Gendered Distinctions of Interjection Usage in the YouTube Context

Xiaohua Wang[1,a,*,†], Jingjing Zheng[2,b,†], Ziqi Li[3,c,†]

[1]School of International Studies, Hangzhou Normal University, Hangzhou, 311121, China
[2]School of Chengyi, Jimei University, Xiamen, 361021, China
[3]School of Foreign Languages, Shandong Normal University, Jinan, 250000, China
a. wangcharlotte236@gmail.com, b. jingjzh260@163.com, c. ziqili_sdnu @163.com
**corresponding author*
[†]Xiaohua Wang, Jingjing Zheng, and Ziqi Li contributed equally to this work and should be considered co-first authors.

Abstract: Linguistic gender research has originated from Western academia and gone through centennial and continual development throughout the birth and growth of linguistics. Lately, the interjection words, serving as an observable but marginal lexical index, have played an increasingly significant role in many aspects of linguistic studies for recent decades, for example, in the way the experts analyze the vocal expression of emotion, study the cross-lingual analysis in Japanese and English, and many other applications. This paper adopts three classifications of interjection words – "noise-like" primary interjections, "word-like" primary interjections, and secondary interjections – to further explore the relationship between gender and each classification of interjections on the platform of YouTube. The randomly chosen videos of YouTubers, fifty male YouTubers, and fifty female ones, are transcribed into words and analyzed systematically with the assistance of the structured questionnaire and the corpus analysis. Consequently and generally, male YouTubers using the interjections are significantly different from those female YouTubers on word frequency (p<0.05) which agreed with the data from "noise-like" primary interjections and secondary interjections but disagreed with "word-like" primary interjections. It suggests that gender does influence the common usage of the interjections significantly, but there are still some exceptions, like the "word-like" primary interjections.

Keywords: gender differences, interjection, social media, Corpus Linguistics

1. Introduction

In the digital age, YouTube has become a significant platform for global video sharing where users express their opinions and form a diverse community. Besides, the usage habits of interjection words on YouTube are also changing with the development of society. By studying gender differences of using interjections in YouTube, it can help understand online language habits, enrich linguistics knowledge, and promote harmony in online communities [1]. Cantina proposed an analysis of YouTube comments, stating that female users use more interjections when expressing emotions [2]. Colley and Todd proposed that male users tend to use exclamation marks when expressing anger and

excitement [3]. These studies provide us with a preliminary understanding of gender differences, but there are also certain limitations.

In this field, significant differences have been observed in the use of interjections between male and female English speakers. Corpus analysis also plays an important role in the fields of sociolinguistics and psycholinguistics. Through analyzing the language usage in different corpora, we can study issues such as social variations of language, social identity, language attitudes, and other aspects. Meanwhile, corpus analysis can also be used to study psycholinguistic issues such as language acquisition, the psychological mechanism of language processing, and language disorders. Corpus analysis provides us with a scientific, systematic, and objective research method, helping us better understand and apply languages.

With the development of computer technology and big data, the application scope of corpus analysis will be more extensive and will have a greater impact on the research of linguistics and related fields. This study utilized two quantitative research methods - questionnaire surveys and corpus analysis - to obtain objective and reproducible data. Due to the vast number of words in each category, it is impractical to cover all primary "noise-like" interjections, "word-like" primary interjections, and secondary interjections when studying gender differences. The questionnaire was designed using the Likert scale, and 140 native English-speaking respondents were surveyed using a quota sampling method. Furthermore, this article employs chi-square analysis to examine data such as the frequency of various exclamation words used by men and women.

This study adopted a more rigorous research method and considered more influencing factors when studying gender differences in interjections on the YouTube platform. The research findings from it indicate that gender differences are more complex in online communication, providing a new perspective of language communication, social power structures and natural language processing for us to understand the role of gender in the digital age.

Initially, the research significance of gender disparities in the utilization of interjections resides in its potential to illuminate the fundamental mechanisms of language and communication. Through scrutinizing the distinct ways in which men and women employ interjections, researchers can acquire insights into the social, cultural, and psychological factors that exert an influence on linguistic behavior. Comprehending these gender differences can have practical implications for domains such as education, psychology, and linguistics. Exploring gender-specific patterns in the usage of interjections might also contribute to a more profound understanding of social dynamics and power structures within interpersonal interactions. This research could furnish valuable information regarding how individuals express emotions, assert authority, or navigate conversational spaces in accordance with their gender identity. Additionally, it may offer cues about the role of societal norms and expectations in shaping linguistic practices. Moreover, investigating gender differences in interjections has implications for natural language processing (NLP) technologies and machine learning algorithms. By recognizing distinct patterns in how men and women use interjections across different contexts, NLP systems could be designed to better understand and respond to diverse communicative styles [4].

