Chapter 11

A comparative corpus analysis of the frequencies and functions of *so* between native and non-native speakers of English in Hong Kong

Phoenix Wai Ying Lam

Orcid.org/0000-0001-8885-1800

Abstract

Multifunctional forms such as *so*, *well* and *like* in English are ubiquitous in discourse. In spoken discourse where they are syntactically optional and carry no or little propositional meaning, these frequently-occurring items are commonly known as discourse markers, pragmatic markers or discourse particles, and serve a variety of important discourse functions in the textual, interpersonal and interactional domains, sometimes even concurrently in one instance. While previous studies on these items have provided invaluable insights into their functions and register variation, such research work was mostly based on data of native speakers. Much less is known about how and the extent to which non-native speakers use such multifunctional forms. In turn, it remains unclear whether these forms constitute part of the pragmatic toolkit for second/foreign language speakers. Drawing on a corpus of spoken English collected in Hong Kong from a range of situational settings, the present study investigates and compares the frequencies and functions of the multifunctional form *so* by Hong Kong Chinese non-native speakers of English and native speakers of English. Findings from the study show similarities in the ratio of the discourse use to propositional use of the word between the two groups of speakers. However, discrepancies are observed in the frequency of certain discourse functions by the two groups. These results are considered in relation to such critical issues as linguistic performance, pragmatic competence and cultural preference. For language teachers and literacy educators, the pedagogical implications of how such comparative findings can be meaningfully deployed to enhance second/foreign language education, especially in terms of the pragmatic aspect of communication, are discussed.

key words: corpus, discourse markers, discourse particles, Hong Kong, pragmatic markers, so

Introduction

Multifunctional forms such as *so*, *well* and *like* in English are ubiquitous in discourse. In spoken discourse where they are syntactically optional and carry no or little propositional meaning, these frequently-occurring items are commonly known as discourse markers, pragmatic markers or discourse particles, and serve a variety of important discourse functions in the textual, interpersonal and interactional domains, sometimes even concurrently in one instance. While their ubiquity, versatility and importance in spoken discourse are generally well-acknowledged, their role in language and literacy education remains under-researched. This is reflected in the relatively small number of studies on these items based on data of non-native speakers. As such, much less is known about how and the extent to which non-native speakers use such multifunctional forms. In turn, it remains unclear whether these forms constitute part of the pragmatic toolkit for second/foreign language speakers.

Drawing on a corpus of spoken English collected in Hong Kong from a range of situational settings, the present study investigates and compares the frequencies and functions of the multifunctional form *so* by Hong Kong Chinese non-native speakers of English and native speakers of English. Given the global role of English as a lingua franca, the dichotomy between native and non-native speakers of a language is increasingly being challenged (see, for example, Prodromou 2003, on the notion of a "successful user of English" and Davies 2003, on the myth of native speakers). It is therefore important to stress at the onset that in the present study the term 'native speakers of English' is simply used to refer to speakers whose first language is English, as opposed to speakers whose first language is not English, without any political or attitudinal implication. Findings from this comparative corpus analysis will lead to the discussion of pedagogical implications for second/foreign language education, especially in terms of the pragmatic aspect of communication.

Studies on *so* as a discourse marker

This section first reviews large-scale studies conducted on *so* as a discourse marker focusing on its overall meaning and function. It then discusses research work on *so* in specific settings and genres based on data from native and non-native speakers respectively.

The monosyllabic word *so* is one of the most frequently occurring words in the English language, and yet few studies have examined it compared with other discourse markers such as *well*. For instance, *so* is not included in the list of "interactional signals and discourse markers" by Stenström (1994, p. 59). The study of *so* can be traced back to its function as a discourse connective. For Grice (1989), *so* is one of the non-truth-conditional discourse connectives which are used to perform higher-order speech-acts, i.e. to signal how speakers comment on some more basic or lower-order speech-acts. Specifically, *so* is associated with the higher-order speech-act of "explaining" (Grice 1989, p. 362). Closely related to the Gricean analysis of *so* as a connective is the study of *so* under relevance theory. For Blakemore (1988), *so* is a constraint on relevance. She argues that the main function of *so* is to guide the listener "to establish an inferential connection" (Blakemore 1988, p. 193), with four different uses of *so* depending on its co-text. In conjoined utterances, *so* connects two propositions together and suggests either a causal effect or a deductive consequence between them. When *so* occurs initially without an explicit linguistic antecedent, it signals that the utterance it prefaces is relevant to the situational context at hand. When *so* is used in responses as a stand-alone utterance or as in *so what*, it implies that the speaker is "unable to see the significance of what someone has said" (Blakemore 1988, p. 189). By imposing a constraint on relevance, *so* minimizes the processing effort involved in interpretation, and thus enhances the efficiency of the cognitive system (Sperber and Wilson 1995).

