Institut für Sprache und Information
Heinrich-Heine-Universität Düsseldorf
Strukturkurs Chinesisch
Univ.-Prof. Dr. Kilu von Prince
4. Semester

A quantitative analysis on Singlish's Sentence Final Particles

Alexander Clemen
2832399
alexander.clemen@uni-duesseldorf.de
5. Fachsemester
February 11, 2022

Contents

1 Introduction

Singlish[1]'s Sentence Final Particles (henceforth SFP) has been a fascinating field for linguists for over 50 years. Not just because Singapore has a unique history and geographical location, making it a melting pot for cultures and languages, but also because gathering data is relatively easy. One problem with seasoned researchers heading into the field is, they obtained knowledge that may taint their view of foreign language phenomena and as a consequence e.ġ. particles are more likely to be determined a misleading origin. Gupta (1992) provides two examples: Baskaran (1987: 110), who attributes the most common SFP *ah* to *Tamil*, while Gupta (1992) herself determines Cantonese as *ah*'s language of origin. Richards & Tay (1977) and Kwan-Terry (1978) have differing opinions over which Chinese language (Hokkien or Cantonese) is more influential regarding pragmatic particle replication in Singlish. In this term paper, instead, I will ascribe particle preferences solely on the speaker's language proficiency by utilizing corpus data. With this in mind, I will try to investigate the following research question:

Are there Sentence Final Particle (SFP) preferences dependent on the speaker group's L1 in Singlish?

Gupta (1992) was the first to raise a similar question on a passing comment in her groundbreaking paper "The pragmatic particles of Singapore Colloquial English". 20 years later, Smakan & Wagenaar (2013) tried to answer the question quantitatively and indeed found differences in particle use – not SFP use – dependent on "ethnic awareness". The researchers found a significant difference in frequency for the particles *ah* and *lah* compared to other observed particles, but no overall significant particle preference dependent on "ethnic clues". For individual particles they found a minor preference for the most frequent particle *ah* for Malay speakers and few occurrences of Chinese speakers using the aforementioned *ah* and *lah*. In general, Chinese speakers seemed to prefer particles with, in the authors' words, clear Chinese origin. Smakan & Wagenaar (2013: 315) limited their corpus study to eight particles (*lah*, *ah*, *what*, *lor*, *hor*, *leh*, *meh*, and *ma*) taken from Lim (2007). Other researchers have covered the field of meaning and use regarding some of Singlish's particles *lah* (Besemeres & Wierzbicka, 2003), *sia* and *sial* (Khoo, 2012), *what* (Kuteva *et al.*, 2018), *one* (Zhiming, 2009), *can* (Hiramoto, 2012), *know* (Wee, 2003), and *already* (Ziegeler & Lee, 2020; Zhiming, 1995). The aforementioned Lim (2007) and Gupta (1992) classified Singapore's particles "exhaustively". No one seems to take a naive corpus linguistic approach which could yield rare or unknown ethnic related particles with the consequence that L1-SFP preferences become more visible.

In my term paper I expect to find more SFPs than Gupta (1992: 33), who worked on recordings from Chinese speakers, and Smakan & Wagenaar (2013), who restricted themselves to the predefined set of eight particles. Second, I expect to show clear L1-SFP preference because (i) there is no foreigner involved in gathering the data, (ii) young speakers are recorded and youth language particles may be present, and (iii) there are only

[1] Singlish is also called Singapore Colloquial English (Gupta, 1992), Colloquial Singapore English (Smakan & Wagenaar, 2013), or Singapore English (Hiramoto, 2012) but I prefer Singlish, because those who speak it call it that way.

inter-group pairings, which promotes L1 related particle use (Smakan & Wagenaar, 2013). Lastly, due to the inter-group pairings I expect the results to contain code switching.

The main motivation for the present topic stems from my personal connection to the country and the strange but fascinating observation "Singaporeans end many of their sentences with the SFP *lah*." While answering the question "Where does *lah* come from?" is reserved for more large scale projects, Gupta (1992) posed interest in discovering the source of Singapores' SFPs and 30 years later there seems to be no naive quantitative account of particle classification even though data science concerning corpus linguistics has transformed tremendously.

I will begin with a brief overview of Singapore's location and history, determining the origin of Singlish's unique SFP-fingerprint and its unclear status as a creole language. In the following Section 2.2 I will explain the parameters of this corpus study followed by my findings, Section 2.3, which are discussed in Section 2.4. Finally, the last Section (3) presents the conclusion and ideas for further research.

