languages

MDPI

Article

"She's Like Why You Speak English While Dreaming?": A Corpus-Based Study of Quotative Markers Used by Chinese Speakers of L2 English

Delin Deng

Department of Linguistics, University of Florida, Gainesville, FL 32611, USA; ddeng@ufl.edu

Abstract: Based on sociolinguistic data collected from 29 Chinese L1 speakers of English in the US, this article investigated the quotative markers used by non-native speakers in their direct speech reporting. By conducting regression analysis, both linguistic factors (the tense of the verb, the subject pronouns, the mimetic status of the quotation, the content of the quotation, and the presence of discourse markers) and extralinguistic factors (gender, extracurricular contact with native speakers, and length of stay in the target country) were examined and further discussed. Our results revealed that factors such as the content of the quotation, the tense of the verb, and the presence of other discourse markers significantly influence the choice of quotative markers in non-native speech. We demonstrated that non-native speakers use a variety of quotative verbs in their direct speech reporting. Nonetheless, because of L1 impact, they rely primarily on the dominant form *say*. We also noticed that while non-native speakers might not be accurate in their words when reporting direct speech, they use prosodic cues to make their storytelling more convincing and authentic.

Keywords: quotative markers; direct speech reporting; non-native English; corpus

check for updates

Citation: Deng, Delin. 2023. *"She's Like Why You Speak English While Dreaming?"*: A Corpus-Based Study of Quotative Markers Used by Chinese Speakers of L2 English. *Languages* 8: 51. https://doi.org/10.3390/languages8010051

Academic Editor: Virginia Yip

Received: 27 October 2022
Revised: 24 January 2023
Accepted: 3 February 2023
Published: 8 February 2023

Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Reported speech is a well-studied phenomenon in linguistics. Traditionally, two types of reported speech are distinguished: direct speech reporting and indirect speech reporting. Direct speech corresponds to the original words being quoted, while indirect speech is the transformation of direct speech by changing the temporal adverbs, pronouns, and/or wording of the direct speech. For the current work, I will only focus on direct speech reporting.

When introducing the reported content, different verbal forms can be used. Here are some examples in English[1]:

(1) People just ***say***: "Do you have one dollar or two dollars? I need it for food." (DYJ_2018)
(2) You will just ***think***: "Ok, let me to go to the United States." (FYL_2018)
(3) He ***asked*** me: "Can I rob you?" (JP_2018)
(4) But her father ***told*** her: "No no no. Just follow your boyfriend." (WHQ_2018)
(5) A lot of them will ***go***: "Holly!" (ZJS_2018)
(6) She's ***like***: "why you speak English while dreaming?" (HQ_2018)

However, it can also be the case that no verbal form is present. The speakers use other acoustic cues, such as a longer pause before the embedded clause or a sudden change in pitch in the embedded clause, to mark the quoted content. For example:

(7) Oh hmm, yes, they do. But in a good way **Ø:** "Oh you are from China? It's cool." (HQ_2018)

As a result, in this work, we define quotative markers as any grammatical device used to introduce quoted speech. The quoted speech here includes both verbally uttered words and one's internal thoughts.

While most of the work that has so far been performed on quotative markers concentrates on the study of the use of one or several particular quotative marker(s) or the functional description of quotative markers (see, for example, Levey et al. 2013; Diskin 2017; Davydova 2019; Magliacane and Howard 2019, etc.), the contribution of the current work is trifold: first, by taking the variationist point of view in its extended version, which involves examining different variants sharing the same function in discourse, we explored the quotative verbal forms used by non-native speakers in informal settings and their distribution. Second, we complement the existing literature by providing data on Chinese L1 speakers of English in the US. Third, we opened a critical discussion between the functions of quotative markers and their importance in social interaction, which could further our understanding of the acquisition of some socio-pragmatic competence by L2 speakers in the target country. As different quotative markers might be acquired at different stages during the learning process, such analysis allows us to see how different factors could play together to affect this acquisition.

Here, we will attempt to answer the following questions: Is the use of quotative markers by Chinese L1 speakers of English as diverse as that by native speakers in their direct speech reporting, or do they only rely on one dominant form? Do Chinese L1 speakers use innovative quotative markers (see, for example, Labov et al. 2008), such as *be like* in English? What linguistic factors could impact their use of quotative markers, such as the tense of the verb, the subject pronouns, the mimetic status of the quotation, the content of the quotation, the presence of discourse markers, etc.? Is there any influence of extralinguistic factors on their use of quotative markers, such as gender, extracurricular contact with native speakers, length of stay in the target country, etc.?

Therefore, the structure of the current work will be laid out as follows: in Section 2, we will review the previous work on quotative markers in both native and non-native speech. Section 3 will present the methodology of the current work, including the information of the corpus and dataset, the participants, the tokens, the linguistic and extralinguistic factors, and the statistical analysis. In Section 4, the results will be tabulated and discussed. Section 5 will conclude the current work and outline future implications.

2. Literature Review

2.1. Quotative Markers in Native Speech

Quotative markers are less studied in L1 Chinese compared to other linguistic features. Much of the literature has focused on the distinction between direct and indirect speech (See, for example, Kuo 2001, 2007; Ng 2011, etc.). Marginal studies could be found discussing the quotative markers used by Chinese native speakers in speech reporting. The only work investigating different verbal forms used as quotative markers in Chinese was written by Hwang (2000). In his study, by comparing four Chinese classic literature books from four different historical periods in China, he found that only saying verbs (such as "say"), informative verbs (such as "tell"), and cognitive verbs (such as "think" or "believe") were used as quotative markers together with other complementizers to report speech. This use first occurred with saying verbs in ancient times. It was not until modern times that this use spread to cognitive verbs.