2. Literature Review

2.1. Gender linguistics

Gender linguistics is a disciplinary domain that conducts an in-depth exploration of how language mirrors and molds gender identities and roles. In this domain, certain eminent researchers have offered valuable perspectives on the intricate connection between gender and language through their pioneering endeavors. For instance, Judith Butler is renowned for her theory of gender permittivity, emphasizing that gender is constructed through continuous linguistic and bodily practices rather than

being inherent at birth. Additionally, Dana Harris, in her research, has disclosed the dynamic process of gendered language, namely, how gender is expressed and sustained through language utilization in diverse social and cultural contexts.

Kapoor believed that Gender linguistics, as an interdisciplinary field of study, not only focuses on how language reflects and shapes societal gender perceptions, but also explores how language practices can promote gender equality [5]. In everyday communication, people are gradually realizing that using more gender-neutral language can reduce the constraints on individual gender identities and help break traditional gender norms. For example, in the workplace environment, adopting more inclusive and equitable expressions can eliminate stereotypical impressions of specific gender roles, thereby promoting a more just and equal direction for organizational and societal structures [6].

Furthermore, as social attitudes and values continue to evolve, there are new demands for language usage. Stange stated that more and more scholars are paying attention to how language practices can challenge inherent, one-sided, or discriminatory cognitive models of gender roles in traditional cultures [7]. They are committed to finding and establishing a mode of communication that not only respects each person's self-identity but also avoids imposing any biases or stereotypes that may have negative effects on others.

Although these researchers have offered rich theoretical frameworks, their studies also have some limitations. For example, although Butler's theory profoundly challenges traditional gender concepts, it might overlook the role of biological factors in gender formation. Claidière hold the opinion that focusing on the process of engendering might not fully consider individual differences and atypical gender expressions [8].

Based on these studies, this research aims to explore another dimension of gender linguistics, namely how to identify and understand the language expressions of non-binary genders in different sociocultural environments. The perspective adopted in this research is based on the recognition that gender is a diverse concept that not only includes the traditional categories of male and female but also should encompass non-binary genders and other gender identities. The significance of this perspective lies in its challenge to the binary assumption of gender and the promotion of the recognition and respect for gender diversity. There are multiple reasons for adopting this perspective in this research. Firstly, as society's understanding of gender diversity deepens gradually, acknowledging and respecting the language expressions of non-binary genders has become an important social issue. Secondly, from an academic research perspective, including non-binary genders in the research scope of gender linguistics helps enrich our understanding of the relationship between gender and language. Finally, from a practical application perspective, providing appropriate language expressions and identity recognition for non-binary genders is of great significance for building an inclusive and equal society. Therefore, this research adopts an inclusive and non-binary perspective in the field of gender linguistics, aiming to provide new insights and thoughts for the theoretical framework and practical application of gender linguistics through the study of language expressions of non-binary genders.

2.2. Interjection words

Greek grammarians classified interjections as a branch of adverbs because they were similar in grammatical function to other adverbs, all having an impact on the "determination" of verbs, even at the level of understanding. For interjections or interectio, Latin grammarians regarded them as an independent word class. This transformative classification ensured that the number of "partes orationis" in Latin remained at the "magical number 8", replacing the article category in Greek, as Latin itself had no articles. Priscian discovered in his research that interjections did not need to be syntactically combined with other parts of the sentence, thus defining them as "pars orationis

significans mentis affectum voce incondite", that is, "a word class expressing emotions through non-fixed forms".

Syafiq indicated that the main interjections of "noise-like" words usually refer to words used in communication to express strong emotions or attitudes. Their pronunciation and meaning often differ from standard words. These interjections play a role in emphasizing and enhancing the tone of language communication, making the expressed emotions clearer and more prominent. For example, in Chinese, "āiyā", "òò", "hēngheng", etc. can all be regarded as interjections of "noise-like" words [9]. The main interjections of "word-like" words refer to words used in language expressions to indicate basic semantics such as affirmation, negation, surprise, and questioning.

