

Original Paper

Register Variation in Chinese English and Language Variation between Chinese English and American English: A Case Study of *Perhaps* and *Maybe*

Yun Xia^{1*} & Qian Li²

¹ Nanfang College of Sun Yat-Sen University, Guangzhou 510970, China

² Guangdong University of Foreign Studies, Guangzhou 510420, China

* Yun Xia, Nanfang College of Sun Yat-Sen University, Guangzhou 510970, China

Received: July 18, 2020

Accepted: July 27, 2020

Online Published: August 8, 2020

doi:10.22158/selt.v8n3p118

URL: <http://dx.doi.org/10.22158/selt.v8n3p118>

Abstract

The study aims to explore the register variation in Chinese English and language variation between Chinese English and American English. A corpus-based and comparative methodology was used to analyse the discourse features of Chinese English in the use of the lexical items perhaps and maybe. The major findings of the study can be stated as follows: 1) the more formal word perhaps is used more frequently than the informal word maybe in all the four genres in Chinese English. This shows that the text of Chinese English is generally in a more formal style. 2) In the Chinese English text, the ratios of the standard frequency of perhaps to maybe are greater than those in American English in the all the four genres. This indicates that the text in Chinese English is generally in a more formal style than that in American English. 3) In the Chinese English text, the informal word maybe is used less frequently than in the American English text. This is a sign that Chinese English is more formal than American English.

Keywords

register variation, language variation, corpus linguistics, discourse analysis, Chinese English

1. Introduction

The English discourse studies at home and abroad center on the native speakers' English with little attention given to other varieties of English. From their research content, the discourse studies on Chinese English can be roughly classified into three groups: cohesion and coherence, discourse structure and genre.

Cohesion and coherence are the main topics of Chinese English discourse. Duan (2006) makes a contrastive study of the differences between English and Chinese in the use of conjunctions, and argues that the use of English conjunctions by Chinese speakers is influenced by their mother tongue. Li (2006) demonstrates the discourse features of Chinese English in the lexical, syntactic and semantic connection with examples. Xu (2010) examines the discourse characteristics of Chinese English from three perspectives: cohesion, coherence, and schema through his collection of data. Some scholars investigate the discourse organization structure (Li, 2006) and the discourse organization style (Zhang & Ning, 2005) through a comparison between Chinese and English. Shen et al. (2016) analyze the similarities and differences between Chinese English and native English in the organization of discourse structure in the text of letters and argumentative essays.

In a word, the above studies on Chinese English discourse are based either on a single text or a small number of texts, among which Xu's study (2010) involved the largest number of texts, but only 20 newspaper articles and 12 novels were included. In other words, none of these studies used large corpora.

Corpora are beneficial to discourse analysis. Due to limited number of texts used in their research, researchers in the past had to investigate certain discourse features by analyzing very few example sentences from the texts. Consequently, their conclusions are not universal. For example, the characteristics of certain lexical or grammatical cohesion drawn from an individual text may not exist in other texts. A corpus-based method enables researchers to extract language use from a large number of representative corpus data through existing corpus data retrieval tools, on the basis of which common Chinese English discourse features can be identified and summarized.

As the core features of discourse, "coherence and cohesion, contextual characteristics, and interactivity" (Liang et al., 2010, p. 214) are the main content of corpus-based discourse studies. In addition, a corpus-based method is also suitable for the study of discourse features such as "register variation, interdisciplinary academic discourse contrast and corpus stylistics" (Liang et al., 2010, p. 215). Therefore, this paper will examine the use of the two lexical units *perhaps* and *maybe* in the Written Corpus of Chinese English (WCCE) and the Corpus of Contemporary American English (COCA) to explore their discourse features as regards register variation and language variation.