2 Body

2.1 Preliminaries

Location, History, and Grammar Transmission Singapore is an sovereign city-state in the center of Southeast Asia. Its location marks the bottleneck of shipping routes between India and America, India and China, as well as China and Europe. This situation blesses Singapore with considerable wealth and diversity today but, metaphorically speaking, also made it the sirloin for colonizers. After its first discovery by the Malay at around 650 AD, Singapore underwent waves of colonization by the Portuguese (16^{th} century), the Dutch (17^{th} century), and the British (19^{th} century) Abshire (2011). The prospering British settlement attracted workers from Southern China (Cantonese, Mandarin, Hokkien, Hakka, and Teochew speakers (henceforth Chinese speakers)), Malaysia (Malay speakers) and India (Tamil speakers). Within 50 years, Singapore's population jumped from around 1,000 in 1819 (Turnbull, 1989: 5) to around 100,000 in 1871 (Yeoh, 2003: 317). This unnatural population increase confined on a small space could be described as a melting pot from which the contact language Singlish emerged. At first glance, one could describe Singlish a creole language – a contact language with native speakers (Sebba, 1997: 15) – but unlike individual conducting forced labour in other British colonies guest workers in Singapore had social status (Ziegeler, 2015). The "rules" of language contact situations apply nonetheless. Blasi *et al.* (2017: 727) argue that "the majority of creole grammars have been transmitted [...] either from their lexifiers or substrates, or through later contact". If we were to translate this finding to Singlish, then SFPs in 19^{th} century English or Chinese, Malay, and Tamil have been transmitted to early Singlish. Contemporary literature agrees that Singlish is rich in SFPs. Still the question remains: are all of these contact situation SFPs shared among all Singaporeans or are there preferences dependent on ethnic clues.

Working Definition My definition of Sentence Final Particle is determined by the technical limitations of the *National Speech Corpus* (NSC). In the corpus particles, and only particles, are put into square brackets and there is no punctuation as a means to determine the end of the sentence. As such "sentence final" is equated to "utterance final" since every utterance final word will also be sentence final. Manual analysis is needed to check for code switching preceding to the particle, which exceeds the scope of this term paper and as such is neglected. However, a paragraph is designated to the code switching limitation. As such:

> A Sentence Final Particles (SFP) is every word in square brackets at the end of an utterance.

Additionally, for ease of reading from here on, the term *particle* will be synonymous with the term *sentence final particle* or *SFP*.

2.2 Method

Corpus data from the *National Speech Corpus* (NSC) was extracted and processed with *Sublime Text*, *Excel*, *R* and *Inkskape*. The NSC is governed by the *Info-communications and Media Development Authority* (IMDA) of Singapore to solve automatic speech recognition inaccuracies, since English spoken in Singapore deviate from training data stemming from e. g. the UK or the USA. The corpus consists of 1.2 TB annotated open source speech data spread across six parts ((i) phonetically-balanced scripts, (ii) prompted recordings, (iii) conversational data, and more.[2]), each encompassing 1,000 hours of material (Infocomm Media Development Authority, 2022b). Part three was chosen since it seemed most promising not to reflect artificial circumstances. The corpus includes "conversations [between two individuals] covering daily life and of speakers playing games provided" (Infocomm Media Development Authority, 2022a). The following meta-data is present in the corpus, from which each four Chinese, Malay, and Tamil speaker pairs were selected. The criteria are ranked in order of importance and the selections were intended to maximise SFP occurrence.

- first language: only Malay/Chinese/Tamil pairs to maximise L1 bias,
- language proficiency: only Malay/Chinese/Tamil and English pairs to maximise L1 bias,
- ethnicity: only pairs of speakers with the same ethnicity to maximize the ethnicity bias (see (Smakan & Wagenaar, 2013: 319),
- age: 1X or 2X (10 to 29 years old) since young speakers tend to speak more colloquially,
- relationship [between interlocutors]: friend since friends tend to speak more colloquially,
- occupation: mainly student since students tend to speak more colloquially,
- education: was kept as similar as possible across all individuals,
- gender: balanced

24 individuals (3 first languages $\times$ 2 interlocutors $\times$ 4 pairs) were selected from the corpus that fit the criteria best. A data set of 63,224 lines of speech with a total of 2,763 SFPs was compiled. 27 distinct SFPs were gathered using regular expressions in *Sublime Text* and word processing in *Excel*. Since solely particles were annotated in square brackets, I could retrieve every utterance ending in a closed square bracket. *Excel* formulas retrieved the SFPs from the preceding utterance into a neat data frame. An *R* script was utilized to calculate counts and fractions, and to display the data in graphs. *Inkscape* was used to visualize the data in a Venn Diagram.

2.3 Findings

Corpus analysis Table 1 provides a numeric display of total counts and subsets of L1 specific counts and proportionate use of particle. Figure 1 shows the stacked sums of SFP counts for the three L1s Chinese (red), Malay (green) and Tamil (blue). The Venn Diagram, Figure 2, highlights the overlap of particle usage – meaning whether a particle is shared by all Singlish speakers (the center), shared with a second speaker group (the outer overlapping sections), or whether the particle is language background exclusive (the outer sections).

[2] In the course of writing the term paper I lost access to the corpus and could not regain it. As such the descriptions of parts (iv) to (vi) is not presented here.

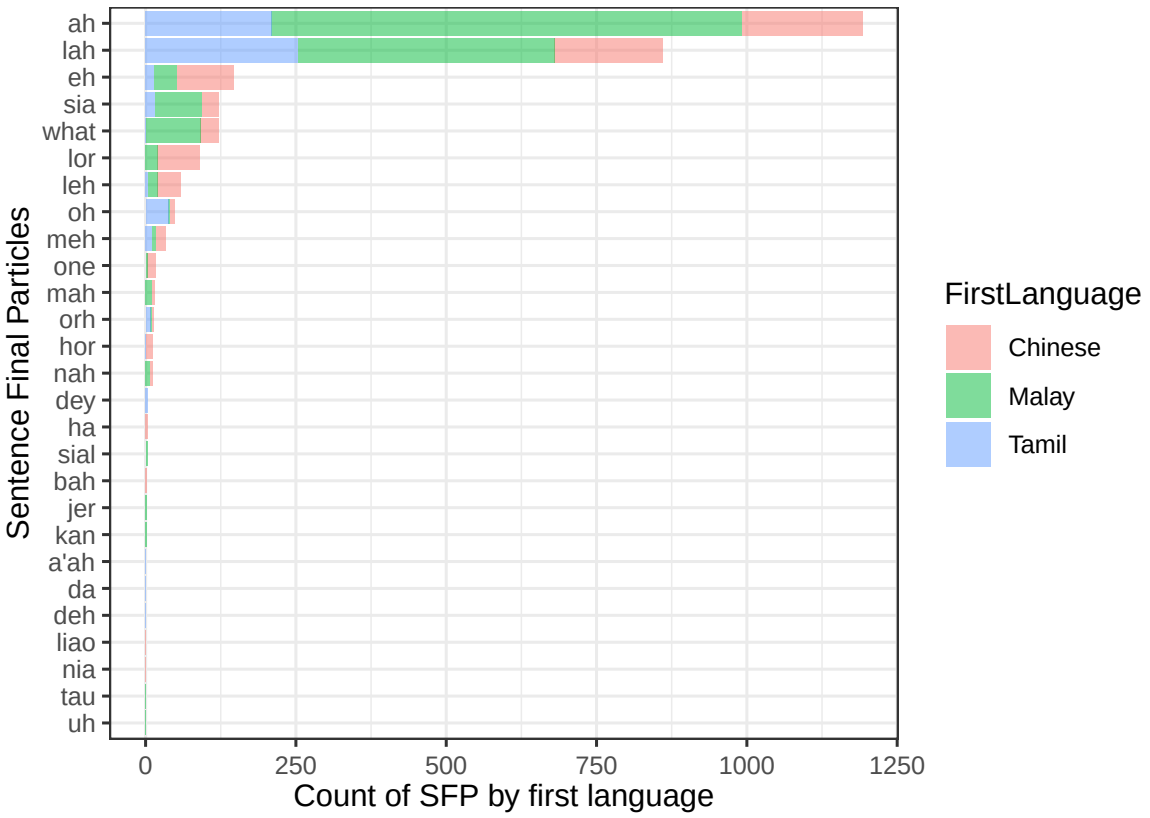


Figure 1: Stacked counts of Singlish's 27 SFPs for the speaker groups Chinese, Malay, and Tamil

The most prominent finding is the amount of SFPs (27) gathered from the corpus analysis. Additionally, the overwhelming frequency of the particles *ah* (1,192 total) and *lah* (861 total), vastly overshadowing the remaining particle landscape of Singlish (see Figure 1). The two particles make up 74% of all particle use across all particles observed. Notably, Malay speakers use *ah* (66% Malay) and *lah* (49% Malay) substantially more than the other two groups (see Table 1). The only other particles that reach a three digit count are *eh* (146 total), *sia* (122 total), and *what* (121 total). Moving on, Table 1 illustrates fractions and L1 preference in more detail. With *sia* (64% Malay) and *what* (74% Malay), covering more than half of particle use, Malay speakers predominately use them compared to Chinese and Tamil speakers. Indeed *what* was only uttered once by a Tamil speaker. *Eh* (64% Chinese), on the other hand, alongside with *leh* (67% Chinese) and *meh* (48% Chinese) are predominately used by speakers with Chinese background. Tamil speakers preferred particles are *oh* (79% Tamil) and *orh* (62% Tamil). Notably, for these two particles, the Malay share, at 4% (2 times) and 15% (2 times), is exceptionally small. Chinese speakers use *orh* comparatively infrequently (3 times), too. For SFPs that are used by all speaker groups (see Figure 2, center) the shared particle was found 39% more frequently in the dominant speaker group than the second most frequent speaker group.

SFP	Total	Chinese	Malay	Tamil
ah	1,192	200 (17%)	782 **(66%)**	210 (18%)
lah	861	181 (21%)	426 **(49%)**	254 (30%)
eh	146	94 **(64%)**	38 (26%)	14 (10%)
sia	122	28 (23%)	78 **(64%)**	16 (13%)
what	121	30 (25%)	90 **(74%)**	1 (1%)
lor	90	70 **(78%)**	20 (22%)	
leh	58	39 **(67%)**	15 (26%)	4 (7%)
oh	48	8 (17%)	2 (4%)	38 **(79%)**
meh	33	16 **(48%)**	6 (18%)	11 (33%)
one	17	14 **(82%)**	3 (18%)	
mah	15	4 (27%)	11 **(73%)**	
orh	13	3 (23%)	2 (15%)	8 **(62%)**
hor	12	11 **(92%)**		1 (8%)
nah	11	4 (36%)	7 **(64%)**	
dey	4			4 **(100%)**
ha	4	4 **(100%)**		
sial	3		3 **(100%)**	
bah	2	2 **(100%)**		
jer	2		2 **(100%)**	
kan	2		2 **(100%)**	
a'ah	1			1 **(100%)**
da	1			1 **(100%)**
deh	1			1 **(100%)**
liao	1	1 **(100%)**		
nia	1	1 **(100%)**		
tau	1		1 **(100%)**	
uh	1		1 **(100%)**	

Table 1: Counts (Fractions) of ovserved SFPs in Singlish

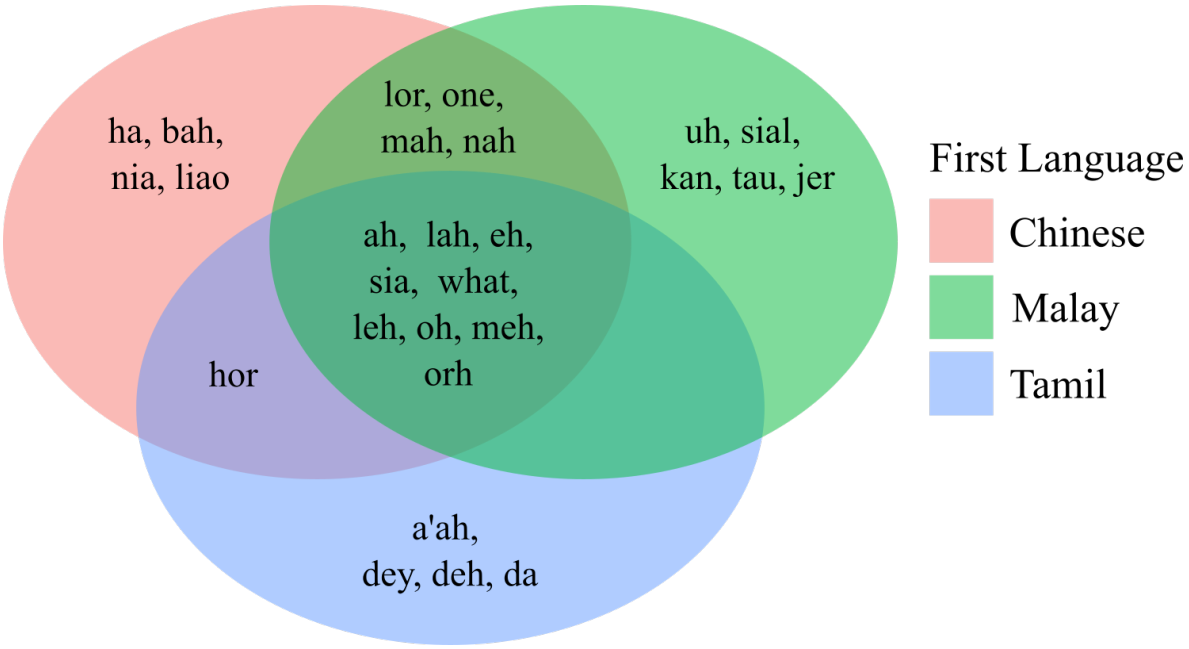


Figure 2: Singapores 27 SFPs and their users Chinese (red, left), Malay (green, right), Tamil (blue, bottom) speakers

Particles used by two speaker groups are for Chinese and Malay speakers *one*, *lor*, *mah*, and *nah*. The frequent *lor* (90 total, of which 78% Chinese) and *one* (17 total, of which 82% Chinese) are used significantly more often by the Chinese speaker group. While the infrequent *mah* (15 total, of which 74% Malay) and *nah* (11 total, of which 64% Malay) are substantially used more by Malay speakers. The only SFP that is shared among Chinese and Tamil speakers is *hor* at an overwhelming quota of 92% to 8%. Indeed, the 8% is made up by a single use of *hor*. Mirroring the analysis above, for particles that are shared by two speaker groups the dominant speaker group uses the given particle 55% more than the other speaker group.

There are four exclusive SFPs for the Tamil speaker group: *dey* (4), *a'ah* (1), *da* (1), and *deh* (1) (see Figure 2, blue section). For Chinese speakers there are also four language exclusive SFPs (red section): *ha* (4), *bah* (2), *nia* (1), and *liao* (1); while Malay speakers uttered five L1 specific particles (green section): *sial* (3), *jer* (2), *kan* (2), *uh* (1), and *tau* (1). Concluding, speaker group exclusive SFPs do exist. But at a total of 24 particles or 0.87% of the total, they make up a fairly small share of Singaporeans SFP use.

