



Gendered Use of Sentence-Final Particles in Taiwan Mandarin: A Corpus Study

Citation

Baumel, Jonathan P. 2020. Gendered Use of Sentence-Final Particles in Taiwan Mandarin: A Corpus Study. Bachelor's thesis, Harvard College.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364706>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

GENDERED USE OF SENTENCE-FINAL PARTICLES IN TAIWAN MANDARIN:
A CORPUS INQUIRY

by

Jonathan Baumel

presented to the Department of Computer Science

and the Department of Linguistics

in partial fulfillment of the requirements

for the degree with honors

of Bachelor of Arts

Harvard College

March 2020

Acknowledgements

I would like to thank my adviser, Prof. C.-T. James Huang, for the invaluable guidance he provided me in my research. I always emerged from our meetings with a deepened understanding of the material and a renewed motivation for the project. Among many other contributions, Prof. Huang had the idea of investigating English intonation, clarified many nuances of SFP pragmatics, and suggested several helpful references. I was grateful that he could bring Taiwanese Mandarin to life for me in Cambridge. I am also indebted to my other Taiwanese informant, Li Jen Tsai, who was exceptionally patient with my repeated requests for her insight. I frequently relied on her thorough and illuminating judgments in my analysis. I am grateful to Prof. H. T. Kung for suggesting how I could expand my discussion to the implications of this research for the field of corpus linguistics. Further, I would like to thank the Taiwanese friends from whom I originally learned Taiwan Mandarin and its sentence-final particles, including but not limited to Li Jen Tsai, Lili Tsai, Szuhan Hu, Alice Zhan, Fei Chen, and Henry Huang. Finally, I would like to acknowledge the impact of my seven months in Taiwan on my acquisition of Mandarin and Taiwan Mandarin. I would not have been able to complete this project without the knowledge I gained through my time abroad. And I am grateful to the people of Taiwan for making my experience there so enjoyable and conducive to learning.

Table of Contents

ABSTRACT	i
1. INTRODUCTION	1
1.1. The basis for gendered usage.....	1
1.2. Gendered usage in Mandarin	3
1.3. Intonation and particles.....	4
1.4. SFPS in <i>Guoyu</i>	6
2. METHODOLOGY	14
2.1. Corpus selection.....	14
2.2. Dimensions of interest	14
2.3. Problems of data collection	15
2.4. Algorithm for data collection.....	19
3. PREDICTIONS.....	21
3.1. The sentence-final particle <i>o</i> (喔).....	21
3.2. The sentence-final particle <i>la</i> (啦)	22
3.3. The sentence-final particle <i>ma</i> (嘛).....	23
3.4. The sentence-final particle <i>ei/ye</i> (欸/耶)	24
3.5. The conversational context	25
4. RESULTS	27
4.1. Cross-context	27
4.2. Romantic opposite gender	29
4.3. Non-romantic opposite gender	30
4.4. Non-romantic same gender.....	31

5. DISCUSSION	32
5.1. Main takeaways	32
5.2. Limitations of this study	33
5.3. Gender and the SFP-intonation spectrum	36
5.4. Cross-linguistic expectations	39
6. CONCLUSION.....	40
REFERENCES	44

Abstract

Discourse particles are markers that convey the attitude of a speaker towards an utterance and serve to connect units of discourse. Previous research on discourse particles has investigated their pragmatic function, syntactic structure, and typological equivalents. Sociolinguistic research has additionally explored gendered associations of these particles. Sentence-final particles (SFPs) are a type of discourse particle and a distinctive feature of Taiwan Mandarin. Despite their use by both male and female speakers, SFPs in the language are often thought to be feminine words because of their role in *sajiao*, a coquettish speech behavior performed by women and children. This study suggests that there is an additional explanation for the gendered association: SFPs fulfill the pragmatic roles of “women’s language” as discussed in previous literature. A connection is proposed between the gendered use of SFPs and the gendered use of intonation in other languages. This study additionally gathers empirical data on SFPs to investigate the gender disparity in their usage. Using a selection of the National Chengchi University (NCCU) Corpus of Spoken Taiwan Mandarin comprising data from 40 speakers, 20 male and 20 female, the rates at which native speakers of Taiwan Mandarin use the SFPs *o* 喔, *la* 啦, *ma* 嘛, and *ei/ye* 欸/耶 were calculated for different conversational contexts. The results of the study show female speakers use all these particles at a higher rate, with the greatest gender disparity shown for *ei/ye*. Furthermore, the conversational context was shown to be related to the usage rates of *ei/ye* for male speakers, who used this particle twice as frequently when speaking with a female romantic partner than with non-romantic friends or peers. These results provide further evidence that SFPs play an important and nuanced role in the linguistic manifestation of gender identity in Taiwan.

1. Introduction

1.1 The basis for gendered usage

Discourse is a complex social behavior that is not only a process of sharing information but also an outward projection of one's identity (Benwell 2006).

Considerations of status, education level, and gender can explain a range of linguistic decisions made by a speaker, and additionally can reflect the values of the broader community to which a speaker belongs. Gender, as one of the most salient features of a speaker's identity, is manifested clearly in the language of male and female speakers. Research such as that of Robin Lakoff has established speech as one important tool that women use to display female identity, in accordance with society's expectations for women (1972, 2004).

In her seminal 1975 book *Language and Woman's Place*, Lakoff identifies a variety of speech patterns that distinguish "women's language" and offers explanations for their ubiquity among female speakers. She claims that failure to use these forms will result in being "ostracized" as unfeminine by both men and women (Lakoff 2004:40). While Lakoff's discussion is primarily grounded in English, this inquiry, taking as a premise that both languages possess gendered usage, aims to demonstrate that correspondence exists between the features of women's speech that Lakoff identifies in English and sentence-final particles in Taiwan Mandarin. The presence of such functional correspondences between linguistically distinct features are thought to be the result of typological factors; English and Taiwan Mandarin differ in many areas, notably in intonation and final particles. As a result, some expressive features of English are not available in Taiwan Mandarin and vice versa, and the languages are forced to harness

different features to fulfill similar gender functions. The proposal that SFPs are one such gendered feature of Taiwan Mandarin will be tested by this corpus study.

Some of the characteristics of women's speech relevant to a discussion of sentence-final particles will be examined now and later returned to when making predictions about the empirical realities of SFP use. One is the use of question intonation in declarative statements, which can be accomplished via tag questions added after the statement, or simply by adding rising intonation to a declarative (Lakoff 2004:49-50). The former is illustrated in (1). The effect is to avoid committing to the statement and thereby to avoid conflict with the receiver.

(1) Philadelphia is a large city, isn't it? (Ainsworth 1993:278)

The rising intonation form, also known as the high rising terminal, or "upspeak," is illustrated in (2b), to be contrasted with (2a), an example of standard rising pitch when uncertainty is present. In (2b), no uncertainty is present, yet the speaker still raises the pitch at the end of the utterance. The acute accent here indicates rising pitch.

(2) (a) Chris isn't *hé*re? (Ainsworth 1993:282)

(b) Speaker 1: Name, please.

Speaker 2: Madison Taylór.

Another characteristic of women's language is hedges: phrases such as "I guess," "I wonder," and "kinda" that reduce the assertive power of a statement (Lakoff 2004:79). These have a similar effect as the intonation changes discussed above; they reduce a speaker's commitment and thereby prevent conflict with the receiver. Lastly, Lakoff claims that female speakers are more polite than men because of society's expectation that women avoid crass language, have good manners, and preserve a higher standard of

etiquette. She proposes three rules for politeness that characterize polite speech: formality, deference, and camaraderie (2004:88). Formality is the principle of maintaining an appropriate level of social distance between the speaker and the receiver. Deference is giving the speaker options, as opposed to imposing the speaker's will on the receiver. Camaraderie is showing sympathy and understanding of the receiver's situation, as well as displaying a friendly affect towards him/her. Lakoff does not state that all three rules need to be present for speech to be polite. On the contrary, she says that formality and camaraderie are mutually exclusive (2004:87-92).

Having now introduced the basics of Lakoff's theory of gendered language, this study will later return to their implications for SFPs in Taiwan Mandarin.

1.2 Gendered usage in Mandarin

In Mandarin, prior inquiries have explored how female identity is reflected among speakers, and SFPs have been a frequent focus of this research (Yueh 2012:12) (Dabney 2014). Shih (1984) found that female mainland speakers use *ba* 吧, *ma* 嘛, *ya* 呀, *ne* 呢, *la* 啦, and *ye* 耶 more frequently than men. Wamsley (2019) found that women exhibited a greater use of *ba* 吧, a particle that solicits agreement from the receiver, while men exhibited a greater use of *ma* 嘛, a particle that he claims insists on the receiver's commitment to the state of affairs. Similar studies of Cantonese (Chan 1999) and Japanese (Uyeno 1971) have found correlations between the gender of the speaker and SFP usage.

The association of SFPs with feminine speech in the minds of native speakers is similarly well-established (Dabney 2014). The origin of this association likely varies by

particle, as each has a distinct pragmatic function. As will be shown in a later section, the function of the SFPs examined in this study can be linked to Lakoff's characteristics of women's language. For instance, the particle *ba* 吧 can be used to soften commands, a goal of women's language described by Lakoff. Another explanation for the femininity of SFPs is specific to Chinese. Namely, the use of SFPs by women in *sajiao* 撒嬌, a speech behavior seen as highly feminine, has cemented their gendered association (Yueh 2012:60). *Sajiao* is characterized by whiny playfulness in which the speaker utilizes the receiver's fondness of her to get her way (Dabney 2014:5). SFPs' role as a major characteristic of *sajiao* has been documented in numerous studies (Yueh 2012) (Dabney 2014).

1.3 Intonation and particles

To clarify the application of a theory of gendered language in English to Mandarin, it is important to consider how these languages differ in relevant features. Specifically, prosodic features have great bearing on what speakers can vary without changing the meaning of an utterance. A typological link between intonation and SFPs has been repeatedly discussed in the literature. Chinese linguist Yuen Ren Chao noted in a 1932 study that "The speech element in Chinese which may be equated to English intonation is the use of grammatical particles, or 語助詞 [*yuzhuci*] ('speech helping words')" (Chao 1932). Cantonese, a Chinese language rich in SFPs with origins in Guangdong province of southern China, was proposed by Yau (1980) as a language on the opposite end of the SFP-intonation spectrum from English (Wakefield 2010:19). Scholarly attempts to document the inventory of Cantonese SFPs have yielded varying

accounts of their number, with as many as 80 particles proposed in the highest accounts (Wakefield 2010:13). English, on the other hand, is not thought of as a language that possesses SFPs, though a few (such as “huh?”) have been proposed (Wakefield 2010:19).

Wakefield (2010), in his study of Cantonese SFPs and English intonation, describes the limits of Cantonese intonation:

Cantonese is a tonal language with six lexical tones [...] Its tonal contrasts involve both height and orientation, severely restricting the speaker’s ability to manipulate pitch within utterances [...] Changing the tone of a Cantonese word alters its intrinsic meaning rather than merely adding connotative meaning. To compensate for this limiting factor, various types of speech acts, speaker stances, and epistemic modalities that are expressed largely through intonation in English have been segmentalized in Cantonese, resulting in a rich variety of SFPs (pp 11).

In contrast, English has no lexical tones and exhibits prosodic freedom across many features, including intonation and stress. It seems the more a language relies on intonation, the less it will employ SFPs. Indeed, the inverse relationship between intonation and SFPs has been described as “beyond doubt” (Cheung 1986:251 in Wakefield 2010:19). So where does Mandarin lie on the SFP-intonation spectrum, and what implications does this parameter have on language and gender? A precise answer must consider the differences in SFPs and prosody between Mandarin as spoken in Taiwan and in China.

Taiwanese Mandarin, or *Guoyu* 國語, possesses the SFPs present in mainland Mandarin, or *Putonghua* 普通話, in addition to other particles unique to Taiwan. While the emergence of these particles has now been documented in the mainland because of the influence of the Taiwanese entertainment industry, they are still strongly associated

with Taiwan, as confirmed by Dabney's comprehensive study on speakers' perception of SFP use (2014:52, 63). It has been noted that many of the SFPs unique to Taiwan Mandarin entered the language via contact with Taiwanese, the other major variety of Chinese spoken in Taiwan (Lin 2014:22-24). While this contact may explain the rich inventory of SFPs in Taiwan Mandarin, their frequency and range of expressive meaning solicits further explanation. Crucially, *Guoyu* and *Putonghua* differ in prosodic freedom. *Putonghua* has a partially developed stress system, and a given sentence will be read by a mainland speaker with less regular syllable duration, intensity, and pitch than the same sentence read by a Taiwanese (Tseng 2004). Taiwan Mandarin has a smaller pitch range and fewer neutral-tone syllables (Huang 2013:21). It also lacks post-focus compression, a prosodic method of conveying focus found in Standard Mandarin and many other languages such as English and French (Xu 2011). With only a minor utilization of non-lexical intonation, Taiwan Mandarin lies closer to Cantonese on the SFP-intonation spectrum than Standard Mandarin.

1.4 SFPs in *Guoyu*

While the inventory of SFPs in *Guoyu* is extensive, this study will only examine the particles *o* 喔, *la* 啦, *ma* 嘛, and *ei/ye* 欸/耶. While other research has considered the Standard Mandarin particles *ba* 吧, *ne* 呢, *a* 啊, and *le* 了, this study is focused on the particles that are associated the most with Taiwan. As will be discussed further below, certain uses of *o* 喔, *la* 啦, *ma* 嘛, and *ei/ye* 欸/耶 are distinctly Taiwanese. Let us first examine these particles one by one.

1.4.1 The sentence-final particle *o* 喔

Prior research on SFPs represented by the character 喔 often classify it into two distinct varieties: the unmarked *o* and the marked *o* (Dabney 2014:26). The unmarked variety is used to indicate the speaker's own realization of information.

- (3) 「原來是這樣喔」
Yuanlai shi zheyang o
 So be this-way O
 “So that’s how it is”

The unmarked *o* has a “flat, low pitch” (Wu 2005:970 in Dabney 2014:26). It is present in *Putonghua*, where it is more commonly represented with the character 哦, which also exists in secondary use in Taiwan.

The marked *o* is the distinctly Taiwanese version and has a variety of use cases. It is characterized by “a rising or falling-rising pitch contour” and can also be represented by the character 哦 (Wu 2005:970 in Dabney 2014:26). The marked *o* is necessarily addressee-oriented; it comments on or highlights information relevant to the addressee.

- (4) (a) 「小心燙喔!」
Xiaoxin tang o!
 Be-careful hot O!
 “Be careful! It’s hot.”
- (b) 「好好保重喔!」
Haohao baozhong o!
 Good take-care-of-yourself O!
 “Take care of yourself!”
- (c) 「他明天就會來看你喔」
Ta mingtian jiu hui lai kan ni o
 He tomorrow just will come see you O
 “He will come to see you tomorrow”

In (4a), the speaker warns the receiver about the state of affairs. Here, 喔 serves to cue the receiver that new information has been communicated. (4b) exhibits a different

use of the same particle, which here does not indicate the presence of new information but rather adds a tone of endearment to the utterance. In (4c), *o* highlights the importance of the information to the receiver and opens the assertion to the receiver's response.

1.4.2 The sentence-final particle 啦 (*la*)

The next SFP is 啦 (*la*), which has a variety of uses and is loosely categorized as an intensifier. It often suggests annoyance or playfulness on the part of the speaker. It is important to distinguish this particle from *Putonghua* 啦, which is merely a contraction of grammatical particle 了 *le* and affirmation particle 啊 *a*, as shown in (5a) below (Lin 2014:87). While this use of 啦 also exists in *Guoyu*, other usages are more common. Notably, because the SFPs 了, 啊, and 啦 are optional for grammaticality in most cases, it is sometimes impossible to distinguish between the 了 + 啊 deep structure of (5a) and that of other 啦-constructions, like surface forms (5d) and (5e). Later in this section, it will be argued that considering all senses of an SFP as one is warranted for the purposes of this study.