2. Register Variation of *Perhaps* and *Maybe* in Chinese English

Register variation refers to the differences in discourse features of a lexical unit in text of different genres. This paper mainly focuses on the variation in the use of the lexical units of *perhaps* and *maybe* in the four genres of newspapers, magazines, fictions, and academics. Genre in discourse studies can refer to "the oral style, written style, or other styles such as news, fictions, academics, sometimes the discourse types such as legal documents, academic papers, cover letters. In academic discourse, it can also refer to the micro-discourse types such as abstracts, introductions, research methodologies,

discussions, conclusions and acknowledgments” (Xu, 2019, pp. 19-20). Genre in corpus linguistics refers to five different types of text: spoken language, fictions, newspapers, magazines and academics. The present research uses the Written Corpus of Chinese English (WCCE) as the observed corpus and the Corpus of Contemporary American English (COCA) as the reference corpus. WCCE collects written text from the following four genres: magazine, newspaper, fiction and academic. It has the same genres as (COCA) except the spoken parts. Table 1 shows the data of the two corpora.

Table 1. Data of WCCE and COCA

	WCCE	COCA
Language	Chinese English	American English
Time	Synchronic	Synchronic
Text type	Written	Spoken and Written
Genre	Newspaper, magazine, fiction, academic	Spoken, newspaper, magazine, fiction, academic
Token	13,962,102	560,000,000

From Table 1, it can be seen that the two corpora are comparable in terms of their size, genre, time and text type.

Table 2 shows the original frequency and standard frequency of *perhaps* and *maybe* in the four different genres in WCCE, where standard frequency refers to the frequency of occurrence in every million tokens.

Table 2. Frequency of *Perhaps* and *Maybe* in WCCE

Words	Newspapers		Magazines		Fictions		Academics	
	Original frequency	Standard Frequency	Original frequency	Standard Frequency	Original frequency	Standard Frequency	Original frequency	Standard Frequency
perhaps	340	24.36	123	8.81	892	63.90	318	22.78
maybe	200	14.32	28	2.01	588	42.12	20	1.43
Ratio	1.79		4.4		1.5		15.9	

In Table 2, the frequencies of *perhaps* are higher than those of *maybe* in all the four genres, which is the most prominent discourse feature of the two lexical units. In other words, in written Chinese English, the word *perhaps* is not only used more frequently than *maybe* in academics, newspapers and magazines, but also more frequently in informal fictions. Therefore, we can draw the conclusion that Chinese English is generally written in a formal way.

There are two reasons for that. First, *maybe* is a spoken word, which is frequently used in spoken language, while WCCE is a written corpus, which results in the lower frequency of *maybe* in the corpus.

On the contrary, *perhaps* is a relatively formal word, which appears more frequently in written English. The overall frequency of *maybe* is lower than that of *perhaps* in all genres except for the spoken language. It follows that the frequency of *perhaps* is higher than that of *maybe*, which accords with the overall discourse characteristics of written Chinese English. Second, in the field of English teaching, teachers often complain that Chinese English learners often fails to write their English writings in an appropriate style. No matter in formal or informal writing, they tend to use more formal words. This writing style of Chinese English learners may continue to appear in the text of Chinese English. As a result, in the text of Chinese English, the word *perhaps* has a greater frequency than the word *maybe* in various genres.

The second remarkable register variation feature in Table 2 is that among all the four genres, the ratio of standard frequency of *perhaps* and *maybe* is the highest in academics. It is 15.9 (22.78:1.43), which is much higher than that of the other three genres. This shows that speakers of Chinese English tend to use more formal words in their academic writings. Besides, due to the large difference in the frequency ratio between the two lexical units, it may also suggest that Chinese English academic writing presents a more serious but less flexible stylistic feature. In the academic genre, we randomly selected some concordance lines containing *maybe*, and found that it appears more in some relatively relaxed topics, such registers as culture, history, etc. Below are the example sentences taken from WCCE.

Ex.1 His poems are somewhat simple, *maybe* even a little crude, in language, but are heroic and forceful in manner in expressing his personal ambitions.

Ex. 2 Whenever I go to Beijing, I go and visit the palace. *Maybe*, this influence came from the book *A Dream Back to Qing Dynasty* that I finished in a single reading.

Ex. 3 “He never added funds to his accounts. He just kept buying new phone numbers -- *maybe* for some telemarketing stuff so that people can’t track him down”.

Ex. 4 Walking through the big courtyard at night, *maybe* you want to join the party of the bands.

Ex. 5 The one named by craftsmen and ordinary people was earthenware pot Liu (now Dashaguo) Hutong, *maybe* there once lived a Mr. Liu who sold earthenware pots.

In the above examples, Example 1 was from an article introducing the history of Chinese literature. Although it belongs to academics, it is relatively less formal, that’s why the informal word *maybe* is used. Example 2 came from an article about architecture. This sentence is colloquial. Besides, the first person pronoun is used in the sentence. Therefore, the use of *maybe* is consistent with the style of the whole discourse. Example 3 from an article about political life is a quotation and is quite colloquial. An empathic pronoun *you* in Example 4 contributes to shortening the distance between the author and the reader so that *maybe* is compatible with the register. Example 5 comes from an article introducing Beijing architecture. Similarly, the use of the informal word *maybe* may make the buildings more attractive to the readers. Combined with the above examples, we can conclude that in the academic articles, sometimes colloquial words, such as *maybe*, can achieve better results. It follows that we should choose the appropriate style according to the context.

The third noteworthy feature of register variation is the lowest ratio of standard frequency between *perhaps* and *maybe* in the fiction genre, which is 1.5 (63.90:43.12). This shows that fictions are the genre in which informal words are used more frequently than in other genres. It is in line with the requirements of fiction writing. First of all, fictions need to be close to life and their readers, so the style is not as formal as that in academic or magazine articles. In addition, fictions may include a large number of dialogues which are the written records of the spoken language. We find that the word *maybe* was used in many dialogues in the fictions in WCCE. Here are some of the examples taken from WCCE.

Ex. 6 “What can I do, eh?” asked Nan. “*Maybe* go to school again?”

Ex. 7 “Why’d you cook it if you knew it was already dead?” “I thought...*maybe* only just die. *Maybe* taste not too bad. But I can smell, dead taste, not firm”.

Ex. 8 “Does he own that whole building?” “I don’t know. *Maybe* so”.

Ex. 6 is a conversation between two friends talking about their future plans. In the conversation, Nan asked his friends for advice, and her friend gave her the advice in the form of a question containing the word *maybe*. Both sides are in an informal and casual register where the speaker uses the spoken word “eh” and the hearer uses the interrogative to make suggestions. Ex. 7 is a conversation between the mother and her child about cooking a dead crab. The conversation is between family members. Both are closely related, so the hearer used a series of sentences with incomplete structure to answer the speaker’s questions. Ex. 8 is a dialogue between brother and sister, so both sides communicated in an informal way. These examples show that informal register is more frequently used in fictions than in the other three genres, which is determined by the unique textual features of fictions.

The fourth salient feature of register variation is that the ratio of standard frequency between *perhaps* and *maybe* is basically the same in newspaper and fiction genre. Both are relatively low with the ratio in newspaper genre being slightly higher. This shows that in Chinese English, the newspaper genre is as informal as the fiction genre. By examining the actual examples in the corpus, we find that the word *maybe* mostly appears in quotations in the newspaper text. Below are some examples extracted from WCCE.

Ex. 9 “*Maybe* the upward turning point of the economy still has to wait”, Lu said.

Ex. 10 “*Maybe* it wasn’t a problem for our grandparents to have a lot of children, but times have changed, and the gap between rich and poor is getting too wide”, she said.

Ex. 11 “But no one showed up, *maybe* because the bitcoin’s value has surged recently”, said Wang Feilang, vice-president at Shanda Tiandi, a cultural property development arm under the Shanda Group.

Ex. 12 “I always think you can improve. My ambition is always to be the best, although *maybe* you reach a level that is difficult to improve on. *Maybe* at the moment I have better prepared and it could be one of the best moments in my career. I am in great form and I hope to stay like that until the end of my career”, he commented.

From the above examples, we can see that in order to improve the objectivity of news reports and the accuracy of information, newspaper articles often quote the original words from the source of information. Consequently, the text in the newspaper genre tends to become colloquial.

3. Register Variation of *Perhaps* and *Maybe* in American English

COCA is composed of spoken language, newspaper, magazine, fiction and academic texts, with a total of 560 million tokens, evenly distributed among the five genres. Table 3 lists the original frequency and standard frequency of *perhaps* and *maybe* in COCA.

Table 3. Frequency of *Perhaps* and *Maybe* in COCA

words	Spoken		Newspapers		Magazines		Fictions		Academics	
	Original frequency	Standard frequency	Original frequency	Standard frequency	Original frequency	Standard frequency	Original frequency	Standard frequency	Original frequency	Standard frequency
Perhaps	20391	36.09	17252	30.53	23838	42.19	29091	51.49	25953	45.93
Maybe	50999	90.26	17211	30.46	18897	33.45	69275	122.61	3374	5.97
Ratio	0.4		1		1.3		0.4		7.7	

From Table 3, the following characteristics of register variation between *perhaps* and *maybe* in American English could be drawn. First, in the academic genre, *perhaps* is used 7.7 times more frequently than *maybe*, which suggests that *perhaps* is a more formal word mainly used in formal style, while *maybe* is seldom used in academic genre. Second, the spoken language genre has the lowest proportion of standard frequency of *perhaps* to *maybe*, being 40%, which suggests that *maybe* is an informal word mainly used in the informal register. Thirdly, although the proportion of standard frequency of *perhaps* to *maybe* between the fiction and spoken language is both 40%, the standard frequency of *maybe* in the fiction genre (122.61) is higher than that in the spoken genre (90.26), which may indicate that the fiction genre tends to be a more informal stylistic. Fourthly, the proportion of standard frequency of *perhaps* to *maybe* in newspapers and magazines is similar (100% vs. 130%), which indicates that these two genres share similar register, but the latter is higher. So the magazine genre is more formal than the newspaper genre.

4. Language Variation between Chinese English and American English

In this paper, we have examined the register variation characteristics of *perhaps* and *maybe* in Chinese English and American English respectively. Next, we will make a contrastive study of their register variation between these two English varieties. In order to investigate the differences more clearly and directly, Table 4 lists their standard frequencies and ratios in Chinese English and American English.

Table 4. Standard Frequency and Ratio of *Perhaps* and *Maybe* in Chinese English and American English

Corpus	word	Newspaper		Magazine		Fiction		Academics	
		Standard frequency	ratio	Standard frequency	ratio	Standard frequency	ratio	Standard frequency	ratio
WCCE	perhaps	24.36	1.7	8.81	4.4	63.90	1.599	22.78	15.9
	maybe	14.32		2.01		42.12		1.43	
COCA	perhaps	30.53	1	42.19	1.3	51.49	0.4	45.93	7.7
	maybe	30.46		33.45		122.61		5.97	

By comparing the data in Table 4, the main features in language variation could be concluded as follows. First, in the fiction genre, the ratio of *perhaps* to *maybe* in WCCE outnumbers that in COCA by four to one on average (1.599/0.4). And the ratio of the standard frequency of *maybe* in American English is 290% (122.61/42.12) as big as that in Chinese English, while the standard frequency of *perhaps* in the two varieties does not show prominent distinction, which suggests that *maybe* is more frequently used in the American English than in Chinese English. In other words, it might reflect that American English tends to be in a more informal style in the fiction genre.

Second, in the magazine genre, the ratio of *perhaps* to *maybe* in Chinese English and American English is 3.4 (4.4/1.3), which ranks in the second place in the difference between the two varieties. As Table 4 indicates that the standard frequencies of *perhaps* and *maybe* in American English are both higher in the magazine genre than those in Chinese English. Since the standard frequency of *maybe* in American English is quite higher than that in Chinese English, which is 16.5 times as big as that of the latter, American English is more informal in the magazine genre. This finding corroborates the previous study that in American English, “more informal and spoken-like language has seeped into the general writings of the magazine” (Lindquist, 2009, p. 61).

Thirdly, in the academic genre the difference in the use of *perhaps* and *maybe* in Chinese English and American English is quite significant as the ratio of the standard frequency of *perhaps* to *maybe* in Chinese English and American English is about 2.1 (15.9/7.7). In the academic genre *maybe* is used less frequently in Chinese English than in American English. Therefore, it could be concluded that American English is in a more informal style in the academic genre than Chinese English.

Fourth, in the newspaper genre the ratio of standard frequency of *perhaps* to *maybe* in Chinese English and American English is the lowest of all the four genres, which is 1.7 (1.7/1). That is to say, it is slightly higher in Chinese English than in American English. This shows that although the register of Chinese English in the newspaper genre is more formal than that of American English, the register of newspaper genre is the closest to that of American English in all the four genres.

In a word, the ratio values of the standard frequency of *perhaps* to *maybe* in Chinese English are greater than those in American English in the all the four genres: fiction, academics, magazine and newspaper, which indicates that the text in Chinese English is in a more formal style than that in American English in general. At the same time, it is found from Table 4 that in Chinese English the standard frequencies of *maybe* are lower than those in American English in all the four genres, which in other words shows that Chinese English users tend to use the informal word *maybe* less frequently in their text than American English users. As a result, its register is generally more formal than that of American English.

5. Conclusions

This paper investigates the frequencies and distribution of *perhaps* and *maybe* in different genres in WCCE and COCA. Based on the statistical data, it analyzes the register variation of *perhaps* and *maybe* in different genres, as well as the comparison of the language variation between Chinese English and American English. Generally speaking, Chinese English is more formal in register, and also its style is more formal than that of American English in all genres.

The main contributions of this study are as follows. First, it shed new light on register variation in Chinese English. In the field of Chinese English, no literature has been found on discourse analysis of Chinese English with authentic and representative data from a corpus of Chinese English. Second, it contrasted and summarized the discourse features of Chinese English and American English in the use of *perhaps* and *maybe*, which contributes to the study of World Englishes. Third, it has made a discourse analysis of Chinese English and American English based on corpora, which is a methodological contribution to the study of Chinese English. Fourth, the paper investigated the discourse features of Chinese English which is meaningful to Natural Language Processing with regard to the understanding of natural language.

Acknowledgments

This work was supported by the Humanities and Social Science Planning and Funding Office of Ministry of Education of the People's Republic of China (Grant reference 15YJA740048: China English or Chinglish: A Study Based on China English Corpus), and the Department of Education of Guangdong Province (grant reference 2017WTSCX32: A Study on the Definitions of Chinese-English Dictionaries for CFL Learners). We would like to express our gratitude to them for their support.

References

- Duan, L. (2006). A study of "Chinese English" from the perspective of Chinese and Western cultures and thoughts. *J. of Hangzhou Teachers' College*, 3, 261-264.
- Li, S. (2006). *Chinese English in the prospect of English globalization and localization*. Yinchuan: Ningxia People's Press.

- Liang, M., Li, W., & Xu, J. (2010). *An Introduction to Corpus Application*. Beijing: Foreign Languages Teaching and Research Press.
- Lindquist, H. (2009). *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.
- Shen, Y., Zhang, Y., & He, Y. (2016). *A Hundred Years of English Nativization in China*. Beijing: Qinghua University Press.
- Xu, J. (2019). *Corpora and Discourse Study*. Beijing: Foreign Languages Teaching and Research Press.
- Xu, Z. (2010). *Chinese English: Features and Implications*. Hong Kong: Open University of Hong Kong Press.
- Zhang, F., & Ning, H. (2005). Cultural and linguistic regeneration of Chinese English. *J. of Yantai Teachers' College*, 1, 91-93.