However, no literature has historically documented the use of other categories of verbs as quotative markers in Chinese. Compared to other languages, the number of categories in this use in Chinese seems to be extremely limited. For example, the most frequently discussed quotative marker, e.g., *be like* in English, would be examples of using static verbs as quotative markers. As illustrated earlier by Example (5), action verbs, such as "go", could also be used as quotative markers. These usages were never documented or discussed in Chinese.

Meanwhile, most works on quotative markers in Chinese were performed qualitatively. Little is known about how both social and linguistic factors could influence the use of quotative markers in Chinese native speech. However, this should be particularly interesting to variationists working in this domain.

In the case of English, most studies focused on a limited subset of quotative markers, mainly the innovative forms, such as *be like*. An extensive body of literature could be found on the study of the quotative use of the innovative form *be like* in American English (see, for example, Blyth et al. 1990; Buchstaller 2001; Cukor-Avila 2002; Dailey-O'Cain 2000; Tagliamonte 2016; Sedlaczek 2020, etc.). Studies on *be like* could also be found in other varieties of English across the world, including Australian English (Winter 2002), British English (Tagliamonte and Hudson 1998, 1999), Canadian English (D'Arcy 2004; Tagliamonte and D'Arcy 2004; Tagliamonte and Hudson 1998, 1999), and Scottish English (Macaulay 2001). Meanwhile, even though comparatively less, discussion on other quotatives can also be found in the literature. For example, we see some studies on the use of the action verb *go* as a quotative marker in American English (Butters 1980; Schourup 1982; Romaine and Lange 1991), studies on the quotative use of *be all* in California (Waksler 2001) and New York (Singler 2001).

Other empirical studies demonstrated that using quotative markers in native English speech is conditioned by both linguistic and social factors. On the one hand, various research has demonstrated this impact by focusing on different linguistic factors. For example, Blyth et al. (1990) reported that *say* and *go* are more frequently used in the third person singular, while *be like* occurs rarely in the third person singular but commonly in the first person singular. Romaine and Lange (1991) found that *say* is more used to report the speech of others, while *be like* is more often used to report one's inner speech. However, Ferrara and Bell (1995) believed that the use of *be like* has expanded to the third person and to introduce others' speech. Barbieri (2005a) found that *be like*, *go*, and *be all* are more used in the present tense, while *say* is more frequent in the past tense.

On the other hand, many studies have proven the influence of social factors on the variable use of quotative markers. For example, Blyth et al. (1990) argued that *be like* was primarily associated with teenagers and college students. However, Singler (2001) showed that *be like* is mainly favored by speakers in their late twenties to early thirties. As for the gender effect, Dailey-O'Cain (2000) and Ferrara and Bell (1995) both reported that the gender effect has neutralized. Barbieri (2005b) further demonstrated that *be like* and *go* are more likely to be used by young females aged sixteen to twenty-six and males aged twenty-seven to forty. However, Sedlaczek (2020) argued that the youth effect on the rise of *be like* might be a stereotype. She believed that the use of *be like* has become more settled in English in our time.

While the linguistic and social factors are indisputably crucial to the variable use of quotative markers in native speech, it seems reasonable to question whether L2 speakers of English's use of quotative markers are also variable and conditioned by the same set of linguistic and/or social factors.

2.2. Quotative Markers in Non-Native Speech

This discussion has also expanded to research in SLA. Davydova and Buchstaller (2015) explored the acquisition of the constraints that govern the variable use of the innovative quotative form *be like* in the speech of German L1 speakers learning English. It was found that for German L1 students of English, mimetic quotes[2], historical present, exposure abroad, and gender are statistically significant for the use of quotative *be like* in non-native speech. This study supplied further evidence on the impact of social factors on the use of quotative markers in non-native speech.

By adopting a variationist approach, Davydova (2019) focused on the use of quotative marker *like* by both native and non-native speakers of English. In her study, she looked at three factors that might influence this use: grammatical subject, temporal reference, and mimesis. As a result, she found that in L1 English, the neuter pronoun *it*, historical present tense, and mimetic reconstruction favor the use of quotative *like*; in L2 English, only the neuter pronoun *it* and the mimetic reconstruction are significant for its use. Her study shed new light on the acquisition of socio-pragmatical variables in non-native speech. While some linguistic factors are significant to the use of a socio-pragmatical variable in native

speech, they do not necessarily have the same effect on non-native speech. Other factors that are more relevant to non-native speech might influence this use.

Diskin (2017) studied the use of *like* as quotative markers in the speech of 42 adult L2 speakers of English, including 17 Chinese (6 male, 11 female), 19 Polish (9 male, 10 female), and 6 Irish (3 male, 3 female). In her study, she examined linguistic and social factors that might influence this use: position and function of *like*, gender and proficiency of the speakers, and length of residence in the target country. Her results indicated that after three years of residence in Ireland, non-native speakers' rates of use of *like* matched those of native speakers (an effect more significant than proficiency). Male speakers are less likely to use *like*. Clause-initial and clause-medial positions were adopted relatively unproblematically by non-native speakers. Her study confirmed earlier findings in Davydova and Buchstaller's study. The duration of stay in the target country, the gender of the speakers, the mimesis of the quoted content, and the tense of the quotative markers significantly influence this variable use in non-native speech.

Magliacane and Howard (2019) also confirmed that the study abroad experience could benefit L2 learners of English by increasing the frequency of the use of *like* (though not restricted in quotative use) in their oral production, even though a short stay of six months might not allow them to approach native speech frequency. However, it is reasonable to assume that a more extended stay in the target community could yield a different result. Meanwhile, they also proposed that native speech exposure might be crucial to acquire socio-pragmatic competence, such as the use of *like* in non-native speech. It calls for a further examination of the frequency of interaction with native speakers in a natural setting, which should surely be one of the focuses of the current study.

However, despite the fruitful work performed on L2 speech, as seen in these studies, most studies only focused on discussing the influence of various factors on the use of *be like* in non-native speech. Nothing has been said about their influence on the use of other quotative markers in non-native speech. On the one hand, we still do not know what other forms non-native speakers use in speech reporting and the general distribution of these forms in their speech; on the other hand, since action verbs or static verbs were not documented to be used as quotative markers in Chinese but in English, it is worth examining if Chinese L1 speakers learning English acquire this use in their non-native speech at all. Meanwhile, since various factors were proven to be significant to the use of *be like* in non-native speech, it becomes relevant to know if they are also influential to the choice among different variants in non-native speech.

Therefore, in the current study, we address the research gap in the existing literature by looking into corpus data of L1 speakers of Chinese learning English in the States. More precisely, we focus on answering the following research questions: Is the use of quotative markers by Chinese L1 speakers of English variable in their direct speech reporting, or do they only rely on one dominant form? Do Chinese L1 speakers use innovative quotative markers, such as *be like* in English? What linguistic and social factors influence their use of quotative markers? Is there any evidence for L1 impact on their use of quotative markers in non-native speech?

3. Methodology

3.1. Corpus and Data

The dataset we used in this article is extracted from our private learners' corpus built in 2018. The corpus consists of 29 interviews conducted in English with L1 speakers of Chinese learning English in the US.

The interviews are all semi-directed conversations between the interviewer and the interviewees while following the protocols of traditional Labovian sociolinguistic interviews (Labov 1978), which involve informal conversations between the interviewer and the interviewees so that more natural oral data can be solicited. Each conversation varies from 25 min to 48 min. Thus, in total, we have 16 h of recording. A list of questions was prepared prior to the interview for this task. However, these interviews were not

a simple question-and-answer exchange between the interviewer and the interviewee. These questions were used merely as a medium to mimic the natural conversation flow and were loosely based on a script, including (1) questions designed to gather personal information from the interviewees, as well as to familiarize them with the interview format; (2) questions on memorable events in life such as childhood memories, theft-related experience, first days in the US, etc.; (3) a series of questions on their English learning and their opinion on English language use. The fact that the same set of questions was used for every interviewee allowed us to compare different language uses by different speakers under the same context.

3.2. Speakers

The 29 speakers in our corpus are all non-native speakers of English who reside in the US at the time of the interview. Among these 29 speakers are 11 male speakers and 18 female speakers, aged between 22 and 36 years old. They were all born in mainland China and came to the US after puberty. All except for one arrived in the States at the college level. One female speaker arrived in the States at the age of 16. All of them started to take English as one of the compulsory courses in elementary school in China. Before coming to the States, they all passed the TOEFL exam with a score higher than the minimum score required by universities. They all speak mandarin Chinese as L1. All of them were recruited via the snowball technique, which means the interviewees helped the interviewer to ask their friends to participate in the following interviews. Therefore, all interviewees are either the friend of the interviewer or the friend of the interviewees. Their participation was entirely voluntary.

However, it is worth pointing out that these speakers mainly differ from speakers in other studies because they are not exclusively students. The student group consists mainly of international students studying in US universities. Of those who work, as shown in Table 1, 12 have already graduated from school and started their professional careers. Their professions range from interns to employees. However, they all pursued higher education in US universities before joining the workforce. This diversity allows us to gain more insight into the actual use of English by non-native speakers outside the academic environment.

Table 1. The detailed information of the participants.

Speakers (Initials)	Gender	Length of Interview	Total Number of Words Produced	Length of Stay in the US	Age	Profession
HQ	f	28 min 12 s	2503	0.7 year	28	student
JP	f	27 min 22 s	1776	1 year	23	student
SS	f	34 min 45 s	3525	1.5 years	27	student
WDF	m	30 min 56 s	1917	3 years	26	student
FYL	f	35 min 35 s	3091	4 years	31	student
LJX	f	28 min 08 s	1874	4 years	27	student
WSY	f	31 min 38 s	2255	4 years	26	research analyst
YW	m	34 min 45 s	3085	4 years	29	student
ZZ	f	34 min 45 s	3753	4 years	22	student
MCL	m	33 min 10 s	2274	5 years	27	student
XJC	f	25 min 30 s	1897	5 years	32	student
QY	f	35 min 44 s	2983	6 years	32	online consultant
ZLL	f	31 min 22 s	2735	6 years	29	Chinese teacher
DYJ	m	35 min 11 s	3282	7 years	28	intern
LLY	f	45 min 49 s	5209	7 years	23	student
WM	f	36 min 33 s	3407	7 years	32	student
YB	f	36 min 43 s	3076	7 years	29	student
HWL	m	47 min 49 s	4411	8 years	31	senior engineer
SR	f	24 min 13 s	1497	8 years	30	bank employee
WHQ	m	25 min 27 s	2293	8 years	26	student
YJ	f	31 min 12 s	2245	8 years	30	data analyst
ZG	m	43 min 11 s	3940	8 years	31	risk analyst
ZJ	m	47 min 49 s	3383	8 years	33	data analyst
ZX	f	26 min 35 s	1788	8 years	31	student

Table 1. *Cont.*

Speakers (Initials)	Gender	Length of Interview	Total Number of Words Produced	Length of Stay in the US	Age	Profession
MF	m	26 min 26 s	1335	9 years	34	software designer
SL	m	26 min 19 s	1477	9 years	31	student
WD	m	31 min 55 s	2940	10 years	33	software engineer
ZL	f	33 min 58 s	3030	10 years	36	project manager
ZJS	f	39 min 53 s	3685	13 years	32	student

3.3. Tokens

The principal author manually identified all occurrences of quotatives in the corpus by using both the transcription and the aligned audio. Only tokens introducing direct "reported speech, sounds, gesture and thought by self or other", as defined by Pichler (2010, p. 588), were included in our analysis. Acoustic cues, such as changes in pitch and longer pauses before the quoted material, as well as the temporal markers in the quoted material and the grammatical pronoun of the quoted material, were used to ensure more accurate identification of tokens. Occurrences involving indirect speech reporting, whether with or without the presence of "that", were excluded from our analysis. Since our speakers are non-native, the temporal marker or grammatical pronoun might not be 100% accurate as clues. In the case of doubt, the tokens were excluded from our final analysis.

Therefore, in total, we included 183 occurrences of quotatives in our corpus. The following forms were identified: *say*, *think*, *ask*, *be*, *do*, *tell*, *go*, *be like*, *feel*, *introduce*, *invite*, *worry about*, *complain*, *reply*, *describe*, *shout*, *share*, *see*, *call*, *cite*, *yield*, *find*, *add*, and null marker (absence of quotative marker). However, due to the small number of occurrences, we coded *feel*, *introduce*, *invite*, *worry about*, *complain*, *reply*, *describe*, *shout*, *share*, *see*, *call*, *cite*, *yield*, *find*, and *add* together as "other forms" in the following sections for analysis.

3.4. Linguistic Factors

As shown in Table 2, for the linguistic factors, we examined the tense of the quotative verbs, the grammatical subject of the quotative verb, the mimetic reconstruction status of the quoted clause, the content of the quotation, and the presence of discourse markers.

Table 2. Linguistic factors to be examined.

Predictor	Levels
Tense of the verb	present historical present past future
Grammatical subject pronoun	first person second person third person
Mimetic reconstruction	mimetic non-mimetic
Content of quotation	internal monologue direct speech
Presence of discourse markers	*just* *like* absent

For the tense of the quotative verbs, we have four different levels: present tense, historical present tense, past tense, and future tense. Historical present tense differs from the present tense in that although the speaker is referring to a past event, the speaker chooses to use the present tense as if it is happening at the time of the storytelling. For example:

(8) first two days I ***win*** the two hundred dollars I ***think*** oh I am lucky and first come here yeah but when last day I leaving I just ***want to*** try so I totally ***lose*** and (MCL_2018)

Here, the speaker described one of his gambling experiences in Las Vegas. Even though the event has already finished, he used a series of verbs in the present tense, attempting to reconstruct the scene for the listeners.

For the grammatical subject pronouns, we mainly look at whether the subject is in first, second, or third person without distinguishing between singular and plural forms. By examining this factor, we expect to see if certain quotative verbs are more associated with a specific subject pronoun.

As for the mimetic reconstruction status of the quoted content, we can distinguish between mimesis and non-mimesis. Hence, the mimesis can be told mainly by a sudden change of prosodic cues. It is usually a sudden rise in pitch at the beginning of the embedded clause (the quoted content). For the coding, we mainly rely on the acoustic perception of the principal investigator on whether there is a change in pitch in combination with/without a longer pause before the embedded clause.

For the content of the quotation, we have two levels: either it is said by others, or it is the internal monologue by the speakers. For example,

(9) so that's why she does that and I think that's a good idea so people wouldn't be ***regretting*** that oh I forgot to say how I feel how I love this person almost the families (LLY_2018)

(10) so I always have an issue in class when I try to say something my professor will encourage me to ***say like*** hey Qian you need to speak loudly (HQ_2018)

As shown by these two examples, (9) would be coded as internal monologue since "oh I forgot to say how I feel" is part of the person's internal monologue, while (10) would be coded as direct speech as "hey Qian you need to speak loudly" is what was said by the professor.

Finally, during coding, we noticed that some quotative markers were used together with discourse markers, such as *just* or *like*. We wanted to see if all quotative markers could have this use or if only a subset of quotative markers could be used in combination with discourse markers. Hence, for the presence of other discourse markers, we will examine whether the quotative marker is accompanied by a discourse marker, such as *just* or *like*. For example,

(11) if you ***say like*** you all have a good day like you all they'll ***say like*** yall yeah like for some vocabulary (>FYL_2018)

(12) like you know most like 27 like you know your parents like every day will ***just tell*** you ok you need to find ok your partner and get married I guess it's not just my hometown it's the I will say (FYL_2018)

3.5. Extralinguistic Factors

Since our speakers are between 22 and 36 years old, the age of arrival might differ. However, since all speakers arrived in the States after puberty and at least at college age, in the current work, the age of the speaker is not a variable under scrutiny. However, as shown in Table 3, for the extralinguistic factors, we will examine the gender of the speaker, the number of years they have spent in the US, and the self-reported extracurricular contact with native speakers.

Davydova and Buchstaller (2015) argued that, as in native speech, gender is a significant factor in the use of quotatives in non-native speech. Male speakers demonstrate a different preference for quotatives than female speakers. Therefore, in this work, we also examined how females might differ from male speakers in their use of quotatives and whether either of them prefers a particular category.

As demonstrated by Davydova and Buchstaller (2015), exposure abroad is a significant factor in the use of quotatives by non-native German speakers of English. Diskin (2017) also proved that after three years of residence in the target country, non-native speakers'

use of quotative *like* matches that of English native speakers. Therefore, we also looked at the length of stay in the target country in the current work. Since our speakers vary a lot in their duration of stay in the USA, it allows us to group our speakers into three groups: 1–3 years, 4–6 years, and more than 7 years.

Table 3. Extralinguistic factors examined.

Predictor	Levels
Gender	female male
Length of stay (years)	1–3 4–6 7–9 10+
Extracurricular contact with native speakers	rare occasional frequent

Even though previous work has never addressed the extracurricular contact with native speakers as a potentially influential factor for the use of quotatives by non-native speakers, this factor was proven to be a significant factor in the use of other discourse-pragmatic variables in non-native speech (see, for example, Sankoff et al. 1997; Lasan and Rehner 2018, etc.). Therefore, based on the self-reported extracurricular contact with native speakers by speakers during the interview, we distinguished among three levels: frequent, occasional, and rare contact with native speakers outside the classroom.

3.6. Statistical Analysis

For this article, we used the regression model carried out in the R environment using Rbrul (Johnson 2009). The model distinguishes the following levels for statistical significance: $p > 0.05$, not significant; $p < 0.05$, significant; $p < 0.01$, very significant; $p < 0.001$, highly significant.

4. Results

4.1. General Distribution

Table 4 shows the distribution of each quotative marker in our corpus. The percentage was calculated by the number of occurrences of each quotative marker over the total number of all forms combined. As shown in Table 4, we identified 183 quotative markers over 16 h of recording in our corpus. We can see that the quotative marker *say* is still the dominant form in direct speech reporting. Leaving other forms aside, the null marker, in which case the speaker does not use any quotative marker at all, is the second preferred form by non-native speakers in our corpus. For example:

(13) and then **Ø** yeah but you sound America so um I was like ok I guess I do go to the US for college I think I pick up a lot of kind of (LLY_2018)

Here, the speaker talked about her conversation with an Uber driver when she traveled to South Africa. To quote what was said by the driver, she did not employ any quotative verb. Instead, she appealed to prosodic cues to change her intonation to let the hearer know that the following content was said by the driver.

Table 4. General distribution of different forms in our corpus.

	Say	*Other*	*Ø*	*Be Like*	*Tell*	*Ask*	*Think*	*Be*	*Talk*	*Do*	*Go*	*Speak*
N	70	32	19	18	15	10	8	6	2	1	1	1
%	38.25	17.49	10.38	9.84	8.20	5.46	4.37	3.28	1.09	0.55	0.55	0.55
Total	183											

Be like is another frequently used form by our speakers. As *be like* is one of the characteristics of youth speech, considering the relatively young age of our speakers, it is not surprising to see a preference for this marker. However, it is worth noting that we distinguish *be like* (example 14) from *be* without the presence of *like* (example 15).

(14) um the driver asked me are you an America like do like do you live in America I ***was like*** why you say that I am Chinese right like you can tell I am Chinese (LLY_2018)
(15) I use a lot ***is*** wait I am confused then my friend like you are always confused and (LLY_2018)

As shown by Table 4, *be* is accompanied by *like* in most cases. Only 6 out of 183 occurrences show *be* used as a quotative marker in direct speech without *like*.

Similar to what was observed in native speech (see, for example, Butters 1980; Blyth et al. 1990; Ferrara and Bell 1995; Singler 2001, etc.), we also have examples of action verbs, such as *go* or *do*, used as quotative markers; for example,

(16) so even though I most of the time ***do*** um how's going good what happened is like how's going good you I'm good too and then you leave (LLY_2018)
(17) a lot of them will ***go*** holly (ZJS_2018)

Nevertheless, this kind of usage is highly marginal in our corpus. We only have one occurrence of each.

Since the occurrences of other forms are of relatively small numbers, for statistical purposes, we compare the dominant form *say* with all other forms so that our dependent variable is binary.

4.2. Cross-Tabulation

Table 5 presents the detailed cross-tabulation information of all factors for the distribution of occurrences between *say* and other forms. This is important because it guarantees that no category contains 0 occurrences, which would later affect the result of the statistical model.

Table 5. Cross-tabulation of all factors.

Linguistic Factors			
	No. of Occurrences of All Other Forms	**No. of Occurrences of *Say***	**Total No. of Occurrences in the Corpus**
Content of quotation			
Direct speech	86	68	154
Internal monologue	27	2	29
Total	113	70	183
Tense of the verb			
Future	3	9	12
Historical present	8	7	15
Present	64	26	90
Past	38	28	66
Total	113	70	183
Presence of Other Discourse Markers			
Just	5	4	9
Like	41	9	50
Absent	67	57	124
Total	113	70	183
Mimetic reconstruction			
Mimetic	107	62	169
Non-mimetic	6	8	14
Total	113	70	183
Grammatical subject pronoun			
First	40	18	58
Second	13	2	15
Third	60	50	110
Total	113	70	183

Table 5. *Cont.*

Extralinguistic Factors			
	Other	***Say***	**Total**
Gender			
Female	87	55	142
Male	26	15	41
Total	113	70	183
Length of Stay			
1–3	23	19	42
4–6	36	19	55
7–9	40	22	62
10+	14	10	24
Total	113	70	183
Extracurricular contact with native speakers			
Frequent	62	37	99
Occasional	41	23	64
Rare	10	10	20
Total	113	70	183

The cross-tabulation shows that the majority of internal monologue uses quotative verbs other than the dominant form *say*. As for the presence of the discourse marker, *like* tends to appear together with verbs other than *say* in direct speech reporting. The first-person and second-person pronouns seem to prefer using other forms, while the third person does not show any particular preference. However, all these need to be confirmed by the statistical model.

4.3. Regression Analysis

Table 6 presents the detailed regression analysis results produced by Rbrul for all factors. The factor of participants is treated as a random factor. Our results indicated that only three factors were tested to be statistically significant for the use of quotative marker *say* against all other variants: the content of the quotation ($p = 3.69 \times 10^{-5}$), the presence of other discourse markers ($p = 0.00938$), and the tense of the verb ($p = 0.0491$). Based on the *p*-value, the content of the quotation is highly significant to the choice between the dominant form *say* and its variants, the presence of other discourse markers is very significant, and the tense of the verb is significant. All other factors are not statistically significant for the choice of quotative markers.

Table 6. Statistical results for all factors.

	Quotative Markers in Direct Speech Reporting: Say/Other Forms		
Input prob.		0.234	
Total number		183	
Log. likelihood		−103.14	
	F. w	%	N
Content of quotation		$\mathbf{p = 3.69 \times 10^{-5}}$	
Direct speech	0.767	44.2	154
Internal monologue	0.233	6.9	29
Presence of Other Discourse Markers		$\mathbf{p = 0.00938}$	
Just	0.595	44.4	9
Absent	0.585	46	124
Like	0.326	18	50
Tense of the verb		$\mathbf{p = 0.0491}$	
Future	0.782	75	12
Historical present	0.471	46.7	15
Past	0.407	42.4	66
Present	0.314	28.9	90

Table 6. *Cont.*

	Quotative Markers in Direct Speech Reporting: Say/Other Forms		
Extracurricular contact with native speakers		Not significant [$p = 0.695$]	
Rare	[0.58]	50	20
Frequent	[0.513]	37.4	99
Occasional	[0.407]	35.9	64
Grammatical subject pronoun		Not significant [$p = 0.794$]	
Second	[0.568]	13.3	15
Third	[0.496]	45.5	110
First	[0.436]	31	58
Gender		Not significant [$p = 0.582$]	
Female	[0.542]	38.7	142
Male	[0.458]	36.6	41
Length of Stay		Not significant [$p = 0.815$]	
4–6	[0.553]	34.5	55
1–3	[0.532]	45.2	42
10+	[0.511]	41.7	24
7–9	[0.405]	35.5	62
Mimetic reconstruction		Not significant [$p = 0.361$]	
Non-mimetic	[0.583]	57.1	14
Mimetic	[0.417]	36.7	169
Participants		Random [not tested]	

Our results revealed that direct speech largely favors the dominant form *say*, while internal monologue disfavors it. As for the presence of other discourse markers, *like* disfavors the dominant form *say* while it favors other variants. Our earlier observation by cross-tabulation on these two factors is confirmed. However, the grammatical subject pronoun is not as significant as we assumed by cross-tabulation. In addition, the statistical results further confirmed that the future tense is the most associated with the dominant form *say*, while the present tense is the least associated with the dominant form *say*.

5. Discussion

To answer our research questions: first, is the use of quotative markers by Chinese L1 speakers of English variable in their direct speech reporting, or do they only rely on one form? As demonstrated by our results, the use of quotative markers in the speech of Chinese L1 speakers learning English is variable in the sense that we do see the use of verbal forms other than *say* as quotative markers. Even though the traditional quotative marker *say* is still the dominant form, it only comprises 38.25% of all forms used. We still notice an important use of the null marker, *be like,* and other forms.

Second, do Chinese L1 speakers use innovative quotative markers, such as *be like* in English? The answer is undoubtedly yes. In our corpus, 9.84% of quotative markers were found to be *be like*. As shown in the previous literature, *be like* is the form used mostly by young speakers in their late twenties to early thirties, which matches the overall age range of the non-native speakers in our corpus (aged 22 to 36). As almost all speakers (except for one) in our corpus started their journey in the United States at the college level, where the use of *be like* in native speech peaks, it is reasonable to suggest that our non-native speakers might acquire the use of this quotative marker during their college time through their interaction with native peers in school and continue to maintain the use of this form even after graduation. This also seems to provide evidence for the importance of age of arrival and the native interactional pool. This should at least be valid for the use of certain forms. It is more likely for non-native speakers to acquire forms used mostly by native speakers with whom they interact the most in daily communication. This finding also partially supports what was suggested by Magliacane and Howard (2019) in that the type

of interaction with native speakers should be crucial for non-native speakers to approach the native frequency, even though they may not reach the same level as native speakers.

Third, what linguistic and social factors influence the speakers' use of quotative markers? Previous studies on the relation between the acquisition of socio-pragmatic competence and length of stay in study abroad demonstrated divergent results regarding the effect of the length of stay in the target country on this acquisition. For example, Taguchi (2013) showed that the study abroad experience positively affects the production of pragmatic routines. Other researchers, however, demonstrated the contrary. For example, Bardovi-Harlig and Bastos (2011) proved that the length of stay alone in the target country is not significant to the acquisition of socio-pragmatic competence for L2 speakers, while it is indeed the intensity of interaction with native speakers that is more crucial to this acquisition (see, also, Bardovi-Harlig 2013; Taguchi and Roever 2017).

In contrast, our results showed that none of the social factors tested in the current study seems to be significant to the variable use of quotative markers in non-native speech. This suggests that the acquisition of socio-pragmatic competence might be more complex than we thought. While some aspects might be related to the length of stay in the target country or the intensity of interaction with native speakers, some might not be related to a unique factor. The current data might not allow us to examine the phenomenon further. A longitudinal study on the same group of speakers in the future might shed more light on this acquisition process.

However, linguistic factors do seem to prevail in this variable choice. What stands out as a factor is the tense. While Barbieri (2005a) found that *be like*, *go* and *be all* are more used in the present tense, while *say* is more frequent in the past tense, this does not seem to be the pattern observed here in non-native speech. Our results indicated that in non-native speech, *say* is dominantly used in future tense and disfavors all other tenses. As the influence of tense was not discussed or considered in previous research on this topic in L2 speech, this finding does open a new direction for further queries of the observed pattern. For example, why in non-native speech, *say* is associated with the future tense, while native speakers use it more with the past? What makes future tense different from other tenses in non-native speech? One possible assumption might be related to the difference in temporal perception. It might be the case that native speakers of English have a different sense of temporality. Due to our limited data in the current study, it is impossible to validate this hypothesis. It is worth further examination in subsequent studies of the same topic on non-native speech.

Fourth, is there any evidence for L1 impact on the use of quotative markers in non-native speech? The answer is: perhaps. As seen in the literature, action verbs, such as *go*, have long been documented in English native speech. However, we only had one occurrence of *go* used as a quotative marker in our corpus. Almost all speakers, except for one, do not use action verbs at all as quotative markers. As shown in the literature, in L1 Chinese, action verbs are never used as quotative markers. It is reasonable to believe that the lack of verbs in this category in our data might be due to the impact of L1. However, due to the lack of data in L1 Chinese for comparison in the current work, it is impossible to confirm our hypothesis.

Meanwhile, our data also opened another direction for discussion: the importance of age of arrival for the acquisition of certain forms in L2 speech, for example the form "go". The only speaker who used this form was the one who came to the States at a younger age (around age 16) to attend high school. Notice that this is right around the "critical period," as proposed by various researchers in language learning (see, for example, Sankoff and Blondeau 2007). According to the theory, this is the period where an individual calibrates his/her speech according to peers. If speakers arriving at the target country after the critical period do not use action verbs as quotative markers, could this mean that the acquisition of action verbs as quotative markers, at least for Chinese L1 speakers learning English, is related to the age of arrival? However, since we only have one case of *go* as a quotative marker in our corpus, it is impossible to draw any conclusion more conclusive. It is undoubtedly preferable to look into this direction for future studies.

6. Conclusions

As Diskin (2013, p. 69) put it, "migration is a unique set of life experiences that inevitably influence a person's outlook, attitudes and behavior. It is, therefore, a vastly dynamic context within which to examine language change at the intersection of social change." In this article, by investigating data collected from 29 Chinese L1 speakers of English in the US, we examined the quotative markers used by non-native speakers in their direct speech reporting. Our results showed that the content of the quotation, the presence of other discourse markers, and the tense of the quotative verb significantly influenced the choice of quotative markers in non-native speech.

At the same time, we showed that non-native speakers do use forms other than *say* as quotative markers in their speech. Meanwhile, non-native speakers also use other prosodic cues to contextualize their storytelling as native speakers do. A lack of extralinguistic factor influence might be explained by the strong influence of L1. We also proposed to examine other aspects of extralinguistic factors, such as the degree of integration. However, such factors should be carefully measured since the self-reported data may not be accurate. For future studies, it would be advisable to conduct a cross-language examination of quotative markers used by speakers with the same L1 background but different L2 to confirm any potential impact of L1. Another possible track would be to pursue longitudinal studies on the same set of speakers over the lifespan to see how the changes in social life could affect their use of quotative markers.

Funding: This research received no external funding.

Institutional Review Board Statement: The study did not require ethical approval. Ethical review and approval were waived for this study as it involved no experimenting with humans. The presented analyses were carried out on an already existing corpus of recordings at the University of Florida. The practice included informed consent from the recorded participants and reassurance by the institution that the recordings were anonymous and strictly confidential, meant to be used only for research purposes by the institution.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 Examples (1)–(7) are taken from our corpus. The acronyms in the brackets are the name initials of the speakers.

2 Non-mimesis is broadly defined as quotes performed by the speakers with a normal "voice", whereas mimesis involves quotes "set off from the surrounding speech in terms of suprasegmental features (such as prosody, pitch, accent, and so on) (Davydova and Buchstaller 2015, p. 446).

References

Barbieri, Federica. 2005a. Quotative use in American English: A corpus-based, cross-register comparison. *Journal of English Linguistics* 33: 222–56. [CrossRef]

Barbieri, Federica. 2005b. Revisiting the Effect of Sex and Age on Quotative Use in American English: A Corpus-Based Study. Paper presented at the AAACL/ICAME Joint Conference, University of Michigan, Ann Arbor, MI, USA, May 11–15.

Bardovi-Harlig, Kathleen. 2013. Developing L2 pragmatics. *Language Learning* 63: 68–86. [CrossRef]

Bardovi-Harlig, Kathleen, and Maria-Thereza Bastos. 2011. Proficiency, length of stay, and intensity of interaction, and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics* 8: 347–84. [CrossRef]

Blyth, Carl, Sigrid Recktenwald, and Jenny Wang. 1990. I'm Like, "Say What?!": A New Quotative in American Oral Narrative. *American Speech* 65: 215–27. [CrossRef]

Buchstaller, Isabelle. 2001. An Alternative View of Like: Its Grammaticalisation in Conversational American English and Beyond. *Edinburgh Working Papers in Applied Linguistics* 11: 21–41.