For Schiffrin (1987), *so* conveys a central meaning of result. Studying *so* together with *because* under the notion of coherence, Schiffrin (1987) gives a comprehensive account of *so* in three aspects: a complementary marker of main idea units, a marker of resultative relations, and a turn-transition device which signals a potential speaker change. Also within the framework of coherence, Redeker (1990) posits that as a marker of pragmatic structure, *so* indicates a sequential relation between "successive elements in a chain of events" (Redeker 1990, p. 373) or prefaces a conclusion made by the speaker. In her later work, Redeker (2006,

p. 339) describes the use of turn-internal instances of *so* as "attentional cues" and focuses on its function in marking transitions in discourse.

More recent studies of *so* as a discourse marker have examined it in specific settings and genres. In police interviews in England, Johnson (2002, p. 103) observes that the word is often used to preface questions by the interviewers as a "topic developer" or "topic sequencer" to mark topic transitions. Importantly, she also suggests that this use of *so* can serve the pragmatic effects of challenging the interviewees. It is therefore not only the function of structural organisation in the textual domain that *so* achieves but also the function of summative evaluation in the interpersonal domain in this kind of asymmetrical interaction. In a series of university computer science seminars in Australia, Rendle-Short (2003) argues similarly that *so* orients the listener to the overall structure of monologic talk. Her analysis shows that *so* is "position sensitive" (Rendle-Short 2003, p. 55) and its role and function vary according to whether it occurs at the beginning, in the middle, or at the end of a section of talk. Specifically, *so* has an orientation function and marks a new topic at the start of a section. By contrast, it serves an explanatory function and signals a digression from the main topic of talk in the middle of a section. At the end of a section, it prefaces a resolution often in the form of personal assessment. The word therefore also has important roles to play in both the textual and interpersonal domains in academic monologues. In American conversations, Bolden (2009, p. 977) notes that *so* operates also in the interactional domain of communication by prefacing "new and resumed pending interactional agendas". As such, the word is used to signpost not only the textual boundaries of topic but also the initiation of new interactional courses of action or the resumption of delayed courses of action. In these uses, *so* occurs in the sequence-initial position and helps answer the question of "why that now" in the conversations (Bolden 2009, p. 996). Most recently in a range of spoken and written registers in New Zealand English, Yin (2019) shows the variation in the pattern of use and meaning of *so*. Academic lectures and conversations are found to have a higher occurrence of the item than the other three registers in comparison. They are also the only two registers in which *so* conveys the meaning of "returning back to the main thread" (Yin 2019, p. 36). As such, these studies have illustrated the variety of contexts and co-texts in which *so* occurs and the range of textual, interpersonal and interactional functions that *so* accomplishes as a discourse marker as used by native speakers.

In the relatively small number of dedicated studies of *so* based on data of non-native speakers, contradictory results have been reported concerning the frequency of use of this linguistic item. By using a number of sources including spoken and written corpora of British English and Chinese learner English supplemented by English textbooks and Chinese conversations, He (2002) shows that *so* is used more frequently in the speech and writing of Chinese EFL speakers compared with native speakers of English. Possible reasons for the more frequent use and the different uses of the word are suggested to be associated with the unawareness of stylistic difference in spoken and written English, limited exposure to English, textbook influence and negative transfer from learners' first language. By contrast, less frequent use of *so* is observed in the speech of German speakers. In a study comparing the use of discourse markers by American and German students based on conversations collected from a movie re-telling experiment, Müller (2004) reports that the German EFL speakers use *so* much less frequently than the Americans. A possible reason for the difference is the phonological and semantic similarity between *so* and the German word *also*. In order to

avoid the German-sounding association when speaking English, German EFL speakers thus use *so* less frequently. In a follow-up study, Müller (2005) explains the differences which are statistically significant in the discourse functions of *so* between the two groups of speakers in more detail. Specifically, the sequential, resultative and the summarising functions of *so* are found to be used more frequently by American speakers than by German speakers. The most significant difference is found in the use of *so* to signal a sequential relationship. Compared with the Germans, the Americans use this function eight times more often. This suggests that the difference in the use of *so* between the two speaker groups may be the greatest on the textual level. Also focusing on European learners of English, Buysse (2012) examines how *so* as a multifunctional discourse marker is used by Belgian and British students. While all the ten functions as identified in the study and mapped onto the Hallidayan ideational, interpersonal and textual domains are found in the data of both speaker groups, some functions are more commonly used by one group than the other. Specifically, the elaborative and self-corrective functions are significantly more frequent in the speech of Belgian learners, whereas the sequential use of *so* is more popular among British speakers. Regarding the overall frequency of use of the marker, Buysse (2012) also reports conflicting results with Müller (2004) in that the Belgian learners use *so* considerably more often than their native peers. These studies thus have suggested that the language background of speakers influences the frequency of use and function of *so*, and that more comparative studies of speakers of different mother tongues are needed to further our understanding of the crucial but at times subtle variations in the use of this multifunctional marker.