Secondary interjections refer to interjections that are used relatively less in language communication but can still express certain emotions and semantics in specific situations. They may be some special words or variations of "noise-like" and "word-like" interjections. The use of secondary interjections is usually related to specific contexts, regional dialects, or personal habits. Kurpiel proposed that they are usually complete words with specific grammatical functions and semantic connotations. Compared with the interjections of "noise-like" words, these interjections are more standardized and closer to commonly used words in daily life [10]. For example, in Chinese, "ho", "bù", "jīngyà" and "ma", etc. can all be classified as the main interjections of "word-like" words.

However, these studies have not completely solved the problem of the definition of interjections. In the middle of the 20th century, the language philosopher John Searle proposed the "non-verbal" feature of interjections. He believed that the difference between interjections and other vocabularies was that interjections had no clear meaning but were non-verbal sounds used to express emotions. This view has, to a certain extent, enriched the study of interjections, but still has certain limitations. The historical evolution of the nature of interjections can be traced back to ancient language studies. In ancient Greece and Rome, linguists began to classify and study interjections. In conclusion, the definition of interjections has undergone a long exploration and evolution, and different scholars have put forward their own views from different perspectives. Based on a comprehensive analysis of existing research results, this study adopts the view that interjections are non-verbal sounds expressing emotions and attempts to further explore the nature and function of interjections on this basis.

3. Research

3.1. Research design

The quantitative methods, encompassing a questionnaire survey and corpus analysis were employed in this study to acquire objective and trustworthy analysis findings [11].

A questionnaire survey was conducted to determine the representative and typical interjection words as research targets since it is not practical to encompass all the words within each category when conducting a study on the gendered differences in the use of "noise-like" primary interjections, "word-like" primary interjections and secondary interjections. The questionnaire was designed using the Likert scale, and the 140 respondents were selected by quota sampling method.

The corpus was used to retrieve the frequency of target interjections already identified in previous questionnaires and to analyze the statistical relationships between gender and interjection frequency. The corpus was derived from 80 videos posted on YouTube which were selected using a purposive sampling method. A chi-square test was employed to analyze the corpus using IBM SPSS Statistics 27.

3.2. Research questions

This study aims to explore the gendered distinctions of interjection usage in the YouTube context. Two research questions were addressed in this survey.

Q1. With interjection words as an observable but marginal lexical phenomenon of gendered differences on YouTube, is there a significant difference in its usage between male and female English-speaking YouTubers in the learning area?

Q2. Is its significance related to the category of interjection words, i.e., "noise-like" primary interjections, "word-like" primary interjections, and secondary interjections?

3.3. Research participant and sampling

In the questionnaire survey, 140 participants, 69 women, and 71 men participated. A quota sampling was used to ensure the gender balance of respondents [12]. Respondents consisted of English speakers from diverse age groups, educational backgrounds, and geographical regions to guarantee the universality and representativeness of the results.

In the corpus study, the corpora were derived from 80 videos posted on YouTube. The videos, 40 of which were posted by male YouTubers and 40 by female YouTubers, were manually selected using purposive sampling to improve research efficiency and make it easier to access valuable corpus [12]. The criteria for video selections are as follows:

(1) Topic: The topics of the videos selected are study-related, which are emotionally neutral, minimizing the influence of the topic on the use of interjections.

(2) Video release date: The videos selected were released within the last ten years to guarantee the effective value of this research for contemporary society.

(3) Popularity: The selected videos all received more than 1,000 likes, which enables the collected corpora to be more likely to conform to the commonly used usages of interjections and such usages are acceptable to the public.

(4) Video duration: The lengths of the chosen videos all ranged from 4 to 20 minutes to guarantee the equilibrium of the corpus sizes between males and females, and to enhance the likelihood of the presence of abundant interjections in the chosen videos.

3.4. Research instrument

The questionnaire was designed based on the Likert scale which is a useful and reliable instrument for self-efficiency and can quantify the questionnaire results [13]. 24 questions were involved in the questionnaire, centered around the usage of pre-selected interjections, such as ha, oh, uh. These interjection words were chosen as typical ones according to the existing literature [14]. The designed questionnaire was published on Qualtrics, a globally available online survey tool, and was disseminated on mainstream websites such as Reddit, Twitter, and Hellotalk.

