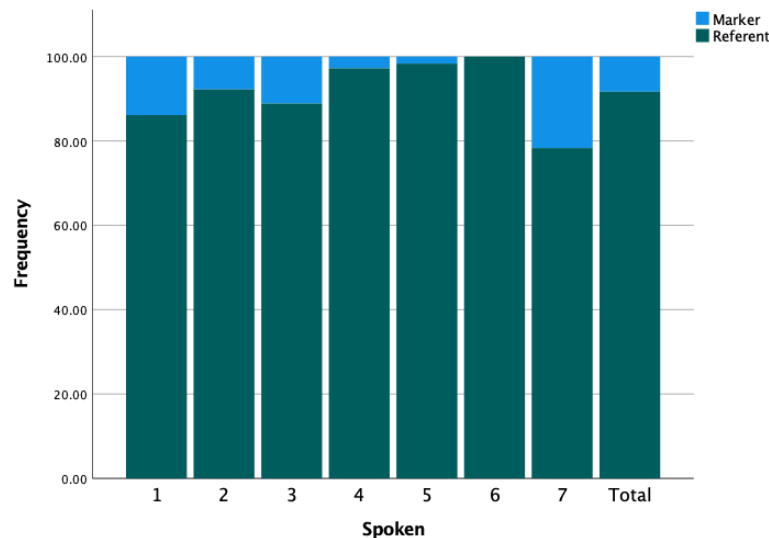


Figure 5: Frequency of the function of “you” in each spoken.

The bar chart illustrates that Chinese students do not often use “I mean” and “you know” as discourse markers in academic speaking. Especially in Spoken 6, the monologic spoken, the frequency of “I mean” is the lowest among the 7 Spoken items, even without the use of “you know”. According to Erman (2001), discourse markers can emphasize the affective function of the power of speech acts and have the effect of eliciting responses from listeners. Therefore, it is reasonable that discourse markers are used less in monologue speaking in Chinese students’ academic speaking, because the focus of monologue is on expressing one’s own meaning but not on triggering the interaction of the listener. However, in interactive speaking, discourse markers are used less frequently. This verifies hypothesis (2) and also further illustrates the reasons for the occurrence of the phenomenon by situation.

5. Conclusion and recommendation

The results of this study showed that there was no significant difference in the frequency of “I” and “you” in Chinese students’ academic speaking, although “I” was still used more frequently than “you”. When the corpus was divided into interactive and monologic discourse patterns, “you” appeared quite frequently in the former, while “I” appeared more frequently in the latter. This is consistent with Morell’s (2001) results in the introduction, where “you” is more common in interactive lectures and “I” is more popular in non-interactive lectures. Meanwhile, the frequency of “I mean” and “you know” is lower than the overall frequency of MICASE, which indicates that Chinese students do not pay much attention to the use of discourse markers in their oral expressions. This may be due to the lack of proficiency in English as EFL or the neglect of interaction by reading from the script. Therefore, in the teaching of academic speaking for Chinese students, the instruction of pragmatic markers can be strengthened. And students are encouraged to go into the authentic English language environment and learn relevant pragmatic markers to make the speaking more fluent and fuller.

Meanwhile, this study has some limitations. Although MICASE is an official and authoritative corpus of spoken English for academic purposes, The University of Michigan English Language Institute last updated it in 2019, so the corpus lacks data from the last three years. However, considering the impact of the epidemic and the decrease in the number of Chinese students choosing to study abroad in the last three years, the data can still cover most Chinese students. In addition, this study considers 7 spoken items, which may not be a sufficient amount of research, but still a more general conclusion can be drawn through corpus analysis. Moreover, the data in this study are authentic in source and rigorous in processing, which can provide reference data for other studies in the same direction. In addition, the causes of the characteristics related to Chinese students’ academic speaking in English are worth studying. This may stem from Chinese students’ language acquisition process, language habits, readiness and identity perception. Future research could conduct in-depth qualitative studies on the causes of this language characteristic in order to help Chinese students better achieve academic English.

References

- [1] Biber, D., & Quirk, R. (2012). *Longman grammar of spoken and written English*. Longman.
- [2] Bourdieu, P., Passeron, J. C., Nice, R., Bourdieu, P., & Bottomore, T. B. (2014). *Reproduction in education, society, and culture*. Sage.
- [3] Braasch, A., Povlsen, C., & Nesi, H. (2002). An English spoken academic word list. In *Proceedings of the tenth EURALEX international congress* (Vol. 1, pp. 351–358). essay, Copenhagen.
- [4] Chafe, W. (1986). Academic Speaking. In *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society* (pp. 26–40). essay, Berkeley Linguistics Society.

- [5] Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, 7(2), 113–127. <https://doi.org/10.1093/applin/7.2.113>
- [6] Chen, C., Härdle, W. K., & Unwin, A. (2016). Computational Statistics and Data Visualization. In *Handbook of Data Visualization* (pp. 4–5). essay, Springer.
- [7] Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics*, 33(9), 1337–1359. [https://doi.org/10.1016/s0378-2166\(00\)00066-7](https://doi.org/10.1016/s0378-2166(00)00066-7)
- [8] Gómez, I. F. (2006). Interaction in academic spoken English: The use of ‘I’ and ‘you’ in the MICASE. *Educational Linguistics*, 35–51. https://doi.org/10.1007/978-0-387-28624-2_3
- [9] Hasko, V. (2020). Qualitative corpus analysis. *The Encyclopedia of Applied Linguistics*, 1–7. <https://doi.org/10.1002/9781405198431.wbeal0974.pub2>
- [10] Jenkins, J. (2014). *English as a lingua franca in the International University: The Politics of Academic English language policy*. Routledge.
- [11] Jenkins, J. (2014). *English as a lingua franca in the International University: The Politics of Academic English language policy*. Routledge.
- [12] Korthagen, F. A. J. (2004). In search of the essence of a good teacher: Towards a more holistic approach in teacher education. *Teaching and Teacher Education*, 20(1), 77–97. <https://doi.org/10.1016/j.tate.2003.10.002>
- [13] Rounds, P. L. (1987). Multifunctional personal pronoun uses in an educational setting. *English for Specific Purposes*, 6(1), 13–29. [https://doi.org/10.1016/0889-4906\(87\)90072-x](https://doi.org/10.1016/0889-4906(87)90072-x)
- [14] Schourup, L. (1999). Discourse markers. In *Lingua* (Vol. 107, pp. 227–265). essay, Amsterdam: North-Holland Pub. Co.
- [15] Stage, F. K., Manning, K., & Love, P. (2013). Document analysis. In *Research in the College Context: Approaches and Methods* (pp. 83–96). essay, Routledge.
- [16] The University of Michigan English Language Institute. (2019). *Micase online home page - University of Michigan*. Michigan Corpus of Academic Spoken English. Retrieved March 25, 2022, from <https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase>