





Is it in Colloquial Singapore English

WILKINSON DANIEL WONG GONZALES ,
MIE HIRAMOTO , JAKOB R. E. LEIMGRUBER  AND
JUN JIE LIM 

What variation can tell us about its conventions and development

1. Introduction

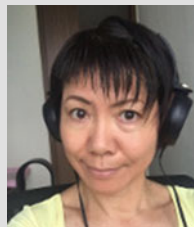
Colloquial Singapore English (CSE, commonly known as Singlish) is a linguistic variety used in Singapore, a Southeast Asian nation home to three major ethnic groups: the Chinese (74.35% of the citizen and permanent resident population), the Malays (13.43%), and the Indians (9%) (Singapore Department of Statistics, 2019). It is one of the best known post-colonial varieties of English and has been documented since the emergence of the field of world Englishes (e.g., Greenbaum, 1988; Richards & Tay, 1977). Linguistically, the grammar and lexicon of CSE are systematically imported from other non-English languages used in the island nation (Leimgruber, 2011). From a creolist perspective, it can be viewed as an English-lexifier creole that contains influences from Sinitic languages such as Hokkien, Cantonese and Mandarin, as well as Malay, Tamil and other varieties in the Singapore language ecology (McWhorter, 2007; Platt, 1975). Several distinct features across various levels of language have been investigated in CSE, including phonetics (Starr & Balasubramaniam, 2019), morphosyntax (Bao, 2010; Bao & Wee, 1999), semantics (Hiramoto & Sato, 2012), and pragmatics (Hiramoto, 2012; Leimgruber, 2016; Lim, 2007).

In this paper, we investigate a feature of CSE that has received relatively little attention in the literature – the question marker *is it*. We first demonstrate how *is it* is used in standardized English (Section 2), before relaying previous insights and our initial observations regarding the use of *is it* in CSE (Section 3). From there, we state our



WILKINSON DANIEL WONG GONZALES is an Assistant Professor at The Chinese University of Hong Kong (Department of English). His research interests include World Englishes, sociolinguistics, language variation and

change, language contact, and language documentation. Email: wdwgonzales@gmail.com



MIE HIRAMOTO is an Associate Professor in the Department of English, Linguistics and Theatre Studies at National University of Singapore. Her research interests are sociolinguistics and linguistic anthropology, in

particular, contact linguistics as well as language, gender, and sexuality. Email: ellmh@nus.edu.sg



JAKOB R. E. LEIMGRUBER is Professor at the University of Regensburg, Germany. His research focusses on world Englishes and on English in multilingual contexts. He is the author of *Singapore English: Structure, Variation, and*

Usage (2013) and *Language Planning and Policy in Quebec: A Comparative Perspective* (2019). Email: jakob.leimgruber@ur.de

doi:10.1017/S0266078422000141

English Today page 1 of 14 (2022). Printed in the United Kingdom © The Author(s), 2022. Published by Cambridge University Press 1

This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

<https://doi.org/10.1017/S0266078422000141> Published online by Cambridge University Press



JUN JIE LIM is a PhD student in linguistics at the University of California San Diego. His research interests are syntax and its interfaces, variation and change, and fieldwork. His current work focuses on Khalkha Mongolian and Singlish.

Email: jjlim@ucsd.edu

goals, hypotheses and methodology (Section 4). This is followed by an analysis of CSE *is it* (Section 5) and a general summative discussion of *is it* and its development (Section 6). We conclude the paper in Section 7.

2. Standardized English *is it*

In standardized English yes-no questions, *is it* – as a single constituent or linguistic unit – seems to only function as a question tag. It takes one of two forms: canonical tags and non-canonical tags. Canonical tags are expected to conform to structural rules (Quirk et al., 1985: 810). First, they should comprise an auxiliary verb and subject, in that order.

Second, the subject/expletive of the tag must be a pronoun that is, by default, identical to the subject/expletive of the statement and that matches it in gender, person, and number.

The verb also must be of the same tense as the verb of the main clause.

Third, if the statement is negative, the tag should be positive, and vice versa.

Finally, the tag should be appended to the right periphery of the statement.

For example, in (6), the tag *is it* is characterized as canonical because not only does it consist of an auxiliary verb and subject in the correct order, the expletive, *it*, in the tag is identical to the expletive pronoun in the preceding statement. Further, the tag appears in the positive – the reverse of the preceding negative statement. And finally, the tag appears at the right end of the sentence.

Non-canonical tags, on the other hand, do not always need to have the auxiliary verb and subject/expletive (e.g., *is . . . right?*). They also might not follow the polarity rule (e.g., *is . . . is it?*). The *is it* tag in (7), for example, is classified as non-canonical because the statement and tag do not agree in polarity, even if all other conditions of a canonical tag are fulfilled.

-
- | | |
|--|---|
| (1) <i>She is hungry, isn't she?</i> | (auxiliary verb + subject) |
| * <i>She is hungry, she isn't?</i> | (subject + auxiliary verb) |
| (2) <i>She is hungry, isn't she?</i> | (third person singular feminine) |
| * <i>She is hungry, isn't he?</i> | (gender mismatch; feminine/masculine) |
| * <i>She is hungry, isn't you?</i> | (person mismatch; third person/second person) |
| * <i>She is hungry, aren't they?</i> | (number mismatch; singular/plural) |
| (3) <i>She is hungry, isn't she?</i> | (non-past/non-past) |
| * <i>She is hungry, wasn't she?</i> | (tense mismatch; non-past/past) |
| (4) <i>She is hungry, isn't she?</i> | (positive/negative) |
| <i>She isn't hungry, is she?</i> | (negative/positive) |
| * <i>She is hungry, is she?</i> | (identical polarity: positive/positive) |
| * <i>She isn't hungry, isn't she?</i> | (identical polarity: negative/negative) |
| (5) <i>She is hungry, isn't she?</i> | (right periphery) |
| * <i>Isn't she</i> she is hungry? | (left periphery) |
| (6) <i>It's not (all) that unusual for women nowadays to join the police force, is it?</i> | |
| (Quirk et al., 1985: 447, canonical construction) | |
| (7) <i>Your car is outside, is it?</i> | |
| (Quirk et al 1985: 812, non-canonical construction) | |
-

Despite their differences, both canonical and non-canonical tags in English yes-no questions always appear clause-finally (Quirk et al., 1985: 810).

Functionally, both tags are also used to express a verification function; however, non-canonical tags have additional functions that distinguish

them from canonical tags (Columbus, 2010). For instance, the use of the non-canonical tag *yeah* in British English is not only used for verification, but also to acknowledge the listener without meaning to elicit a response, among other functions (Columbus, 2010).

A summary of the similarities and differences between the canonical tag and the non-canonical tag can be found in Table 1.

3. What we know so far about CSE *is it*

Is it in CSE has been characterized by scholars as an invariant tag question, appended at the end of the question (Leimgruber, 2011; Wong, 2014). Functionally, it differs from standardized English in that CSE *is it* does not imply that the speaker expects agreement, i.e., CSE speakers rarely use it to invite the listener to express their opinion (Wong, 2014). Apart from functional differences, CSE *is it* is also structurally different from standar-

dized English *is it*, based on data from the Singapore component of the International Corpus of English (ICE-SIN), a local corpus that was made available in 2002 as part of a global initiative with the goal of collecting material for comparative studies of English worldwide (Nelson, 2002). First, unlike standardized English where canonical tags (henceforth, polarity tags) are more commonly used (Quirk et al., 1985: 812), such usage is rare in CSE: only one out of 59 (1.69%) instances of the polarity tag *is it* found in ICE-SIN can be described as canonical (8).

Instead, non-canonical tags (henceforth, non-polarity tags) are widespread in CSE, taking up 91.5%, or 54 out of 59 instances of *is it* found in ICE-SIN. An example is shown in (9).

Third, based on observations from the ICE-SIN data and the authors' observations of CSE, there seems to be a novel use of *is it* – that of a clause-initial yes/no question marker (henceforth, clause-initial *is it*).

(8) *it's not their problem is it*
<ICE-SIN:S1A-062#89:1:B>

(9) *It is too sweet is it?*
'It is too sweet, isn't it?'
<ICE-SIN:S1A-052#130:1:C>

(10) *Is it everyone likes Jane Austin?*
'Does everyone like Jane Austin?'
<ICE-SIN:S1A-078#172:1:B>

(11) *Is it you think I'm pretty?*
'Do you think I am pretty?'
(spontaneous conversation, 21-year-old Chinese Singaporean, NUS Campus, 2017)

The *is it* in both (10) and (11) does not function as a question tag because it is not syntactically located at the end of the clause. The *is it* observed also does not seem to be the result of a subject-auxiliary inversion operation, where the auxiliary syntactically moves from the preverbal position to the clause-initial position before the subject. This is because the *it* in *is it* is not the subject of the sentence e.g., in (11) – *you* is. *Is it* here instead functions as a single constituent that can be treated as an analog to the interrogative *do*-support construction (e.g., *Do you think...* vs. *Is it you think...*). This is a potential innovation in CSE that appears to be uncommon in the corpus (6.77%; four out of 59) but ubiquitous based on our anecdotal observations.

This preliminary investigation has uncovered two key observations. First, the distribution of polarity vs non-polarity *is it* question tags in CSE seems to be very different from standardized English – while the canonical tag is more frequently used in standardized English (Quirk et al., 1985), the situation is reversed in CSE where the non-polarity tag is more common. Second, there appears to be another as-of-yet unnoticed use of *is it* in CSE, where *is it* is found in clause-initial position. In the following sections, we empirically test these two observations using data from a newly compiled corpus, to see if they hold more generally of the CSE-using population. We also seek to identify the linguistic and social conditions that may play a role in CSE speakers' use of *is it*.

Table 1: Comparison between canonical vs. non-canonical tags (✓ = yes; ? = not necessarily)

		Canonical tag e.g., <i>isn't ... is it?</i>	Non-canonical tag e.g., <i>is ... {is it, right, etc.}?</i>
Structural	Composition (Aux + Subj)	✓	?
	Order (Aux then Subj)	✓	?
	Co-referentiality	✓	?
	Polarity	✓	?
	Clause-finality	✓	✓
Functional	Verification function	✓	✓
	Other non-verification functions (e.g., narrative, emphatic, see Columbus 2010)	?	✓

4. Goals, hypotheses, methodology

In this paper, building on our anecdotal observations and findings in our preliminary corpus analyses, we first aim to establish the distribution of question-marking constituent *is it* in contemporary CSE. To fulfill this goal, we draw on the first version of the 3.6 million-word Corpus of Singapore English Messaging (CoSEM), compiled at the National University of Singapore (Gonzales et al., 2021). CoSEM is a sociolinguistic corpus of online text messaging data collected between 2016 and 2021. Each utterance in CoSEM is tagged with the following social metadata: sex (male and female), age (18 to 69), race¹ (e.g., Chinese, Malay, Indian), and nationality (e.g., Singaporean, Malaysian). CoSEM comprises finger speech – or speech that is ‘less like the written language’ (Faulkner & Culwin, 2005: 182; McWhorter, 2013), providing a novel approach to studying how CSE speakers use *is it* in a conversational context. Second, we aim to identify the linguistic, social, and orthographical conditions under which these question-marking variants are used; specifically, we adopt a statistical approach toward analyzing variation. From our results, we hope to shed some light on the development of *is it* in CSE.