In the corpus study, Feishu, a transcription software, was used to transcribe the video corpora into text corpora, which improved the transcription efficiency. The transcribed texts were stored and retrieved in Antconc which is equipped with a powerful and convenient in-built retrieval function. The data derived from the corpus analysis was analyzed with the help of IBM SPSS Statistics 27 software.

3.5. Research procedure

In the questionnaire survey, respondents were presented with 24 questions such as "I use ha frequently in my daily life", and were requested to choose their level of agreement between strongly disagree (1) to strongly agree (5). A total of 140 valid responses were received. According to Table 1, 2, and 3,

the overall values of each interjection word were calculated and ranked. For each category of interjections, the top four highest-valuing ones, namely, the most frequently utilized ones, were chosen as research targets.

Table 1: The values of "noise-like" primary interjections

Interjection word	Overall value	Average value
Whoops	375	3.679
Woah	391	2.793
Uh	361	2.579
Oh	350	2.5
Er	343	2.45
Ha	314	2.243
Um	313	2.236
Hmm	305	2.179

Table 2: The values of "word-like" primary interjections

Interjection word	Overall value	Average value
Ahem	418	2.986
Boo	414	2.957
Hey	397	2.836
Yeah	394	2.814
Nope	372	2.657
Wow	358	2.557
Yay	351	2.507
Gosh	337	2.407

Table 3: The values of secondary interjections

Interjection word	Overall value	Average value
Alright (All right)	359	2.564
Shit	359	2.564
Right	358	2.557
Well	354	2.529
Fuck (-ing)	352	2.514
God	348	2.486
Ok (Okay)	339	2.414
Damn	339	2.414

Then, we built the corpora that were used to access the frequency of target interjections determined in the previous questionnaire survey. In the corpus construction process, we first transcribed the video corpora into text corpora using the transcription software, Feishu. Additionally, in order to ensure the accuracy and quality of the transcription, these texts were cross-checked manually to add interjection

words that were not recognized by the software and that were incorrectly recognized. After being reviewed and proofread, two corpora were built and stored in Antconc to conduct the research.

During the actual lookup process, it was found that certain target interjections had relatively low word frequencies, or were not presented in the corpora. Therefore, to render the analysis results more scientific and representative, a slight adjustment was made to the original word list in accordance with the actual usage of interjections in corpora. Table 4 shows the final retrieved word list:

Table 4: Two versions of the word list used for retrieval

The category of interjections	Interjection word (original)	Interjection word (revised)
"Noise-like" primary interjections	Whoops	Uh
	Woah	Oh
	Uh	Um
	Oh	Hmm
"Word-like" primary interjections	Ahem	Hey
	Boo	Yeah
	Hey	Wow
	Yeah	Gosh
Secondary interjections	Alright (All right)	Alright (All right)
	Shit	Right
	Right	Well
	Well	Ok (Okay)

The frequency of interjections in the corpora was fetched using Antconc. During the search process, words that are of the same form as interjections but not used as such were manually excluded to ensure the accurate identification of interjection frequency. The occurrence of each interjection within each video was counted and arranged in the format of an Excel sheet.

We used SPSS to statistically process the data. In this process, a chi-square method was adopted to analyze the gendered differences in the frequency of interjection usage, for variances among the study groups are not equal [15].

3.6. Research data

Data in this study derived from a questionnaire, and two corpora, 70626 tokens for males and 74640 tokens for females, designed to study the gender differences in the frequency of interjection usage. Thereafter, the gathered data were processed via a chi-square test.

3.7. Ethical Considerations

Strict ethical guidelines were followed throughout the whole study process. The anonymity and privacy of all respondents in the questionnaire survey were protected. For the video selections, only publicly available videos on YouTube were involved and no personally identifiable information was extracted in the corpora.

4. Results and Discussion

4.1. Data analysis

The first classification in the study is "noise-like" primary interjections which are phonologically aberrant, and thus sound like noises, such as "Uh" and "Um" [16]. The data shown by the chi-square test are tabulated in Tables 5 and 6:

Table 5: General Comparison of Male and Female YouTubers Using the "Noise-like" Primary Interjections on Word Frequency

Word Frequency	X^2	*DF*	*p-value*
The "noise-like" primary interjections	25.722	3	.000
Gender	12.920	1	.000
Word frequency total	188		

$p<0.05$

Table 5 emphasizes whether there is a significant difference in the "noise-like" primary interjection usage between YouTubers of different genders. Overall, 80 YouTube videos are selected about learning from an equal pool of male and female YouTubers and then transcribed into texts to be analyzed by word frequency, altogether appearing 188 times. Specifically, it shows that male YouTubers using the "noise-like" primary interjections were significantly different from those female YouTubers on word frequency ($X^2(1)=12.920$, $p<0.05$). Then it comes to the more concrete data in Table 6.

Table 6: Concrete Comparison of Male and Female YouTubers Using the "Noise-like" Primary Interjections on Word Frequency

Gender	Interjection	Word frequency	Word frequency ratio (in Male)	Word frequency ratio (in Male & Female)
Male	Uh	44	37.9%	91.7%
	Oh	29	25.0%	48.3%
	Um	40	34.5%	52.6%
	Hmm	3	2.6%	75.0%
	Total	116	100.0%	61.7%
Female	Uh	4	5.6%	8.3%
	Oh	31	43.1%	51.7%
	Um	36	50.0%	47.4%
	Hmm	1	1.4%	25.0%
	Total	72	100.0%	38.3%

Table 6 lists how often the male or female YouTubers use the above "noise-like" primary interjection – it is apparent that the average word frequency of the "noise-like" primary interjections used by male YouTubers was relatively higher than those of female YouTubers, differing by about 23.4%. These findings suggested that gender did influence the usage of the "noise-like" primary interjections significantly.

The next classification is the "word-like" primary interjections which are phonologically normal and still have the potential to be used as fundamental forms for normal words, such as "Hey" and "Wow" [16-17]. For example, "Wow" can serve as a verb to show how impressive the people are.

The above two classifications belong to the primary interjections which don't derive from present existing words. Tables 7 and 8 enumerate the data about "word-like" primary interjections:

Table 7: General Comparison of Male and Female YouTubers Using the "Word-like" Primary Interjections on Word Frequency

Word Frequency	X^2	*DF*	*p-value*
The "word-like" primary interjections	1.568	3	.667
Gender	1.005	1	.316
Word frequency total	132		

$p>0.05$

Table 7 underlines if there is a significant difference in the "word-like" primary interjection usage between YouTubers of different genders – it shows that male YouTubers using the "word-like" primary interjections were not significantly different from those female YouTubers on word frequency ($X^2(1)= 1.005$, $p>0.05$). More specific data are presented in Table 8:

Table 8: Concrete Comparison of Male and Female YouTubers Using the "Word-like" Primary Interjections on Word Frequency

Gender	Interjection	Word frequency	Word frequency ratio (in Male)	Word frequency ratio (in Male & Female)
Male	Hey	12	27.9%	37.5%
	Yeah	28	65.1%	31.8%
	Wow	2	4.7%	40.0%
	Gosh	1	2.3%	14.3%
	Total	43	100.0%	32.6%
Female	Hey	20	22.5%	62.5%
	Yeah	60	67.4%	68.2%
	Wow	3	3.4%	60.0%
	Gosh	6	6.7%	85.7%
	Total	89	100.0%	67.4%

According to Table 8, inspections of these two different gender groups indicate that the average word frequency of the "word-like" primary interjections used by male YouTubers was approximately 34.8% lower than those of female YouTubers. These findings suggested that gender did not influence the usage of the "word-like" primary interjections significantly.

The third classification is the secondary interjections. Contrary to the primary interjections, they originate from regular words, mostly nouns and verbs, and come to be used as interjections by virtue of semantics [16-17]. Through the chi-square test, the data are analyzed and tabulated in Tables 9 and 10:

Table 9: General Comparison of Male and Female YouTubers Using the Secondary Interjections on Word Frequency

Word Frequency	X^2	*DF*	*p-value*
The secondary interjections	28.580	3	.000
Gender	19.917	1	.000
Word frequency total	296		

$p<0.05$

Table 9 stresses whether there is a significant difference in the secondary interjection usage between YouTubers of different genders – it reveals that male YouTubers using the secondary interjections were significantly different from those female YouTubers on word frequency ($X^2(1)= 19.917$, $p<0.05$). Here are more concrete data surfaced in Table 10:

Table 10: Concrete Comparison of Male and Female YouTubers Using the Secondary Interjections on Word Frequency