Data and methodology

The Hong Kong Corpus of Spoken English (HKCSE)

The data on which the current study is based is from the Hong Kong Corpus of Spoken English (HKCSE hereafter). As an approximately one-million word collection of naturally-occurring speech which has been transcribed both orthographically and prosodically (see also Cheng, Greaves & Warren 2005, 2008, for further details of the corpus), the corpus primarily consists of intercultural encounters in Hong Kong in a variety of settings between Hong Kong Chinese and speakers of languages other than Cantonese, mostly native speakers of English. Since its introduction, the HKCSE has been an important source of data for a number of intercultural studies of Hong Kong (see, for example, Cheng 2007, Cheng and Warren 2006). In total, the corpus consists of more than 100 hours of recordings from 311 speech events evenly spread across four major settings in which English is spoken in the context of Hong Kong, viz. academic, business, conversational and public. Some examples of text types in the corpus include university lectures, job interviews, casual conversations and political speeches. The corpus thus contains a variety of text types which are of varying degrees of interactivity ranging from the highly scripted monologic public speeches to the spontaneous uninstitutionalized conversations. This makes the HKCSE particularly useful for the purpose of studying the full range of functions that *so* accomplishes.

Participants in the corpus can be classified into three groups according to their language background: Hong Kong Chinese whose first language is Cantonese; native speakers of English; and speakers whose first language is neither Cantonese nor English. A breakdown of

the composition of the corpus by the first language background of speakers is presented in Table 11.1.

Table 11.1. Composition of the HKCSE by the first language background of speakers

	Total (N=311)	Academic (N=29)	Business (N=112)	Conversational (N=71)	Public (N=99)
Hong Kong Chinese	669,431 (70.7%)	173,966 (82.1%)	180,485 (69.7%)	122,373 (47.5%)	192,607 (88.2%)
Native speakers of English	240,847 (25.4%)	25,655 (12.1%)	74,287 (28.7%)	120,084 (46.6%)	20,821 (9.5%)
Speakers of other languages	36,298 (3.8%)	12,244 (5.8%)	4,109 (1.6%)	15,082 (5.9%)	4,863 (2.2%)
TOTAL	946,576 (100.0%)	211,865 (100.0%)	258,881 (100.0%)	257,539 (100.0%)	218,291 (100.0%)

It can be observed from Table 11.1 that the majority of the participants in the HKCSE are either Hong Kong Chinese or native speakers of English. Across the four sub-corpora, Hong Kong Chinese contribute the greatest number of words. Their speech accounts for 70.7% of the total size of the corpus. About a quarter of the corpus consists of speech produced by native speakers of English, most notably from Australia, the United Kingdom and the United States of America. For the present study, no attempt has been made to categorise the native speakers in the HKCSE into different groups in relation to the varieties of English they speak. This is mainly because such fine-grained differentiation between different native speaker groups will return very small samples of native speakers' speech for each variety, which will in turn complicate the comparative analysis and render each native speaker group unrepresentative. In addition, the present study is mainly concerned with the comparison of speakers in terms of the first language they speak in the context of Hong Kong. In this connection, it is assumed that native speakers of English would be more similar to each other linguistically when compared with Hong Kong Chinese (Cheng 2003). Since the focus of the study is on the use of *so* by Hong Kong Chinese when compared with native speakers of English, the distinction between different varieties of native English is considered relatively less essential in the present case. After all, for a corpus which contains an extensive number of texts, it is inevitable that participants may come from different backgrounds and the varieties of English they speak may be influenced by such factors as regional and social variation, even in places where English is predominantly the first language. It is therefore not feasible, if at all possible, to compile a large collection of texts from a wide range of contextual settings with participants from a fairly homogenous speech community, especially in a city like Hong Kong where a mix of ethnic groups is present. It is therefore decided that a

simple dichotomous comparison of the use of *so* between native and non-native speakers of English is more appropriate in the context of Hong Kong.

Methods of data analysis

This study follows a corpus-based approach by first searching for all the instances of *so* in the HKCSE. Through a concordancer, every occurrence of the word has been automatically identified and then manually examined to distinguish the discourse use (D-use) and propositional use (P-use), i.e. the use of *so* as a discourse marker and as an adverb, conjunction or substitute form. Once the discourse uses have been identified, they are collated and analysed together in their linguistic co-text and context to accumulatively arrive at a functional taxonomy. In most cases, the study of each occurrence in its immediate co-text as revealed on the concordance line has been sufficient for the functional analysis, though in a small number of instances the whole texts have been consulted as well to retrieve the wider co-text and to better understand the situational setting. The functional categorisation is therefore data-driven and solely derived from the recurrent patterns observed in the corpus. In the process, the categories developed have been continuously modified to fully capture the range of functions observed in the data. Only a very small number of instances out of the total (N=138, 2%) are functionally unclassified owing to inaudible or unclear speech in the co-text. In addition, every occurrence of *so* has been studied in terms of its speaker's language background. This allows the calculation of the frequency rates of use of the word by Hong Kong Chinese and native speakers of English, and the detailed comparative analysis of the discourse functions of *so* by the two speaker groups both qualitatively and quantitatively. In establishing whether the differences observed between the two speaker groups are statistically significant, the chi-squared test is chosen as it has been widely used in corpus analysis for quantitative data involving nominal categories, especially for the comparison of frequency distributions in different social groups.

Findings

Frequency rate of use of so *as a discourse marker*

In the HKCSE, there are altogether 7,894 instances of *so* produced by Hong Kong Chinese and native speakers of English. Among them, 6,401 instances are uses of *so* as a discourse marker. It is found that the two speaker groups use *so* as a discourse marker at similar frequency rates, with only a slightly higher rate observed in the speech of the latter group when compared with the former. Table 11.2 details the frequency distribution of *so* in the speech of the two speaker groups.

Table 11.2. Frequency distribution of *so* in the HKCSE according to speakers' linguistic background (Hong Kong Chinese vs. native speakers of English)

	Hong Kong Chinese	**Native speakers of English**
Total number of words produced	669,431	240,847
Total number of *so*	5,526	2,368
Number of D-use	4,472	1,929
Number of P-use	1,023	422
D-use / total use (%)	80.93	81.46
D-rate (per 10,000 words)	66.80	80.09

Table 11.2 shows that the total number of words produced by Hong Kong Chinese is more than two times of that produced by native speakers of English in the corpus. Similarly, the total number of *so* produced by Hong Kong Chinese is approximately more than twice compared with native speakers of English. In other words, it seems that *so* is equally common in the speech of both speaker groups. In addition, the discourse-function ratio of *so* is analogous for Hong Kong Chinese and native speakers of English (80.93% vs. 81.46%). This is supported by the fact that the difference in the discourse-function ratio between the two speaker groups is *not* statistically significant ($x^2 \geq 0.487$; d.f. = 1; p = 0.485). In both speaker groups, roughly four out of five tokens of *so* are used as a discourse particle.

While a previous study has suggested that Chinese EFL learners use *so* as a discourse marker more frequently compared with native speakers of English (He 2002), findings from this quantitative analysis of *so* in the HKCSE do not support this view concerning advanced Chinese speakers of English in Hong Kong. On the contrary, it seems from the intercultural corpus that the discourse rate of *so* is in fact slightly higher in the speech of native speakers when compared with that in the speech of Hong Kong Chinese. Approximately 80 tokens of D-use *so* are found per 10,000 words in the speech of native speakers of English, whereas only about 67 tokens of D-use *so* per 10,000 words are found in the speech of Hong Kong Chinese.

Functions of so *as a discourse marker*

When only the discourse uses of *so* are considered in the corpus, six functions of *so* have been identified in the textual, interpersonal and interactional functional domains. Four of them are more frequently realised in the speech of Hong Kong Chinese, whereas two of them are more commonly found in the speech of native speakers of English. Table 11.3 shows the functional distribution of *so* in the HKCSE based on the linguistic background of speakers:

Table 11.3. Distribution of the discourse functions of *so* by linguistic background in the HKCSE (Hong Kong Chinese vs. native speakers of English)[1]

Speaker (Linguistic Background, LB)		Function						Total
		Framing	Linking	Consequential	Responsive	Processing	Turn managing	
Hong Kong Chinese	Count	1706	565	442	1182	336	152	4383
	% within Speaker (LB)	38.9%	12.9%	10.1%	27.0%	7.7%	3.5%	100.0%
Native speakers of English	Count	640	227	187	596	121	109	1880
	% within Speaker (LB)	34.0%	12.1%	9.9%	31.7%	6.4%	5.8%	100.0%
Total	Count	2346	792	629	1778	457	261	6263
	% within Speaker (LB)	37.5%	12.6%	10.0%	28.4%	7.3%	4.2%	100.0%

<u>Functions more frequently realised in the speech of Hong Kong Chinese speakers of English</u>

a. Framing

Of the 6,263 tokens of functionally classifiable instances of *so* as a discourse marker in the HKCSE, there are 2,346 instances serving the textual function of framing. They thus constitute more than one-third (37.5%) of the total, making this function the most prevalent in the corpus. This is consistent with findings reported from previous studies that *so* is a typical marker of transition (e.g. Johnson 2002, Rendle-Short 2003). As a frame, *so* signals textual transitions of topics and of discourse stages. One of the most commonly found transitions marked by *so* is the change of topics. In Example 1 from a conversation, two friends are first having a conversation on learning languages. At the beginning of the excerpt, the native speaker of English (Speaker E) is asking the Hong Kong Chinese (Speaker C) whether he is learning Mandarin. As the talk evolves, the Hong Kong Chinese appears to be reluctant to continue on the topic, which is evidenced by his repeatedly short and negative responses. Notice how Speaker C changes the topic from learning a language to the job of Speaker E's wife by using *so* after a short pause in the conversation. This co-occurrence with pauses before the use of *so* is a common feature of the marker as a frame.

(1)

E: stupid (.) what about you are you learning Mandarin or

C: no

E: not bothered

C: not bothered

E: yea practise your English first right

C: no ((laugh))

(pause)

C: **so** your wife got a job

E: yea

(HKCSE, C109)

This framing function is more frequently realised by Hong Kong Chinese speakers of English (38.9%) when compared with native speakers (34%). The difference between the two speaker groups is statistically significant ($x^2 \geq 13.377$; d.f. = 1; p = 0.000255), suggesting that the former speaker group more commonly uses *so* to mark different types of textual boundaries in the corpus.

b. Linking

Apart from serving as a boundary marker, *so* also achieves the textual function of linking by connecting parts of text together to establish coherence. In the HKCSE, 792 of the 6,263 tokens of functionally classifiable instances of *so* as a discourse marker are used as a linkage device, constituting around one-eighth (12.6%) of the total discourse use. As a connector, *so* either marks a sequential relationship between discourse units, or more often just loosely introduces additional information to the preceding segment of talk. In the latter case, it is roughly equivalent to the appending function of the conjunction *and*. In Example 2 from an academic lecture, the Hong Kong Chinese instructor is explaining the contents of a book chapter. Note his use of *so* in the middle of his talk simply to provide further details of the chapter, i.e. the number of pages it contains.

(2)

C: … and then chapter three tells you the type of transfusions and their applications okay
        so these are twenty-three pages okay chapter three …

(HKCSE, A008)

This textual function of *so* as a linking device, especially for appending or loosely explanatory (cf. Grice 1989) purposes, has not been well-documented in the literature, although the temporal, sequential aspect of *so* is discussed in Redeker (1990) and Müller (2005). In relation to the linguistic background of speakers, there is little difference in the use of the linking function by Hong Kong Chinese speakers of English (12.9%) and native

speakers of English (12.1%) in the corpus. Statistically, the difference between the two speaker groups is not significant ($x^2 \geq 0.794$; d.f. = 1; p = 0.373). As such, no noticeable qualitative and quantitative differences are found between the two speaker groups in terms of their use of *so* for textual connection.

c. Consequential

As discussed earlier in the review of the literature, the use of *so* marking resultative or consequential relationship has been discussed extensively and is often considered its quintessential function. For Schiffrin (1987), for instance, *so* is a marker of result whose extended functions are derived from this core meaning. In the HKCSE, however, instances of *so* clearly signalling a resultative relation between adjoining units are not very frequent. Of the 6,263 tokens of functionally classifiable instances of *so* as a discourse marker, only 629 instances serve the textual function of marking a resultative or consequential relationship. They make up one-tenth (10.0%) of the total discourse use. In these instances, *so* marks the following segment as a result of inference. At times, the inferential relation is strengthened by the use of other linguistic items right after *so*. In Example 3 from a conversation between a Hong Kong Chinese (Speaker C) and a native speaker of English (Speaker E), the former uses *so* and *that's why* together to emphasize that younger drivers in the bus company are a result of higher pay.

(3)

E: dri- better drivers (.) buses are clean

C: well I suppose they pay more **so** that's why they can get the er younger drivers …

(HKCSE, C013)

In these instances where *so* clearly indicates a resultative or consequential relationship, the two segments connected by *so* can be reversed in position with a replacement of *so* by *because*. This reversibility is a unique property of *so* as a marker of result, but not when it serves the other textual functions of framing or linking. This use of *so* thus explicitly expresses a causal or inferential relationship (cf. Blakemore 1988). In terms of speaker group comparison, *so* as a marker of result is employed at very similar rates by Hong Kong Chinese speakers of English (10.1%) and native speakers (9.9%) in the corpus. The very slight difference between the two speaker groups yields no statistical significance regarding this use ($x^2 \geq 0.028$; d.f. = 1; p = 0.868).

d. Processing

Another rather infrequent discourse function of *so* in the corpus concerns the interactional, rather than textual, functional domain. In the HKCSE, *so* is occasionally found as a processing device to indicate that, at the moment of uttering the word, the speaker is engaged in some kind of planning or modification. Only 457 of the 6,263 tokens of the discourse use of *so* achieve this function, making up less than one-twelfth (7.3%) of the total in the corpus. As the speaker undergoes some processing problem in communication which requires more time, *so* is used as a stalling strategy so that the different aspects of interaction, including

timing control, can be managed simultaneously and successfully. In these instances when *so* accomplishes the interactional function of process, it is frequently accompanied by pauses, fillers and false starts. In Example 4 from a conversation, the Hong Kong Chinese is explaining the time she spent overseas during her holiday. Note her use of *so* as a processing device in signalling hesitation, which is reinforced by the co-occurrences of the fillers *mm* and *um*, as well as another discourse marker *well* in its surroundings.

(4)

C: … and er I went to um to the UK for about three weeks just to visit my friends there **so** mm I think um well but if I got the chance I think I will go to Scotland

(HKCSE, C121)

The processing function of *so* has rarely been discussed in previous studies, with the exception of Buysse (2012) who classifies it as a textual device marking self-correction. As Example 4 suggests, however, there are occasions when the use of *so* is not associated with a specific paraphrase or reformulation but simply signals some difficulty in interaction. For this reason, this processing function is classified under the interactional domain in this study. In relation to the linguistic background comparison, processing *so* is more frequently found in the speech of Hong Kong Chinese speakers of English (7.7%) when compared with native speakers (6.4%), though the difference has not reached statistical significance ($x^2 \geq 2.942$; d.f. = 1; p = 0.863). Like the linking and consequential functions then, there is negligible difference in the use of *so* as a processing device between the two speaker groups.

Functions more frequently realised in the speech of native speakers of English

e. Responsive

As the second most prevalent function of *so* as a discourse marker in the HKCSE, the use of the word as a responsive signal occurs 1,778 out of 6,263 times. These instances thus constitute more than a quarter (28.4%) of the total. In these occurrences, *so* shapes a forthcoming response as initiated by prior discourse. While this responsive use of *so* also carries with its meaning an element of inference like its use as a marker of result, it operates in the interpersonal, rather than textual, functional domain. This is because the relationship marked by responsive *so* is between the ongoing discourse and the speaker, rather than between two adjacent segments of discourse. Specifically, *so* as a response marker introduces the speaker's reaction towards some preceding information, be it the speaker's own prior talk, another speaker's contribution, or even some extra-linguistic context. In the corpus, the most common type of interpersonal response prefaced by *so* is in the form of questions. This marks the question as motivated by the preceding discourse. In Example 5 from an academic supervision session on writing skills, a Hong Kong Chinese student (Speaker C) is explaining the kind of help she needs from the instructor, who is a native speaker of English (Speaker E). Notice the use of *so* by the teacher to introduce a follow-up question which seeks clarification in response to the student's request.

(5)

C: er (.) I want to (.) er I want I want for your help (.) on checking my application letter

E: **so** what kind of checking do you mean what do you mean by checking

(HKCSE, A036c)

This responsive function of *so*, which suggests that the speaker is actively taking into account what is happening in the ongoing discourse including what the other interactant(s) say, can be a useful device for establishing solidarity. It is thus frequently found to collocate with acknowledgement tokens and question words. Specifically, the responsive use of *so* occurs after such words as *yes* and *I see* which signal positive responses and receipt of information and understanding, and before such words as *what*, *who* and *how*. Contrary to the use of *so*-prefaced questions by police officers which are face-threatening and serve to challenge others in police interviews as observed by Johnson (2002), *so* may also introduce questions which are face-enhancing and help to create a supportive atmosphere. When the two speaker groups are compared in their use of this function, the native speakers of English are found to use it more frequently (31.7%) than the Hong Kong Chinese (27.0%). This difference is statistically significant ($x^2 \geq 14.505$; d.f. = 1; p = 0.00014), meaning that it is far from random that the responsive function of *so* is more common in the speech of the former speaker group than the latter.

f. Turn managing

The last function of discourse *so* as identified in the HKCSE is also the one which is the most infrequent. In the corpus, *so* as a turn-managing device only occurs 261 out of 6,263 times. Less than one-twentieth (4.2%) of the total, this function of *so* concerns the interactant's desire to hold the conversational floor, to relinquish his/her current turn, or to take over from another interactant. It thus operates in the interactional functional domain. Given its association with possible speaker change, the turn-managing *so* is frequently found at the transition relevance place at the end of a turn construction unit (cf. Schiffrin 1987). When it acts as a turn-yielding signal, *so* often suggests that something is left unsaid and can be inferred by the hearer from the information already given by the speaker. In Example 6 from a service encounter between a Hong Kong Chinese airport ground staff member (Speaker C) and a passenger whose mother tongue is English (Speaker E), the passenger uses *so* at the end of his utterance to yield his turn, while at the same time implies to the ground staff that he thinks his computer needs to be checked. However, the ground staff does not immediately infer this and only picks up this hidden message when the passenger checks with her whether the computer is acceptable as hand luggage.

(6)

C: here you are Mister T__

E: thanks I have a computer in here **so**

C: oh

(pause)

E: you think it's okay

C: eh you want to check in this one

E: yea

(HKCSE, B049)

This turn-managing function is considered a "prompt" by Buysse (2012, p. 1769), who posits that this use of *so* fulfils "little or no function other than to indicate the desire to yield the floor". In the HKCSE, however, *so* is also found to hold or take the floor. Its functional scope is therefore broader in this corpus. Relatively speaking, native speakers of English again use this function more frequently (5.8%) when compared with Hong Kong Chinese (3.5%). The difference between the two speaker groups also reaches statistical significance ($x^2 \geq 17.884$; d.f. = 1; p = 0.00023). This suggests that the former group may be more actively engaged in using *so* to maintain, give up, or assume the speaker role in interactions.

Discussion

Findings from the study have revealed quite similar patterns of use of *so* between the Hong Kong Chinese speakers of English and native speakers of English in terms of the frequency rate and discourse function of *so* in the HKCSE. Overall, both speaker groups use the majority of the instances of *so* in the corpus as a discourse marker, rather than as an item with propositional meaning. No noticeable difference is found between them in the discourse rate of the word. Interestingly, although there is a similar particle in Cantonese which overlaps in function with *so*, Hong Kong Chinese do not avoid using *so* for fear of the Chinese-sounding association, unlike the German speakers of English in Müller's (2004) study.

Not only is the ratio of the discourse use to propositional use of *so* largely comparable in the corpus, the six discourse functions identified are all shared by the two speaker groups. In other words, no discourse function of *so* is unique in the speech of one group but not the other. In addition, fairly strong similarities have been noted in the relative pattern of frequency distribution of the six functions. For instance, framing is the most common function for both groups, whereas turn-managing is the least frequently-occurring. Indeed, the relative order of frequency distribution of the six functions is exactly the same for both groups. However, subtle differences have been observed in the relative frequency figure of these functions, with three out of six of the functional categories displaying statistically significant differences between the native and non-native speakers of English in the corpus. In particular, the textual function of framing is more significantly realized in the speech of Hong Kong Chinese, whereas the interpersonal responsive function and the interactional turn-managing function are more significantly realized by native speakers. The similarities and differences may be explained in terms of linguistic performance, pragmatic competence, and cultural preference.

In a rare study which discusses the self-corrective function of *so*, Buysse (2012) notes that this function is significantly more frequent in the speech of Belgian learners of English than in that of the native speakers, though he also remarks at the same time that the small number of tokens realizing this function in the data makes this finding inconclusive. Similar to Buysse's (2012) finding, the processing function is an infrequent category in the present study. However, there is no statistically significant difference between native and non-native speakers here in relation to the use of this function. One possible reason is that the non-native speakers in the HKCSE are experienced and competent adult users of English engaged in a

wide range of professional settings such as academic lectures, business service encounters and public speeches, whereas the Belgian learners are young adults still in training in their higher education. As such, the Hong Kong Chinese may be more confident in their linguistic performance in English and thus have fewer needs to make self-corrections. In particular, their linguistic competence, in terms of grammatical and textual competence (Bachman 1990), has reached a high level through their extensive on-the-job professional training and experience. The Hong Kong Chinese in the corpus thus are more likely to attend to the textual organization of communication, unlike the German learners of English as reported in Müller (2005), who use *so* less frequently on the textual level when compared with native speakers of English. As the use of *so* as a marker of transition is typical, frequently-occurring and common in many registers and genres (e.g. Johnson 2002, Rendle-Short 2003, Buysse 2012), there is a greater opportunity that the Hong Kong Chinese have been exposed to and are familiar with this function than the other two textual functions. In addition, the framing function is particularly dominant in the academic sub-corpus of the HKCSE in long monologic talks such as lectures and that more speakers of such monologues happen to be Hong Kong Chinese. These reasons may result in their more frequent use of *so* as a framing device.

Compared with linguistic competence, pragmatic competence is harder to master. Bardovi-Harlig (1996, 2001), for instance, has made the point that a satisfactory level of grammatical competence does not always correspond to the same level of pragmatic competence. It is therefore possible that comparatively speaking, the native speakers of English in the corpus are more sensitive to the interpersonal and interactional aspects of communication involving personal evaluation and role change. Further, regarding cultural preferences in relation to Hofstede's cultural dimensions (Hofstede *et al.* 2010), Hong Kong's scores on indulgence and individualism are both low (Hofstede Insights 2020). For the former dimension, it means that Hong Kong people tend to be restrained by social norms and are thus perhaps less vocal in their expression of self-opinion. For the latter dimension, it means that Hong Kong people tend to avoid open conflicts in a collectivist culture and are thus perhaps less likely to hold or take over the conversational floor in interaction. In the Hong Kong context, in particular, the native speakers of English are often in a more powerful position than the locals by assuming such roles as job interviewers or customers in professional settings. They may therefore play a more active and dominant role in communication by giving positive responses to encourage further interaction or by taking or keeping the floor more frequently to assert control. This may result in a significantly higher rate of use of *so* as a responsive signal and as a turn-managing device in the speech of the native speakers of English in the corpus, which in turn means a significantly lower rate of use of *so* achieving these two functions by Hong Kong Chinese comparatively.

Conclusion

This study has examined and compared the frequency rates of use and functions of *so* as a discourse marker by native and non-native speakers of Hong Kong. The findings have revealed both strong similarities and subtle differences between the two speaker groups. These findings have pedagogic implications for both Hong Kong and beyond. Locally in this multilingual and multicultural society, English is a lingua franca among different cultural and

ethnic communities. Comparative findings from the present study, especially taken together with those from previous studies, suggest that language background, both in relation to the level of communication competence and culture preference, influences how frequent and in what ways *so* is used as a discourse marker. Language teachers and literacy educators in the local context, therefore, need to take into account not only what first languages language learners speak but also their proficiency level in English and their cultural origin. In this respect, it should be noted that regional differences may exist between sub-cultural groups even within the same country, as in the case of Hong Kong where people may have varied cultural norms, assumptions and behaviours from other parts of China. Further, the order in which learners are exposed to different functions of *so* also merits careful consideration. The basic, prototypical textual functions of *so* as a boundary marker, for instance, may be more appropriately taught first to beginners based on its high frequency, prevalence and coverage, whereas functions in the interpersonal and interactional domains be introduced later to intermediate or advanced learners whose linguistic competence is relatively well-established but whose pragmatic competence still requires further development.