Hinging on our initial observations, we hypothesize that the non-polarity *is it* and clause-initial *is it* constructions will be more frequently used compared to the polarity *is it*. We also hypothesize that pragmatic factors such as affect and rhetoricity will influence a speaker’s choice of the *is it* variant, given studies reporting associations between pragmatic function and linguistic variables (Pratt, 2021; Siemund, 2018; Wong, 2014). Specifically, based on our preliminary work and consultation with

native CSE speakers in 2018, we hypothesize speakers will be more likely to use clause-initial *is it* if they are indexing ‘playful’ and ‘mocking’ affects and if they do not expect an answer from the addressee (i.e., in a ‘rhetorical’ context). We expect speakers to display an increased use of non-polarity tag *is it* in contexts that are not ‘playful’, ‘mocking’, and/or rhetorical.

We also hypothesize that certain variants will be used by certain speaker groups more than others. In general, we hypothesize that age, nationality, gender, and race will condition the use of CSE *is it*, as these factors have been found to correlate with other features of CSE (Gonzales et al., 2021; Leimgruber et al., 2020). We did not find correlations between orthographical choice (e.g., *is it*, *isit*, *izzit*, *issit*) and the use of polarity and innovative *is it* in the preliminary data, so we hypothesize that orthography will not systematically condition *is it* use. A significant part of the variation in the use of orthographical variants will be attributed to factors such as individual preferences or habits regarding spelling conventions.

To prepare the data for analysis, we collected all tokens of *is it* in CoSEM before filtering out irrelevant instances such as singleton uses (e.g., *Is it?*) as well as instances involving auxiliary inversion (e.g., *Is it dead?*). After the initial selection process, there were a total of 1902 instances of CSE *is it* in the database. We then calculated the raw and relative frequency distributions of the different types of *is it* constructions (polarity, non-polarity, clause-initial). We fit a series of generalized linear regression models with logistic link function on our dataset in the R environment (Bates et al., 2015; Kuznetsova, Brockhoff & Christensen, 2019; R Core Team,

To test our hypotheses on the conditioning effect of social factors on *is it* production, we analyzed the data in relation to the embedded social information of each utterance. Particularly, we analyze four variables – age, race, nationality and sex – by including them as potential predictors in our *is it* models. Age was included as a continuous variable (range = 18 to 69). Race, nationality and sex were included as binary categorical variables (i.e., Chinese vs. non-Chinese, Singaporean vs. non-Singaporean, male vs. female; see Gonzales et al., 2021 for more information regarding the

For rhetoricity, the native-speaker coders examined the utterances after the question containing *is it*. If a yes or no response immediately succeeds this statement, the question is coded as non-rhetorical (12a). Otherwise, it is coded as rhetorical (12b).

- (12) a. A: *LOL the two virgins is it*
 ‘Haha. Do you mean the two virgins?’
 B: *Yup ??*
 ‘Yes?’
 <COSEM:17CM02-731-23CHM-2016> (non-rhetorical)
- b. A: *wa is it u saying i wuliao*
 ‘Wow, are you saying that I am boring?’
 B: *LOL*
 <COSEM:18CF59-2829-20CHF-2017> (rhetorical)

	S	Z
unconcatenated	<i>We eat in the west is it</i> ‘Do we eat in the west?’ <COSEM:17CF10–6652–20CHF–2016>	<i>Iz it LY/F/CH/19 is our OG Acads Director</i> ‘Is LY/F/CH/19 our OG Academic Director?’ <COSEM:17CM02–21841–21CHM–2016>
concatenated	<i>Issit u fly aeroplane</i> ‘Would you like to fly here via airplane?’ <COSEM:18CF52–11845–21CHM–2016>	<i>Eh ichtat is ichtat.edu.com.sg izit?</i> ‘Is iChat ichtat.edu.com.sg?’ <COSEM:18CF54–6912–19CIM–2015>

For affect, the coders examined the utterances surrounding the *is it* question to determine whether it has a ‘mocking’ or ‘playful’ affect. Although both affective meanings overlap, we distinguish between these two as follows. A question is coded as ‘mocking’ if it is deemed to be used by the speaker to make fun of someone or something in an ‘unpleasant’ way.

A question is coded as ‘playful’ (or ‘jocular’) if it is judged to express lightheartedness intended for one’s own or others’ amusement rather than be taken seriously. These two affective meanings were regarded as not mutually exclusive – a question can be just ‘playful’, just ‘mocking’, neither ‘playful’ nor ‘mocking’ or both, as illustrated in (13).

-
- (13) a. A: *May be a B/B+*
 ‘Maybe a B/B+’
 B: *She has flu issit*
 ‘Does she have the flu?’
 <COSEM:18CM10–3741–25CHM–2017> (non-playful, non-mocking)
- b. A: *This is dam brilliant*
 ‘This is damn brilliant.’
 B: *Is it I’m the new face of psych soc? Hahaha!*
 ‘Am I the new face of Psych Soc? Hahaha!’
 C: *is that josh’s shirt design for the competition? hahaha??*
 ‘Is that Josh’s shirt design for the competition? Hahaha?’
 <COSEM:17CM02–18924–21CHM–2016> (playful, non-mocking)
- c. *is it my oily skin too oily everything tak jalan*
 ‘Is my oily skin too oily that nothing worked? (lit. everything did not move)’
 <COSEM:18MF03–15373–23MAF–2017> (mocking, non-playful)
- d. A: *Do you know where we are? Cos im lost in your eyes*
 ‘Do you know where we are? Because I am lost in your eyes.’
 B: *HAHAHA MC issit you pro at pickup lines*
 ‘Haha. MC are you a professional at pickup lines?’
 <COSEM:17IF02–2473–21CHF–2015> (playful, mocking)
-

Note that only the preceding and succeeding utterances were considered when the coders made their judgments. This can potentially skew the results. For instance, in the case of making judgments for rhetoricity, the yes/no response may appear in utterances much later than the one immediately succeeding (e.g., the responder could be interrupted by another interlocutor in the group chat). In the case of affect, the coder might code an utterance as ‘non-playful’ even if the utterance is, due to the lack of evidence in the preceding and/or succeeding utterances. However, we decided to use

this approach to simplify the coding process, with the intention to code more data. We hope that the quantity of data will compensate for any potential noise.

5. Findings and results

5.1 General distribution

We find all three types of *is it* constructions in CoSEM: polarity tag *is it* (14), non-polarity tag *is it* (15), and clause-initial *is it* (16):

-
- (14) *the thing isn’t working yet is it*
 ‘The thing isn’t working yet, is it?’
 <COSEM:18CF41–17499–22CHM–2017>
- (15) a. *Casper asked someone is it???*
 ‘Did Casper ask someone?’
 <COSEM:17CF32–1810–24MAF–2016>

- b. *U watch too much drama isit*
‘Is it because you watch too much [TV] drama?’
<COSEM:18CF41–7664–22CHF–2017>
- c. *You live in east loop isit*
‘Do you live in East Loop?’
<COSEM:18CF41–34767–23CHM–2018>

(16) *Is it u got a thing for me?*
‘Do you have a thing for me?’
<COSEM:19MM01–2587–19MAF–2016>

(17) a. *I dunno isit pregnant or belly*
‘I don’t know if it is because I am pregnant or because I have a belly.’
<COSEM:17CF19–2733–21CHF–2016>

- b. *But not sure is it real.*
‘But I am not sure if it is real.’
<COSEM:18CF52–22058–52CHF–2017>

Apart from the three aforementioned variants, we also find *is it* in embedded interrogatives without a complementizer (17), where *is it* seems to be analog to *whether or not* or *if it is*. This construction falls beyond the scope of our investigation of question marking, as sentences with this construction cannot be interpreted as yes/no questions, unlike the other sentences. We hope to examine it further in the future.

Table 3 summarizes the distributions of *is it* constructions in CoSEM.

Overall, we found several variants of CSE *is it* constructions. We now proceed to investigate whether age, sex, race, nationality, rhetoricity, and affect condition the use of these variants. Due to the relatively small number of tokens, we decided to use all the tokens for the analysis of variation. This meant not being able to control for sample size or proportion differences between social groups (e.g., discarding tokens belonging to one group to ensure balance): age groups

(below 30 = 99%, above 30 = 1%), sex groups (male = 41%, female = 59%), racial groups (Chinese = 80%, non-Chinese 20%), nationality groups (Singaporean = 94%, non-Singaporean 6%). Although a balanced dataset is the gold standard, our unbalanced dataset should not pose too much of a problem for our analysis, as the statistical tests we employed account for it. Sample size is considered alongside magnitude of difference and variability in the data when deciding whether a correlation/effect is statistically significant (see Winter, 2020: 159 for an in-depth discussion); that is, sample size is taken into account in our measure of (effect) confidence – the *p*-value.

5.2 Polarity vs. non-polarity

To determine the factors that influence CSE speakers’ choice of tags, we attempted to model the likelihood of using the non-polarity tag over the polarity variant using social factors. However, because of the lack of data for polarity questions

Table 3: Distribution of *is it* constructions in CoSEM

<i>Is it</i> feature and variants			Raw frequency	Relative frequency
Yes-no question marker	Question tag	Polarity	3	0.16%
		Non-polarity	1509	79.34%
	Clause-initial <i>is it</i>		383	20.14%
<i>Is it</i> in embedded interrogatives without complementizer			7	0.37%
Total instances			1902	100%

($n = 3$, see Table 2), we were not able to create a model nor were we able to comment on the development of *is it* in the context of polarity. The dominance of non-polarity tags over polarity tags in our data indicates that the use of non-polarity tags in CSE is stable.

5.3 Question tags vs. clause-initial *is it*

To determine the factors that influence CSE speakers' choice of *is it* question marker, we modeled the likelihood of a CSE speakers' use of clause-initial *is it* over the use of *is it* tags. We used the manually coded rhetoricity, affect, and orthographic choice factors in addition to the social factors mentioned earlier. We wanted to control for potential individual effects (individual variation) by modeling in random intercepts for participant/interlocutor. This was, however, not possible because there were insufficient *is it* questions per individual in the corpus.

The resulting model showed an effect of rhetoricity, playful affect, age, and nationality on the likelihood to use clause-initial *is it* (Table 4). There was no evidence of mocking affect, race, sex, and choice of orthographical variants (i.e., concatenated/non-concatenated, *s*-/*z*-variants) correlating with the likelihood of using the innovative marker over question tags.

CSE speakers were more likely to use clause-initial *is it* in a rhetorical question (mean = 0.216, SD = 0.411, $n = 1,510$) compared to a non-rhetorical question that has a *yes*, *no*, or *maybe* response (mean = 0.149, SD = 0.357, $n = 382$) (Figure 1).

Speakers were also more likely to use clause-initial *is it* in contexts where they are being playful (mean = 0.386, SD = 0.488, $n = 207$) (18 and Figure 2) and were less likely to use it in other contexts (mean = 0.179, SD = 0.38, $n = 1,685$). Overall, it appears that CSE speakers use the innovative *is it* in playful, rhetorical yes-no questions.

(18) *LOL IZZIT U CHECKING WHO GONNA BE THERE SO U CAN SEE IF U SHOULD HANKYPANKY*
 'Lol. Are you checking who is going to be there so you can see whether you should hanky-panky?'
 <COSEM:18CF50–9557–21CHF–2016>

Younger speakers tended to use clause-initial *is it* more compared to older speakers (Figure 3). One possible reason why this pattern surfaced is because of a potential change-in-progress led by

the youth. Younger speakers may be innovating the conventionalized question marker – the non-polarity tag *is it* – for reasons such as distinguishing themselves from older speakers, etc. This

Table 4: Logistic regression results – likelihood to use the innovative marker over question tags (observations = 1,892, $R^2 = 0.049$, no random effects, multi-level categorical variables coded using Weighted Helmert coding conventions, reference levels and statistically significant p -values in boldface, $\alpha = 0.05$)

Predictors	Log-Odds	SE	CI	p
(Intercept)	0.23	0.62	−0.85–1.58	0.711
Age	−0.08	0.03	−0.14 – −0.03	0.007
Rhetoricity (rhetorical vs. non-rhetorical)	−0.42	0.16	−0.74 – −0.12	0.008
Affect – Mocking (mocking vs. non-mocking)	0.21	0.29	−0.36–0.79	0.467
Affect – Playful (playful vs. non-playful)	0.92	0.23	0.47–1.36	<0.001
Race: Chinese (Chinese vs. non-Chinese)	0.29	0.18	−0.06–0.65	0.118
Nationality: Singaporean (Singaporean vs. Non-Singaporean)	1.42	0.49	0.56–2.51	0.004
Sex (male vs. female)	−0.21	0.12	−0.45–0.04	0.096
Variant – Concatenation (concatenated vs. unconcatenated)	0.13	0.12	−0.11–0.37	0.279
Variant – Consonant choice (s vs. <i>z</i>)	−0.53	0.39	−1.28–0.28	0.18

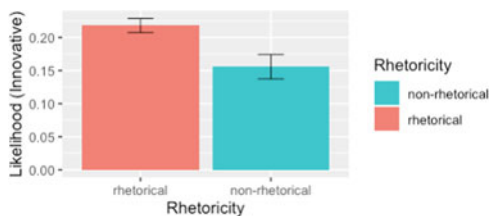


Figure 1. Effect of rhetoricity on likelihood to choose clause-initial *is it*

claim is not far-fetched, as prior sociolinguistic research investigating language change has often reported younger speakers – the ‘movers and shakers in the community’ and ‘people with energy and enterprise’ – innovating language for various reasons like countering adult norms (Maclagan, Gordon & Lewis, 1999: 19).

Apart from younger speakers, Singaporean CSE speakers were more likely to use clause-initial *is it* (mean = 0.212, SD = 0.408, n = 1,785) compared to non-Singaporeans (e.g., Chinese from the People’s Republic of China, Filipino, Taiwanese, see full list in Gonzales et al., 2021) (mean = 0.0467, SD = 0.212, n = 107) (Figure 4). This is perhaps because the construction is linked to the Singaporean identity. Singaporean speakers may have used more clause-initial *is it* constructions compared to non-polarity tag *is it* to stress their Singaporean-ness. Some evidence of this can be found in our consultation with three native CSE speakers, who identified the construction as a Singaporean feature. The correlation between the clause-initial *is it* construction and the Singaporean group provides strong evidence that the fronted construction is a local Singaporean innovation, and not an innovation introduced by non-Singaporeans.

Our inability to find evidence of a racial contrast (Chinese vs. non-Chinese) relating to the likelihood of the use of clause-initial *is it* suggests that the construction is uniformly used across all

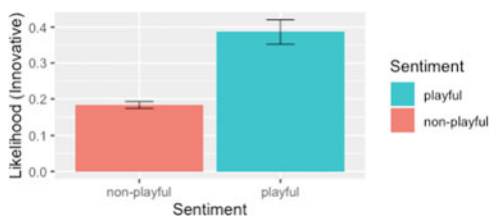


Figure 2. Effect of playful affect on likelihood to choose clause-initial *is it*

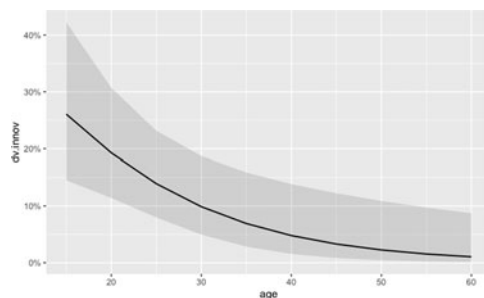


Figure 3. Effect of age on likelihood to choose innovative marker

racial groups in Singapore, of which the Chinese, Malays and Indians are the major ones. It is highly likely that the novel CSE *is it* emerged out of contact between English and the Sinitic substrate languages used by Chinese-dominant speakers of CSE (e.g., Mandarin-dominant speakers) (see Section 6.2). We hypothesize that the feature first became popular among the Chinese CSE-speaking population before it spread to (and stabilized within) the non-Chinese CSE-speaking groups. A potential reason could be because CSE *is it* closely resembles the standardized English tag question *is it*, both in terms of form and function.

In summary, we provided evidence that the use of *is it* in CSE is distinct from the use of *is it* in standardized English: the use of both non-polarity *is it* as well as clause-initial *is it* as default question markers is robust and stable in CSE but not in standardized English. We have also shown that the variation in the use of these question markers is systematically conditioned by rhetoricity, playful affect, age, and nationality. However, we have not found evidence that mocking affect, sex, race, and orthographical choice conditioned this variation.

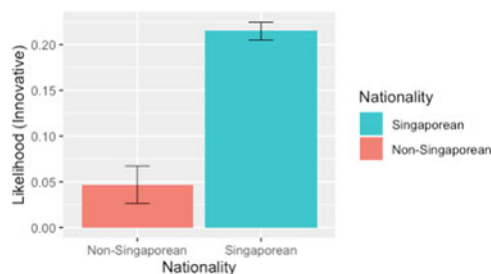


Figure 4. Effect of nationality on likelihood to choose innovative marker

6. General discussion

At the outset of this paper, we sought to identify the distribution of the *is it* question marking variants in CSE. We found that CSE speakers used clause-initial *is it* and non-polarity tags alongside polarity ones, with a significant preference for non-polarity tags, in contrast to standardized English. Another objective was to answer the question of whether the *is it* variants in CSE are conditioned by linguistic, orthographic, and social factors. Given the data used for this study, we were not able to find evidence of linguistic or social conditions affecting the use of a non-polarity construction over a polarity one. We were, however, able to identify correlations between the use of clause-initial *is it* and certain pragmatic factors, such as rhetoricity and playfulness. In addition, our results showed that certain social groups tend to use it more than others. A summary of what has been discussed in the previous sections and how they relate to our hypotheses can be found in Table 5.

Obviously, several of our predictions were not confirmed: several factors (e.g., mocking affect) that were hypothesized to have an effect on a CSE speaker's likelihood of using a particular variant of the *is it* construction turned out not to have one. These factors are represented in gray in Table 5. In any case, we were able to account for at least some of the variation occurring with CSE *is it*. We pinpointed some pragmatic and sociolinguistic factors that condition CSE *is it* use, consequently making it possible for us to comment on its systematicity, as well as speculate its development.

6.1 The systematicity and distinctiveness of CSE *is it*

Based on the patterns of variation in the use of CSE *is it* presented earlier, we propose the following structural conventions: in CSE, the polarity tag is rarely used to elicit a yes-no response. Instead,

one uses the non-polarity tag – the default *is it* question marker. If a speaker wishes to be playful or hopes not to get a response, as in a rhetorical question, they typically use clause-initial *is it* (Table 6).

These conventions are notably different from the conventions of standardized English, which by default use the polarity tag. In CSE, the non-polarity tag is used as the default yes-no question marker in CSE. Another structural convention that distinguishes CSE from standardized English is the use of the clause-initial *is it* question marker, which is not found in standardized English. Given the use of this construction alongside the tag construction, CSE *is it* seems to have developed from a canonical question tag to a general yes-no question marker.

6.2 The development of CSE *is it*

Based on the effects of the social factors identified in the model, we are also able to comment on the development of CSE *is it*. To recapitulate, we found an age effect for the innovation: the clause-initial *is it* is associated with the youth. This suggests that younger speakers were the ones who started and spread the innovative clause-initial *is it* construction. The order in which innovations entered CSE can be deduced by looking at the age-grading results using the apparent time model (Sankoff, 2006) as well as our findings from comparing various corpora.

Using British English, arguably the source variety of CSE, as a point of comparison, we propose that the variety of Singapore English in the early post-colonial period recognized both polarity and non-polarity tag constructions. It most likely had the polarity tag as the 'unmarked' or default variant and the non-polarity tag as the 'marked' or conditioned variant. From the distributional differences between the British English (early

Table 5: Summary of results (gray text = no evidence of effect, contradicting hypotheses)

Feature	Linguistic condition	Social condition
Question tag <i>is it</i> (i.e., clause-final tag yes-no question marker <i>is it</i>)	Non-rhetorical	Older
	Non-playful	Non-Singaporean
	Non-mocking	Non-Chinese Male
Innovative <i>is it</i> (i.e., non-tag, clause-initial yes-no question marker <i>is it</i>)	Rhetorical	Younger
	Playful	Singaporean
	Mocking	Chinese Female

Table 6: Proposed structural conventions in CSE *is it*

<i>Is it</i> question-marking variant	Conventions
Polarity tag	<ul style="list-style-type: none">• Not used
Non-polarity tag	<ul style="list-style-type: none">• Used as default <i>is it</i> question marker• Used when asking for explicit verification
Clause-initial <i>is it</i>	<ul style="list-style-type: none">• Used when not expecting a yes-no-maybe answer (rhetorical)• Used when trying to be playful

CSE) and 1990s CSE, represented by ICE–SIN data, we argue that the polarity vs. non-polarity distribution shifted over time, perhaps before the 1990s. Instead of only being used in specific conditions, the non-polarity variant was generalized to all *is it* constructions. That is, CSE speakers using *is it* constructions eventually used the non-polarity *is it* tag almost exclusively.

However, the use of the non-polarity tag as the default *is it* question marker seems to be undergoing a process of change again recently. Based on findings from CoSEM, clause-initial *is it* emerges

as a competing variant: a conditioned variant that tends to be used more in rhetorical, playful contexts, and by younger Singaporeans. If this pattern of change holds, *ceteris paribus*, it is very likely that the recent fronted *is it* innovation would take over the non-polarity tag as the general *is it* question marker, especially since the marker is found to be an index of Singaporean identity. There is already evidence of this in present day CSE: there are some cases where the fronted innovation is used in non-playful, non-rhetorical contexts, as seen in (19), (20), and (21):

- (19) *Is it I send recordings now?*
'Do I send the recordings now?'
<COSEM:18CM10–9519–21CHF–2017>
- (20) *Issit we don't have film lecture?*
'We don't have a film lecture, right?'
<COSEM:19MM13–15292–23MAMIW–2018>
- (21) *Is it the rain stopping?*
'Is the rain stopping?'
<COSEM:17CF19–2104–23CHM–2016>

We summarize the developmental trajectory of CSE *is it* in Table 7.

Currently, we do not have definitive evidence of what specifically motivated these changes. One possibility is internal change or language drift. The distributional changes could have occurred without any external influence. A more likely

catalyst is language contact. It is possible that the distribution shifted due to the influence of languages spoken alongside CSE, such as Mandarin, although more evidence is needed to substantiate this claim. Part of the Mandarin yes-no question forming system (i.e., the clause-initial A-not-A construction, Table 8) might have been transferred

Table 7: Proposed developmental trajectory (distribution and conditions)

	CSE genesis	1990s	2019	2019 onwards
General	polarity tag	non-polarity tag	non-polarity tag	clause-initial <i>is it</i> (?)
Conditioned	non-polarity tag		clause-initial <i>is it</i>	

Table 8: Comparison of yes-no question-forming strategies between CSE, Mandarin and standardized English (question: ‘Do you like her?’)

Strategy	CSE	Mandarin	English
auxiliary inversion	<i>Do you like her?</i> do 2.SG like 3.SG		<i>Do you like her?</i>
clause-final	<i>You like her isit?</i> 2.SG like 3.SG Q	你 喜欢 她 吗? <i>Nǐ xǐhuān tā ma?</i> 2.SG like 3.SG Q	
clause-initial	<i>Isit you like her?</i> Q 2.SG like 3.SG	你 喜欢 她, 是不是? <i>Nǐ xǐhuān tā, shìbùshì?</i> 2.SG like 3.SG yes-NEG-yes 是不是 你 喜欢 她? <i>Shìbùshì nǐ xǐhuān tā?</i> yes-NEG-yes 2.SG like 3.SG	<i>You like her, don't you?</i>
clause-medial (adverbial)		你 是不是 喜欢 她? <i>Nǐ shìbùshì xǐhuān tā?</i> 2.SG yes-NEG-yes like 3.SG	

to CSE. While the clause-medial strategy seems to have not been selected by CSE speakers, as evidenced by the lack of clause-medial *is it* (e.g., **You isit like her?*) constructions in the corpus, the clause-final and clause-initial strategies seem to have been transferred to CSE, as indicated by the robustness of these strategies in CSE. The account involving Mandarin influence is justified as Mandarin is a dominant language (if not the most dominant non-English language) in Singapore. Mandarin is promoted as the lingua franca of Chinese Singaporeans (see e.g., Lim, Chen & Hiramoto, 2021), who form the majority of the Singaporean population (Singapore Department of Statistics, 2019).

In summary, we have identified a potential contact-induced innovation in CSE: clause-initial *is it*. We argue that this innovation is a recent one, introduced by younger Singaporean CSE speakers. A likely reason for the proliferation of clause-initial *is it* is its ability to function as a salient marker of a youthful, Singlish-speaking, Singaporean identity, similar to how, for example, question tags like *msibá* ‘is it not’ in Lánang-uè spoken in the Philippines are emblematic of the relatively young, hybrid Lannang identity (Gonzales, 2017; Gonzales, 2021). As of this point, we have yet to come across evidence that would allow us to definitively attribute the emergence of this innovation to any one particular factor (or set of factors); however, a language contact account that involves Mandarin seems very likely.

7. Conclusion

Drawing on CoSEM, a 3.6-million-word corpus of CSE finger speech, we investigated the use of *is it* in contemporary CSE. Apart from wanting to establish the distribution of *is it* use in CSE, we also wanted to determine the conditions on which CSE users utilize different *is it* constructions. By analyzing the frequency distributions and statistically modeling the likelihood of CSE users’ choice of *is it* variants, we were able to shed some light on how CSE users construct *is it* questions.

CSE speakers seemingly use different strategies (e.g., polarity tag, non-polarity tag, clause-initial *is it*) to relay the same thought, but a closer look at the variation reveals that CSE speakers have specific conventions for structuring yes-no questions involving *is it*. Specifically, we found that CSE speakers use two *is it* question constructions: (1) clause-final *is it* (non-polarity question tag *is it*) for non-rhetorical or non-playful questions, and (2) clause-initial *is it* for rhetorical, playful questions. We also found that the first construction is associated with older Singaporean speakers whereas the second is used robustly by younger speakers and speakers who identify as Singaporean. There was no evidence of sex, race, orthography, or mocking affect conditioning the choice of construction or strategy.

These (socio)linguistic conventions, like conventions in other languages, are not static. They can change. For instance, although most CSE speakers follow the conventions listed above, we

identified instances where some speakers (younger speakers) are generalizing the use of clause-initial *is it* such that it is used even in non-playful or non-rhetorical contexts, suggesting that there is ongoing change in progress.

Currently, we do not have sufficient evidence to definitively pinpoint the catalysts for change and/or variation in the use of CSE *is it* variants. However, an account involving language contact where Mandarin is a primary contributor seems very likely, based on a comparison of structures between CSE, Mandarin and English. Overall, by analyzing variation in CSE *is it* use, we were able to identify linguistic conventions, observe correlations between social groups and particular *is it* variants (sociolinguistic conventions), and comment on the development of the *is it* question-marking system in CSE.

Acknowledgments

We gratefully acknowledges the support provided for this project by Singapore Ministry of Education Academic Research Fund Tier 1 under WBS R-103-000-167-115 as well as Singapore Ministry of Education Academic Research (Conference) Fund Tier 1 under R-103-000-174-115. We also thank class members of EL3211: Language in Contact and of EL3551: Undergraduate Research Opportunity Program (UROP) at the National University of Singapore. We'd also like to show our sincere gratitude to our research assistants for this paper, Vincent Pak and Lauren Yeo. Last but not least, we are immensely grateful to Andrew Moody and two anonymous reviewers' comments on an earlier version of this manuscript.