Gender	Interjection	Word frequency	Word frequency ratio (in Male)	Word frequency ratio (in Male & Female)
Male	Alright	42	22.2%	73.7%
	Right	52	27.5%	80.0%
	Well	38	20.1%	74.5%
	OK	57	30.2%	46.3%
	Total	189	100.0%	63.9%
Female	Alright	15	14.0%	26.3%
	Right	13	12.1%	20.0%
	Well	13	12.1%	25.5%
	OK	66	61.7%	53.7%
	Total	107	100.0%	36.1%

Based on Table 10, it is obvious that the average word frequency of the secondary interjections used by male YouTubers was approximately 27.8% higher than those of female YouTubers. These findings suggested that gender did influence the usage of the secondary interjections significantly.

Ultimately, Table 11 was drawn to have an overview of contrasting male and female YouTubers using interjections on word frequency. Compared to Tables 5, 7, and 9, Table 11 tended to figure out if gender could have a significant impact on the usage of the interjection words, covering the primary and secondary interjections.

Table 11: General Comparison of Male and Female YouTubers Using Interjections on Word Frequency

Word Frequency	X^2	*DF*	*p-value*
The interjections	22.531	3	.000
Gender	11.653	1	.001
Word frequency total	616		

$p<0.05$

Table 11 shows that male YouTubers using the interjections were significantly different from those female YouTubers on word frequency ($X^2(1)=11.653$, $p<0.05$). These findings suggested that gender did influence the usage of the interjections significantly.

Given these findings, it's clear that our results verify our initial hypotheses in the general aspect. Specifically, results from the "noise-like" primary interjection and the secondary interjection aspect do verify our hypotheses whereas results from the "word-like" primary interjection do falsify them. Therefore, some improvements are still required to better our research.

4.2. Discussion

Generally, this study concluded and suggested that gender did significantly influence the usage of interjections, particularly the "noise-like" primary interjection and the secondary interjection, among YouTubers on the topic of learning. However, there is an important difference in the usage of the "word-like" primary interjection, which didn't show a significant relation between gender and the usage of itself.

Current linguistic researchers are inclined to conclude that gender has played a crucial role in the process of communication [18-19]. More relevant research proposed by Chapman, Gardner, and Lyons presents that when rating positive words and neutral words, there is no significant difference between the genders; however, when rating negative words, it appears that negative words are continuously rated more negatively by young women than by men [20]. This study, to some extent, corresponds to the previous ones since it has generally recognized the significant influence of gender in the field of interjection usage, which is supported by the findings from the "noise-like" primary interjection and the secondary interjection respects, but objected by the findings from the "word-like" primary interjection respect.

These results provide some practical and theoretical insights into the nuanced relationship between gender and interjection usage, elaborately practical insights for communication and education, and theoretical insights for linguistics and sociolinguistics. Initially, understanding how different types of interjections are influenced by gender can offer some refinements for the communicative strategies and the approaches to communication training, because interjections are a critical component of language learning environments. Besides, these findings can contribute to the broader field of linguistics and sociolinguistics by suggesting that while some linguistic behaviors might show gender differences, others may not. This subtle interpretation can lead to more refined and retailed theories by adding how interjection words are used across different social gender groups.

According to this study, the possible factors that contribute to what we observed may be the handpicked research participants and interjection words, and the limited research samples. For the research participants whom we selected, there exist some individual differences, including educational background, speaking habits, and so on. Furthermore, though the investigated interjections are from our extensive online searches and rigorous public questionnaires, they are still restricted and give some limitations to this article. Additionally, it is about our limited research samples – we respectively chose 40 corpuses each for male and female. This sample size, 80 materials in total, may not be large enough, and thus it may lead to inevitable errors in our results.

5. Conclusion

In this research, we studied the gendered differences in interjection usage in the YouTube context. The key findings reveal that, in general, there is a significant difference in the frequency of interjection usage between male and female YouTubers in the learning area, and its degree of significance is related to the category of interjections. Specifically, the average frequency of "noise-like" primary interjections and secondary interjections usage for males is significantly higher than

that for females. However, no significant difference was observed in the frequency of "word-like" primary interjection usage between males and females.