Butters, Ronald. 1980. Narrative Go 'Say'. *American Speech* 55: 304–7. [CrossRef]

Cukor-Avila, Patricia. 2002. *She Say, She Go, She Be Like*: Verbs of Quotation over Time in African American Vernacular English. *American Speech* 77: 3–31. [CrossRef]

D'Arcy, Alex. 2004. Contextualizing St. John's Youth English within the Canadian Quotative System. *Journal of English Linguistics* 32: 323–45. [CrossRef]
Dailey-O'Cain, Jennifer. 2000. The Sociolinguistic Distribution of and Attitudes toward Focuser Like and Quotative Like. *Journal of Sociolinguistics* 4: 60–80. [CrossRef]
Davydova, Julia. 2019. Quotative like in the Englishes of the Outer and Expanding Circles. *World Englishes* 38: 578–92. [CrossRef]
Davydova, Julia, and Isabelle Buchstaller. 2015. Expanding the circle to Learner English: Investigating quotative marking in a German student community. *American Speech* 90: 441–78. [CrossRef]
Diskin, Chloe. 2013. Integration and identity: Acquisition of Irish-English by Polish and Chinese migrants in Dublin, Ireland. *Newcastle Working Papers in Linguistics* 19: 67–89.
Diskin, Chloé. 2017. The use of the discourse-pragmatic marker 'like' by native and non-native speakers of English in Ireland. *Journal of Pragmatics* 120: 144–57. [CrossRef]
Ferrara, Kathleen, and Barbara Bell. 1995. Sociolinguistic Variation and Discourse Function of Constructed Dialogue Introducers: The Case of Be + Like. *American Speech* 70: 265–90. [CrossRef]
Hwang, Jya-Lin. 2000. Historical development of reported speech in Chinese. Paper presented at the 26th Annual Meeting of the Berkeley Linguistics Society, Berkeley, CA, USA, February 18–21; vol. 26, pp. 145–56.
Johnson, Daniel Ezra. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed—Effects variable rule analysis. *Language and Linguistics Compass* 3: 359–83. [CrossRef]
Kuo, Sai-Hua. 2001. Reported speech in Chinese political discourse. *Discourse Studies* 3: 181–202. [CrossRef]
Kuo, Sai-Hua. 2007. Language as ideology: Analyzing quotations in Taiwanese news discourse. *Journal of Asian Pacific Communication* 17: 281–301. [CrossRef]
Labov, Beatriz. 1978. Where does the sociolinguistic variable stop? A response to Beatriz Lavandera. *Working Papers in Sociolinguistics* 44: 1–17.
Labov, William, Sharon Ash, and Charles Boberg. 2008. *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: Walter de Gruyter.
Lasan, Ivan, and Katherine Rehner. 2018. Expressing and perceiving identity and intentions in a second language: A preliminary exploratory study of the effect of (extra) curricular contact on sociolinguistic development. *International Journal of Bilingual Education and Bilingualism* 21: 632–46. [CrossRef]
Levey, Stephen, Karine Groulx, and Joseph Roy. 2013. A variationist perspective on discourse-pragmatic change in a contact setting. *Language Variation and Change* 25: 225–51. [CrossRef]
Macaulay, Ronald. 2001. *You're Like 'Why Not?'* The Quotative Expressions of Glasgow Adolescents. *Journal of Sociolinguistics* 5: 3–21. [CrossRef]
Magliacane, Annarita, and Martin Howard. 2019. The role of learner status in the acquisition of pragmatic markers during study abroad: The use of 'like'in L2 English. *Journal of Pragmatics* 146: 72–86. [CrossRef]
Ng, Eva. 2011. A survey of court interpreters' use of direct versus reported speech in the Hong Kong courts. *Law, Language & Discourse* 1: 29–58.
Pichler, Heike. 2010. Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics* 14: 581–608. [CrossRef]
Romaine, Suzanne, and Deborah Lange. 1991. The use of like as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech* 66: 227–79. [CrossRef]
Sankoff, Gillian, and Hélène Blondeau. 2007. Language change across the lifespan: /r/in Montreal French. *Language* 83: 560–88. [CrossRef]
Sankoff, Gillian, Pierrette Thibault, Naomi Nagy, Hélène Blondeau, Marie-Odile Fonollosa, and Lucie Gagnon. 1997. Variation in the use of discourse markers in a language contact situation. *Language Variation and Change* 9: 191–218. [CrossRef]
Schourup, Lawrence. 1982. Quoting with Go 'Say'. *American Speech* 57: 148–49. [CrossRef]
Sedlaczek, Leona. 2020. *The Rise of the Quotative "Be Like" and Sociolinguistic Stereotypes amongst Young Speakers*. Munich: GRIN Verlag.
Singler, Victor. 2001. Why You Can't Do a VARBRUL Study of Quotatives and What Such a Study Can Show Us. *University of Pennsylvania Working Papers in Linguistics* 7: 257–78.
Tagliamonte, Sali. 2016. *Teen Talk: The language of Adolescents*. Cambridge: Cambridge University Press.
Tagliamonte, Sali, and Alex D'Arcy. 2004. He's like, she's like: The quotative system in Canadian youth. *Journal of Sociolinguistics* 8: 493–514. [CrossRef]
Tagliamonte, Sali, and Rachel Hudson. 1998. *Be Like* et al. Beyond America: The Quotative System in Canadian and UK Youth. *Cahiers Linguistiques d'Ottawa* 26: 55–69. [CrossRef]
Tagliamonte, Sali, and Rachel Hudson. 1999. *Be Like* et al. Beyond America: The Quotative System in British and Canadian Youth. *Journal of Sociolinguistics* 3: 147–72. [CrossRef]
Taguchi, Naoko. 2013. Production of routines in L2 English: Effect of proficiency and study abroad experience. *System* 41: 109–21. [CrossRef]
Taguchi, Naoko, and Carsten Roever. 2017. *Second Language Pragmatics*. Oxford: Oxford University Press.

Waksler, Rachelle. 2001. A new all in conversation. *American Speech* 76: 128–38. [CrossRef]
Winter, Joanne. 2002. Discourse Quotatives in Australian English: Adolescents Performing Voices. *Australian Journal of Linguistics* 22: 5–21. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.