Notes

- 1 We opted to use the term 'race' over 'ethnicity' because it is the conventional term used in Singapore.
- 2 We use the term 'concatenation' broadly.

References

Bao, Z. 2010. 'A usage-based approach to substratum transfer: The case of four unproductive features in Singapore English.' *Language*, 86(4), 792–820.

Bao, Z. & Wee, L. 1999. 'The passive in Singapore English.' *World Englishes*, 18(1), 1–11.

Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. 2015. 'Fitting linear mixed-effects models using lme4.' *Journal of Statistical Software*, 67(1), 1–51.

Columbus, G. 2010. "'Ah lovely stuff, eh?" Invariant tag meanings and usage across three varieties of English.' In S. T. Gries, S. Wulff & M. Davies (eds.), *Corpus-Linguistic Applications: Current Studies, New Directions*. Leiden: Brill, pp. 85–102.

Faulkner, X. & Culwin, F. 2005. 'When fingers do the talking: A study of text messaging.' *Interacting with Computers*, 17(2), 167–185.

Gonzales, W. D. W. 2017. 'Chinese Filipinos tag their questions, *kiam si?* Some notes on tag questions in Philippine Hybrid Hokkien.' *Philippine ESL Journal*, 19, 84–99.

Gonzales, W. D. W. 2021. 'Filipino, Chinese, neither, or both? The Lannang identity and its relationship with language.' *Language & Communication*, 77, 5–16.

Gonzales, W. D. W., Hiramoto, M., Leimgruber, J. & Lim, J. J. 2021. 'The Corpus of Singapore English Messages (CoSEM).' *World Englishes*, 1–18.

Greenbaum, S. 1988. 'A proposal for an international computerized corpus of English.' *World Englishes*, 7(3), 315.

Hiramoto, M. 2012. 'Pragmatics of the sentence-final uses of *can* in Colloquial Singapore English.' *Journal of Pragmatics*, 44(6–7), 890–906.

Hiramoto, M. & Sato, Y. 2012. 'Got-interrogatives and answers in Colloquial Singapore English.' *World Englishes*, 31(2), 198–207.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. 2019. 'lmerTest package: Tests in linear mixed effects models.' *Journal of Statistical Software*, 82(13), 1–26.

Leimgruber, J. R. E. 2011. 'Singapore English.' *Language and Linguistics Compass*, 5(1), 47–62.

Leimgruber, J. R. E. 2016. 'Bah in Singapore English.' *World Englishes*, 35(1), 78–97.

Leimgruber, J. R. E., Lim, J. J., Gonzales, W. D. W. & Hiramoto, M. 2020. 'Ethnic and gender variation in the use of Colloquial Singapore English discourse particles.' *English Language and Linguistics*, 1–20.

Lim, L. 2007. 'Mergers and acquisitions: On the ages and origins of Singapore English particles.' *World Englishes*, 26(4), 446–473.

Lim, J. J., Chen, S. C. & Hiramoto, M. 2021. "'You don't ask me to speak Mandarin, okay?" Ideologies of language and race among Chinese Singaporeans.' *Language & Communication*, 76, 100–110.

MacLagan, M. A., Gordon, E., & Lewis, G. 1999. Women and sound change: Conservative and innovative behavior by the same speakers. *Language Variation and Change* 11: 19–41.

McWhorter, J. 2007. *Language Interrupted: Signs of non-native acquisition in standardized language grammars*. Oxford: Oxford University Press.

McWhorter, J. 2013. 'Txtng is killing language. JK!!!' Online at https://www.ted.com/talks/john_mcwhorter_txtng_is_killing_language_jk/discussion#t-1876 (Accessed October 10, 2020).

Nelson, G. 2002. 'International Corpus of English: The Singapore Corpus User Manual.' Online at https://courses.washington.edu/hypertext/ice-cwb/Manuals/ICE-SIN_Manual.PDF (Accessed September 9, 2020).

Platt, J. 1975. 'The Singapore English speech continuum and its basilect "Singlish" as a "creoloid".' *Anthropological Linguistics*, 17(7), 363–374.

Pratt, T. 2021. 'Affect in sociolinguistic style.' *Language in Society*: 1–26. doi:10.1017/S0047404521000774

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.

R Core Team. 2013. 'R: A language and environment for statistical computing.' Online at <http://www.R-project.org>.

Richards, J. C. & Tay, M. W. J. 1977. 'The *la* particle in Singapore English.' In W. Crewe (ed.), *The English*

- Language in Singapore*. Singapore: Eastern Universities Press, pp. 141–156.
- Sankoff, G. 2006. 'Age: Apparent time and real time.' In K. Brown (ed.), *Elsevier Encyclopedia of Language and Linguistics: Second Edition*. Amsterdam: Elsevier Science, pp. 110–116.
- Siemund, P. 2018. *Speech Acts and Clause Types: English in a Cross-Linguistic Context*. Oxford: Oxford University Press.
- Singapore Department of Statistics. 2019. 'Singapore residents by age group, ethnic group and sex, end June, annual.' Online at <https://data.gov.sg/dataset/resident-population-by-ethnicity-gender-and-age-group> (Accessed October 20, 2020).
- Starr, R. L. & Balasubramaniam, B. 2019. 'Variation and change in English /r/ among Tamil Indian Singaporeans.' *World Englishes*, 38(3–4), 630–643.
- Winter, B. 2020. *Statistics for Linguists: An Introduction Using R*. New York: Routledge.
- Wong, J. 2014. *The Culture of Singapore English*. Cambridge: Cambridge University Press.
-