The same principles underlying the local pedagogic suggestions described above also apply to other contexts where English is a second/foreign language and to the teaching of other discourse markers as well as other multifunctional forms. The introduction of such multifunctional items therefore needs to be tailor-made based on the learners' sociolinguistic and cultural background as well as their academic or professional needs at the specific stage of their language development. At the same time, it should also be emphasised that a lower rate of an item by a particular community of language speakers does not necessarily mean underuse, nor does a different pattern of use always mean misuse. This is especially the case when a community of second/foreign language speakers is compared with a group of native speakers of English arbitrarily serving as the "standard" from a different and irrelevant geographical location. In this connection, a corpus-based, data-driven pedagogic approach drawing from naturally-occurring examples from the particular community where proficient language users actually employ these forms in a range of contexts is advocated. Such a localized approach offers the opportunity to consider the contextual variation concerning the use of these multifunctional forms in the very community where such forms are incorporated into the language or literacy curriculum. More contrastive analyses like the one presented here will further our understanding of the diversity displayed in the use of these items and allow them to be fruitfully included and successfully deployed as part of the pragmatic toolkit for second/foreign language speakers.

References

Bachman, L. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bardovi-Harlig, K. 1996. Pragmatics and language teaching: Bringing pragmatics and pedagogy together. In L. Bouton, ed. *Pragmatics and language learning*. Monograph Series Volume 7. Urbana-Champaign: University of Illinois, Division of English as an International Language, 21-40.

Bardovi-Harlig, K. 2001. Evaluating the empirical evidence of the need for instruction in pragmatics. In K. Rose and G. Kasper, eds. *Pragmatics in language teaching*. Cambridge: Cambridge University Press, 13-32.

Blakemore, D. 1988. *So* as a constraint on relevance. In R.M. Kempson, ed. *Mental representations. The interface between language and reality*. Cambridge: Cambridge University Press, 183-195.

Bolden, G.B. 2009. Implementing incipient actions: The discourse marker 'so' in English conversation. *Journal of Pragmatics*, 41 (5), 974-998.

Buysee, L. 2012. *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics*, 44, 1764-1782.

Cheng, W. 2007. The use of vague language across spoken genres in an intercultural corpus. In J. Cutting, ed. *Vague language explored*. London: Palgrave Macmillan, 161-181.

Cheng, W., Greaves, C. and Warren, M. 2005. The creation of prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic). *International Computer Archive of Modern English (ICAME) Journal*, 29, 5-26.

Cheng, W., Greaves, C. and Warren, M. 2008. *A corpus-driven study of discourse intonation: The Hong Kong Corpus of Spoken English (Prosodic)*. Amsterdam: John Benjamins.

Cheng, W. & Warren, M. 2006. // R you need to be RUTHless //: Entertaining cross-cultural differences. *Language & Intercultural Communication*, 6 (1), 35-56.

Davies, A. 2003. *The native speaker: myth and reality*. Clevedon: Multilingual Matters.

Grice, H.P. 1989. *Studies in the ways of words*. Cambridge, Mass: Harvard University Press.

He, A. 2002. On the discourse marker *so*. In P. Peters, P. Collins and A, Smith, eds. *New frontiers of corpus research*. Amsterdam: Rodopi, 41-52.

Hofstede Insights. 2020. *Hong Kong*. Available from https://www.hofstede-insights.com/country/hong-kong/ [Accessed 5 July 2021].

Hofstede, G., Hofstede, G.J. and Minkov, M. 2010. *Cultures and organizations: Software of the mind* (Rev. 3rd ed.). New York: McGraw-Hill.

Johnson, A. 2002. So...?: Pragmatic implications of *so*-prefaced questions in formal police interviews. In J. Cotterill, ed. *Language in the legal process*. Hampshire: Palgrave Macmillan, 91-110.

Müller, S. 2004. *'Well you know that type of person'*: Functions of *well* in the speech of American and German students. *Journal of Pragmatics*, 36, 1157-1182.

Müller, S. 2005. *Discourse markers in native and non-native English discourse*. Amsterdam: John Benjamins.

Prodromou, L. 2003. In search of the successful user of English. *Modern English Teacher*, 12 (2), 5-14.

Redeker, G. 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14, 367-381.

Redeker, G. 2006. Discourse markers as attentional cues at discourse transitions. In K. Fischer, ed. *Approaches to discourse particles*. Amsterdam: Elsevier, 339-358.

Rendle-Short, J. 2003. 'So what does this show us?': Analysis of the discourse marker 'so' in seminar talk. *Australian Review of Applied Linguistics*, 26 (2), 46-62.

Schiffrin, D. 1987. *Discourse markers*. Cambridge: Cambridge University Press.

Sperber, D. & Wilson, D. 1995. *Relevance: Communication and cognition*. Oxford: Blackwell.

Stenström, A.-B. 1994. *An introduction to spoken interaction*. New York: Longman.

Yin, Z. 2019. *So*? The effect of register and social settings on the meanings of *so* and its pedagogy. *The TESOLANZ Journal*, 27, 32-43.

---

[1] The 138 instances of D-use *so* which are functionally unclassified are excluded from Table 11.3.