This study is noteworthy for its combination of a questionnaire survey and a corpus study, as well as its use of chi-square analysis for data processing. This comprehensive approach offers a more accurate and in-depth understanding of this research topic. In addition, while previous studies approached the gendered characteristics of interjection usage from the perspective of how gender influences the use of interjections with different emotional colors, this research adopts an alternative classification of interjections: "noise-like" primary interjections, "word-like" primary interjections, and secondary interjections, to investigate the impact of gender on their usage. This novel aspect lays the groundwork for constructing a more precise and detailed language theory framework, thereby advancing the development of gendered language.

Theoretically, the findings reported here enrich linguistic and sociolinguistic research, leading to deeper and more comprehensive thinking about the relationship between language behavior and gender. Practically, the study contributes to improving communication training approaches and communicative tactics.

Looking forward, future research could expand the sample size to enhance the generalizability of the results. Additionally, exploring the impact of social factors, such as educational background and age groups, on the degree of gender differences in interjection usage would be a promising direction. This would provide a more comprehensive understanding of the complex interplay between gender and language in a social context.

Acknowledgement

Xiaohua Wang, Jingjing Zheng, and Ziqi Li contributed equally to this work and should be considered co-first authors.

References

[1] *Cheng, N., Chandramouli, R., Subbalakshmi, K. (2011) Author gender identification from text. Digital Investigation, 8: 78–88.*

[2] *Cantina, J. M., Alvarico, P. L. (2022) Gender And Language Variation On The Comments Of Viral YouTube Videos. Sprin Journal of Arts. Humanities and Social Sciences, 1: 341-359.*

[3] *Colley, A., Todd, Z. (2002) Gender-Linked differences in the style and content of E-Mails to friends. Journal of Language and Social Psychology, 21: 380–392.*

[4] *Jeong-Hwa, L., Cha, K. W. (2020) An analysis of the errors in the auto-generated captions of university commencement speeches on YouTube. Journal of Asia TEFL, 17: 143.*

[5] *Kapoor, S. (2022) "Don't act like a Sati-Savitri!": Hinglish and other impoliteness strategies in Indian YouTube comments. Journal of Pragmatics, 189: 4–16.*

[6] *Rubin, D. L., Greene, K. (1992) Gender-Typical style in written language. Research in the Teaching of English, 26: 7–40.*

[7] *Stange, U. (2019) The social life of emotive interjections in spoken British English. Skandinaviske Sprogstudier, 10 : 174–193.*

[8] *Claidière, N., André, J. B. (2012) The transmission of genes and culture: A questionable analogy. Evolutionary Biology, 39: 12-24.*

[9] *Syafiq, B. Z. M., Bakar, K. A. (2024) Language shaming on YouTube: linguistic features, themes, and social implications. Journal of Research in Applied Linguistics, 15: 128-144.*

[10] *Waseleski, C. (2006) Gender and the Use of Exclamation Points in Computer-Mediated Communication: An analysis of exclamations posted to two electronic discussion lists. Journal of Computer-mediated Communication, 11: 1012–1024.*

[11] *Sandelowski, M. (2000) Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. Research in nursing & health, 23: 246-255.*

[12] *Acharya, A. S., Prakash, A., Saxena, P., Nigam, A. (2013) Sampling: Why and how of it. Indian journal of medical specialties, 4: 330-333.*

[13] Maurer, T. J., Pierce, H. R. (1998) A comparison of Likert scale and traditional measures of selfefficacy. Journal of Applied Psychology, 83: 324-329.
[14] Norrick, N. R. (2009) Interjections as pragmatic markers. Journal of Pragmatics, 41: 866–891.
[15] McHugh, M. L. (2013) The chi-square test of independence. Biochemia medica, 23: 143-149.
[16] Goddard, C. (2013) Interjections and Emotion (with Special Reference to "Surprise" and "Disgust"). Emotion Review, 6: 53-63.
[17] Ameka, F. K. (1992) Interjections: The universal yet neglected part of speech. Journal of Pragmatics, 18: 101–118.
[18] Kazemian, R. (2022) Gender Identity in Social Media: A Corpus-based Study of Metadiscourse Features. Corpus Linguistics Research, 7: 97-125.
[19] Zhi, Y., Shu, B., Luo, L., Wang, J. (2022) Gender and linguistic distinctions of information needs in the Chinese context: Conscious and formalized levels. Library and Information Science Research, 44: 101195.
[20] Chapman, R. M., Gardner, M. N., Lyons, M. (2022) Gender differences in emotional connotative meaning of words measured by Osgood's semantic differential techniques in young adults. Humanities and Social Communications, 9: 119-129.