Singlish or Code Switching One downside to the ethnic pairing filter criteria mentioned in Section 2.2 (only e. g. Tamil-Tamil speaker pairs, no e. g. Tamil-Malay speaker pairs) is, interlocutors may be prone to code switching – a multilingual speaker may blend two separate languages in conversation or writing (Sebba, 1997). As such, the critical reader may claim the 24 L1 specific SFPs stem from non-Slinglish utterances. This section will resolve this issue concerning L1 exclusive SFPs. The 24 SFPs were manually checked and the results are displayed in Table 2. All L1 exclusive SFPs (*ha*, *bah*, *liao*, and *nia*) uttered by L1 Chinese speakers were without exception embedded in a Singlish utterance (see Examples (1)).

(1) a. ya lah her sometimes her attitude lah when she tired **bah**.
b. no you yours first **bah**.

Tamil speakers paint a very different picture. *Dey* is embedded in a fully code switched utterance (Example (2-a) as well as a fully functional Singling utterance (Example (2-b)). The corpus provided only one "Tamil-*dey*" and three "Singlish-*dey*s". In regards of semantics, it is unclear to me whether *dey* means the same in the two contexts. The other three L1 exclusive particles (*a'ah*, *deh*, and *da*) are found only in Singlish contexts

L1: Chinese	Singlish	CS
ha (4)	4	-
bah (2)	2	-
liao (1)	1	-
nia (1)	1	-
L1: Tamil	**Singlish**	**CS**
dey (4)	3	1
a'ah (1)	1	-
deh (1)	1	-
da (1)	1	-
L1: Malay	**Singlish**	**CS**
sial (3)	3	-
jer (2)	-	2
kan (2)	2	
tau (1)	-	1
uh (1)	1	-

Table 2: SFPs (Counts) appearing in Singlish or Code Switching (CS) contexts

(see Table 2).

(2) a. eh siikkiram peesa solludaa enakku semmaiyaa pasikkuthudaa saapitduruveen **dey**.
"**Tell him** that he should speak faster, I am starving, I will devour him/something.[3]"
b. what is the nearest petrol station **dey**.

The most inconclusive finding is presented with Malay speakers. *Sial* and *uh* are observed to be only used in Singlish sentences, while *jer* and *tau* is only used in Malay sentences. *Kan* seems to only surface in highly mixed contexts, as illustrated in the two Examples (3-a) and (3-b) ("Malay" words are written in italics). The Malay words *takde* and *tu* are neither found in a conventional dictionary Shellabear (1912), nor any open source translator online. This indicates that the speakers do not code switch to standard Malay, but use a more niche form of Singlish. It is unclear whether one should categorize *kan* as part of the Singlish or quasi-Malay vocabulary. In both example sentences at least half of the sentence is Singlish, but is half "good enough"? The more conservative approach is to place *kan* in a grey area.

(3) a. *takde* thing **kan**.
b. *yang* (uh) got the picture of the (uh) water *tu* **kan**.

2.4 Discussion

Comparison with related literature This study confirms Gupta's and Smakan & Wagenaar's finding: *ah* and *lah* are the most common SFPs in Singlish. Adding to Gupta (1992), who only observed Chinese speakers, Tamil and especially Malay speakers use *ah* and *lah* frequently. Smakan & Wagenaar (2013), who observed particles in sentence initial, medial, and final position, state for *ah* and *lah* a total particle use share of 94%, while I observed a share of 74% for particles in sentence final position. From this we can infer (a) the two particles are even more present in sentence initial and medial position, (b) the remaining particles are less frequently used in sentence final position, or, most likely, (c) a mix of the above. Smakan & Wagenaar (2013) state the origin of *ah* to be unknown but my findings show prevalent preference of Malay speakers, which may hint towards *ah* be of Malay origin. *Lah* is relatively evenly shared among the speaker groups. This is not surprising because *lah* functions as a marker of solidarity (Ziegeler, 2015: 37) or code-marking (Pakir, 1992: 149), illustrating that one speaks Singlish or belongs to the group of Singlish speakers/Singaporeans.

[3] Translated by my good friend Varsha Irver.

Furthermore, Smakan & Wagenaar (2013: 318) reiterate the origin of *lor*, *hor*, *leh*, and *meh* to be Chinese and *what* to be "unknown". My frequency analysis deviates from their analysis in some aspects (compare Table 1 and Figure 3 if needed). In general, we paint a similar picture, but my analysis shows greater use of particles among multiple speaker groups. Smakan & Wagenaar show *hor* and *meh* to be Chinese exclusive, while my data shows usage of both Chinese and Tamil speakers; Smakan & Wagenaar shows *leh* to be Chinese exclusive, while my data suggests *leh* to be used by all speaker groups; and lastly Smakan & Wagenaar has no findings for *mah*, while I found data for both Chinese and Malay speakers with a preference for Malay speakers. We need to bear in mind Smakan & Wagenaar accepted particles in sentence initial, medial, and final position, while I only accepted SFPs. This could be attributed to the difference in corpora or over the last 10 years speakers tend to use more "foreign" particles (here "Chinese particles"). In my opinion, 10 years of language contact is not enough to cause such variance in particle preference and as such the corpus made the difference. Other frequent SFPs (*eh* and *sia*) were overlooked by the researcher, which signals that taking a naive approach might surface greater results. So far, this study has revealed a far greater account of SFPs (27) compared to other researchers (Gupta (1992): 11, Ler Soon Lay (2006): 10, Gupta (2006): 8) because I did not set any formal barriers for excluding potential particles, which may be described as imprudent, but I deem the great number of particles, of which some did not find any recognition in the literature before, a success.

"Shared" SFPs There seems to be a cleft between true shared SFPs, which I will continue calling "shared SFPs" and SFPs that are clearly preferred by one speaker group. The latter group I suggest calling "fractured SFPs". *Lah* (Chinese 21%, Malay 49%, Tamil 30%) and *meh* (Chinese 48%, Malay 18%, Tamil 33%) are the only particles that are shared by all Singaporeans roughly equally. For *lah* the meaning and the reason why it is relatively evenly shared among Singaporeans has already been discussed above. For *meh* I see no reason why it is used evenly among speaker groups. Gupta (1994: 44) believes *meh* expresses surprise or questions a presupposition. This I do not see as a unifying factor for Singapore-wide particle use. Further research is needed to answer the question why *meh* is used by all Singaporeans. For the fractured SFPs (*ah*, *eh*, *sia*, *what*, *leh*, *oh*, and *orh*), the biggest speaker group uses the particle $\geq$38% more than the second biggest speaker group. Taking *eh* as an example: Chinese 64%, Malay 26%, Tamil 10%; $64\% - 26\% = 38\%$. Researchers are not certain about the true origin of some particles (e. g. (Kwan-Terry, 1978: 22) and (Kuteva *et al.*, 2018: 33) on *what*) and therefore I will not attempt to make a declarative statement about origins, but the statistics suggest a correlation of language background and particle preference.

Tamil speaker group The data suggests that Tamil speakers speak a relatively segregated Singlish from the two other groups. Malay and Tamil speakers share no SFP and Chinese and Tamil speakers share only one SFP, *hor*, which was merely used once by a Tamil speaker. For the fragmented particles, Tamil speakers prefer the particles *oh* and *orh* (average Tamil share 70%) and strongly reject the particles *ah*, *eh*, *sia*, *what*, and *leh* (average Tamil share 9%). The low Tamil share for non-preferred particles show that the adoption of "foreign" FSPs seems undesired. Additionally, one could speculate that *oh* and *orh* are related because one can infer from the Latin script, a highly phonetic script, that they are pronounced similarly. Both are also vastly dominated by Tamil speakers, so that one could infer that *oh* and *orh* originated from one or two Tamil lexemes. One of the four exclusive particles (*dey*, *a'ah*, *deh*, *da*) is used in a Tamil sentence (see (2-a)), which suggests that *dey* stems from Tamil. I speculate the phonological similar particles *deh* and *da* to also come from Tamil, but I have no data to back this thought. This language segregation can be backed up with historical evidence. Throughout the development of Singlish, Tamil speakers were quantitatively speaking of minor contribution. In 1924, the share of Tamil speakers in Singapore was only at 7% (Buckley, 1984: 154)) and in 2020 at 6% (Ziegeler & Lee, 2020: 9). This minority status entails that Tamil's influence on Singlish's origin up to today's use has been relatively small.

Malay speaker group Out of the top five most frequent SFPs in Singlish, Malay speakers use four of them the most – the much discussed *ah* and *lah* and the two fractured particles *sia* and *what* (see Table 1). One can safely say Malay speakers use on average more SFPs than Tamil and Chinese speakers. Of the four particles shared by Chinese and Malay speakers, two are preferred by Malay speakers (*mah* and *nah*, average Malay

share 69%) and two are preferred by Chinese speakers (*lor* and *one*, average Chinese share 80%). One could infer that half stems from Malay and half stems from a Chinese language, but it is well documented (Lim, 2007: 469) (Low & Brown, 2005: 179) that *mah* stems from Cantonese *ma*. *Nah* has been mentioned but not ethnically classified (Lim, 2007). Malay's preference for "Chinese particles" is also attested when we look at the fractured particles preferred by Chinese speakers (*eh* and *leh*) of which Malay's share is at 26% on average versus particles preferred by Tamil speakers (*oh* and *orh*) of which Malay's share is at 9% on average. Malay speakers tend to use "Chinese particles" more than "Tamil particles" and in the instances of *mah* even adopt it "as their own". Turning to exclusive particles, there is a high degree of code switching in as such one particle, *kan*, only appears in a Malay type of language that is not translatable by any standard dictionary. To underpin this implicated idea of *slang*, *sial* (engl: *fuck*, as a sentence strengthener) is considered incredibly rude among Singaporeans and is often used by young Malays (Yuan, 2021). Singaporeans with Malay background seem to have a higher degree of colloquial Singlish. Malay's Singlish being at a higher degree of colloquialism is expressed in the higher frequency of particles, the usage of (more) code switching, and the usage of vulgar particles that are otherwise avoided by other speaker groups.

Chinese speaker group Singaporeans with a Chinese L1 seem so speak the most consistent Singlish, meaning there is no evidence of code switching. The two fractured particles they prefer are *eh* and *leh*, of which *leh* is thought to have originated from a Chinese language (Low & Brown, 2005). The origin of *eh* is unclear, however the data also suggests a Chinese origin. With regard to fractured particles that are preferred by other ethnicities (*ah*, *sia*, *what*, *oh*, and *orh*, average Chinese share 22%), Chinese speakers seem to take a confident stance on using them, but there is no preference for ethnicity: Malay preferred particles (*ah*, *sia*, and *what*) with an average Chinese share at 23% and Tamil preferred particles (*oh*, and *orh*) with an average Chinese share at 20%. *Lor* and *one*, the two particles shared with Malay speakers are at 80% average use of Chinese speakers, indicating a clear preference for Chinese-Singaporeans. Wee (2002) draws paralels between the Cantonese particle *lo* and Singslish's *lor* and my data supports this line of argument. Zhiming (2009) identifies the Cantonese *de* forming the grammar, while the English lexeme *one* serves as vocabulary for Singlish's *one*. Once again, the data supports this claim. The Tamil-Chinese shared SFP *hor* is clearly preferred by Chinese speakers. The L1 exclusive paricles *ha*, *bah*, *nia*, and *liao*, when looking at the data, are obviously preferred by Chinese speakers.

3 Conclusion

To conclude, there are clearly SFP preferences dependent on the speaker groups L1 in Singlish. The naive corpus linguistic approach has laid bare 27 different Singlish Sentence Final Particles of which 25 have clear ethnic related preference and two particles (*lah* and *mah*) are shared among Singaporeans roughly equally. Two particles (*ah* and *lah*) make up nearly three quarters of all SFP use, while less than 1% is reserved for L1 exclusive particles. *Ha*, *bah*, *nia*, and *liao* are only used by Singaporeans with Chinese background; *uh*, *sial*, *kan*,*tau*, and *jer* are only used by Singaporeans with Malay background; and *a'ah*, *dey*, *deh*, and *da* are only used by Singaporeans with Tamil background. Tamil speakers speak a relatively segregated Singlish in terms of SFPs, since the overlap between their and other SFPs is by far the smallest. Malay speakers speak the most colloquial Singlish and tend to use Chinese preferred SFPs over Tamil preferred SFPs. Chinese speakers speak the most consistent Singlish and are most confident with using SFPs that are preferred by any other speaker groups. This term paper has complemented previous works even if the naive approach has its downsides.

(4) a. indeed I say clockwise **ah** why you going anticlockwise
b. **oh eh**

There are a total of 14,461 annotated particles and my "utterance final" limitation leaves me at only 2,763 (19%) missing four out of five particles, of which many are sentence final as example (4-a) shows. Another downside is the case of what I call "double SFPs" as shown in example (4-b). Here, only *eh* but not *oh* is extracted. A more sophisticated algorithm might solve this issue in the future. Lastly the issue of code switching was only resolved for the exclusive particles but not the grand total of 2,763. They would all have to be manually

checked, but this exceeds the extend of this study. Furthermore, the selection of my 24 individuals is heavily biased towards SFP occurrence and is potentially not representative for all Singlish speakers.

All in all, I did find more SFPs than previous authors, in most instances a clear L1-SFP preference was shown, and code switching was apparent in my results, as expected. It would be interesting to see if these results are replaceable if the above mentioned downsides are resolved or a greater data set or corpus is used. I remain curious what the next 50 years of Singlish research can show us.

Word count: 4,144 words

References

Abshire, Jean E. 2011. *The history of Singapore: Jean E. Abshire.* The Greenwood histories of the modern nations. Santa Barbara, Calif: Greenwood.

Baskaran, Lohanayahi. 1987. *Aspects of Malaysian English syntax.* Ph.D., University of London, London.

Besemeres, Mary, & Wierzbicka, Anna. 2003. The meaning of the particle lah in Singapore English. *Pragmatics & Cognition*, **11**(1), 3–38.

Blasi, Damián E., Michaelis, Susanne Maria, & Haspelmath, Martin. 2017. Grammars are robustly transmitted even during the emergence of creole languages. *Nature human behaviour*, **1**(10), 723–729.

Buckley, Charles B. 1984. *The anecdotal history of old times in Singapore: From the foundation of the settlement under the honourable the East India company on Feb. 6th, 1819 to the transfer to the colonial office as part of the colonial possessions of the Crown on April 1st, 1867.* Singapore: Oxford University Press.

Gupta, Anthea Fraser. 1992. The pragmatic particles of Singapore colloquial English. *Journal of Pragmatics*, **18**(1), 31–57.

Gupta, Anthea Fraser. 1994. *The step-tongue: Children's English in Singapore.* Multilingual matters, vol. 101. Clevedon England and Philadelphia: Multilingual Matters.

Gupta, Anthea Fraser. 2006. Epistemic modalities and the discourse particles of Singapore. *Pages 243–263 of:* Fischer, Kerstin (ed), *Approaches to Discourse Particles.* Amsterdam: Brill.

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

Infocomm Media Development Authority. 2022a. *About.txt.*

Infocomm Media Development Authority. 2022b. *National Speech Corpus - Infocomm Media Development Authority.*

Khoo, Velda Yuan Ling. 2012. *The sia particle in colloquial Singapore English.*

Kuteva, Tania, Rhee, Seongha, Ziegeler, Debra, & Sabban, Jessica. 2018. On sentence-final "what" in Singlish: Are you the Queen of England, or what? *Journal of Language Contact*, **11**(1), 32–70.

Kwan-Terry, Anna. 1978. The Meaning and the Source of the "la" and the "what" Particles in Singapore English. *RELC Journal*, **9**(2), 22–36.

Ler Soon Lay, Vivien. 2006. A relevance-theoretic approach to discourse particles in Singapore English. *Pages 149–166 of:* Fischer, Kerstin (ed), *Approaches to Discourse Particles.* Amsterdam: Brill.

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

Low, Ee Ling, & Brown, Adam. 2005. *English in Singapore: An Introduction.* Singapore: McGraw-Hill Education (Asia).

Pakir, Anne. 1992. Dictionary entries for discourse particles. *Pages 143–152 of:* Pakir, Anne (ed), *Words in a cultural context.* Singapore: Unipress.

Richards, Jack C., & Tay, Mary W.J. 1977. *The la particle in Singapore English: The English language in Singapore.* Singapore: Eastern Universities Press.

Sebba, Mark. 1997. *Contact Languages: Pidgins and Creoles.* Macmillan Modern Linguistics. London: Macmillan Education UK.

Shellabear, William Girdlestone. 1912. *Malay-English Vocabulary: Containing Over 7000 Malay Words or Phrases With Their English Equivalents, Together With an Appendix of Household, Nautical and Medical Terms*. 2 edn. Singapore: Cornell University Library.

Smakan, Dick, & Wagenaar, Strephanie. 2013. Discourse particles in Colloquial Singapore English. *World Englishes*, **32**(3), 308–324.

Turnbull, C. M. 1989. *A history of Malaysia, Singapore and Brunei*. Sydney: Allen & Unwin.

Wee, Lionel. 2002. Lor in colloquial Singapore English. *Journal of Pragmatics*, **34**(6), 711–725.

Wee, Lionel. 2003. The birth of a particle: know in Colloquial Singapore English. *World Englishes*, **22**(1), 5–13.

Yeoh, Brenda S. A. 2003. *Contesting space in Colonial Singapore: Power relations and the urban environment*. Singapore studies in society and history. Singapore: Singapore University Press.

Yuan, Kae. 2021. *Discussing Singspores SFPs: In Person*.

Zhiming, Bao. 1995. Already in Singapore English. *World Englishes*, **14**(2), 181–188.

Zhiming, Bao. 2009. One in Singapore English. *Studies in Language*, **33**(2), 338–365.

Ziegeler, Debra. 2015. *Converging grammars: Constructions in Singapore English*. Language contact and bilingualism, vol. 11. Boston: De Gruyter Mouton.

Ziegeler, Debra, & Lee, Sarah. 2020. Lexical Retention in Contact Grammaticalisation: Already In Southeast Asian Englishes. *Glot International*, **12**(3), 737–783.

A Appendix

Table 5. Distribution of particles (N = 1,427); number of users (at least one occurrence) of each particle per ethnicity

Particle	Origin	Chinese (19 speakers)		Indian (34 speakers)		Malay (23 speakers)	
		N	%	N	%	N	%
ah	unknown	17	89	29	85	17	74
lah		16	84	28	82	16	70
what		4	21	12	35	1	4
lor	Chinese	10	53	5	15	2	9
hor		4	21	-	-	-	-
leh		4	2	-	-	-	-
meh		1	5	-	-	1	4
ma		-	-	-	-	-	-

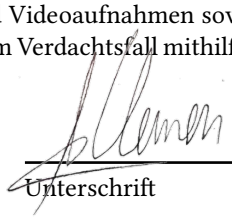
Figure 3: Smakan & Wagenaar's distribution of eight Singlish particles

Eigenständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen sind, habe ich in jedem einzelnen Fall unter Angabe der Quelle kenntlich gemacht. Dies gilt auch für verwendete Zeichnungen, Skizzen, Ton- und Videoaufnahmen sowie graphische Darstellungen. Ich erkläre mich damit einverstanden, dass meine Arbeit im Verdachtsfall mithilfe einer Plagiatssoftware überprüft wird.

Düsseldorf, 11.02.2022

Ort, Datum



Unterschrift