- (5) (a) 「我在上班啦」
Wo zai shangban la
 I at start-shift LA
 "I'm at work (now)"

(5)(a-f) demonstrate that 啦 has a variety of conversational functions. It would be an oversimplification to state that 啦 is an affirmation particle or an intensifier, and an analysis of the various cases is needed to arrive at a sufficiently precise characterization of its meaning.

(5) (b) 「討厭啦」
Taoyan la
 Annoying LA
 “So annoying”

(c) 「你先啦」
Ni xian la
 You first LA
 “You go first”

The 啦 in (5b) expresses annoyance on the part of the speaker. The 啦 in (5c) is a suggestion, similar to *ba* 吧 but indicative of low social distance between the speaker and addressee.

(5) (d) 「不要啦」
Buyao la
 Don't want LA
 “I don't want to”

(e) 「好啦」
Hao la
 Good LA
 “Fine”

(f) 「沒關係啦」
Meiguanxi la
 Doesn't matter LA
 “It's OK”

In (5d), 啦 is a softener that in context allows the speaker to reject a proposal without signaling negative emotion. The 啦 in (5e) carries a degree of coyness or reluctance. It expresses that in the matter at hand, the speaker's acceptance was not to be taken for granted. In (5f), 啦 affirms the speaker's lack of negative emotion after a statement or incident that had the potential to cause a negative reaction. And (5a) is, in at least one reading, a fusion of the modal particle 了 and the affirmation particle 啊. In summary, 啦 does not have a constant meaning, but it is speaker-oriented (Lin 2014:133) and

specifically seems to comment on the speaker's mood towards the state of affairs (with the exception of cases like 5c).

1.4.3 The sentence-final particle 嘛 (*ma*)

The next SFP under consideration is 嘛 (*ma*). This particle has been variously described as an indicator of obviousness, or alternatively, of the speaker's insistence that the receiver commit to the state of affairs (Wamsley 2019:5). However, not all instances of the particle are so forceful. A more nuanced interpretation is needed. Data such as (6) do seem to indicate unsurprising information, but perhaps a better description is that the particle indicates that the information preceding it is likely already known to the receiver. In this way, such a use of 嘛 is the pragmatic opposite of 喔 as used in (4c).

- (6) 「他美國人嘛，沒有健保」
Ta meiguoren ma, meiyou jianbao
 He American MA, not have National Health Insurance
 “As you know, he's American, so he doesn't have National Health Insurance”

The acknowledgment of the receiver's awareness of information is frequently a polite tactic, similar to the English phrase “as you know.” To fail to acknowledge the receiver's prior awareness of information runs the risk of suggesting that the receiver was truly unaware, which could be construed as condescending. Additionally, there are separate uses of 嘛 that bear little relation to this one. For instance, the particle commonly appears in *sajiao*, a behavior common among young women and children used to get the receiver to do what the speaker wants (Yueh 2012:26). In this context, the particle can be a marker of petulant insistence, and often has a lengthened phonetic realization.

- (7) 「我吃不下了嘛」
Wo chibuxia le ma
 I eat-not-down LE MA
 “I can’t eat anymore”

In (7), 嘛 is drawn-out and has a falling tone. The fact that the speaker feels full cannot be described as obvious information or an acknowledgment of the receiver’s knowledge, as in the previous example. Instead, this *sajiao* use of 嘛 follows new information. Its function may be described as an insistence or plea that the receiver understand the speaker’s mental state.

- (8) 「人家就是很想你嘛」
Renjia jiu shi hen xiang ni ma
 People just be very miss you MA
 “I just miss you a lot”

Another use of 嘛 in *sajiao* is (8), in which the particle indicates playful insistence or obviousness. This 嘛 also has some indication of annoyance, expressing that the receiver should already know the information preceding the particle, or should not be forcing the speaker to communicate it in such a direct way. Previously, the description of 嘛 as a particle that indicates obviousness was deemed incomplete. It is true that the particle in (8) expresses obviousness, though crucially the tone is whiny and playful, not assertive or domineering. Indeed, *renjia*, a pronoun which when used in the first-person is considered strongly feminine, would not be used in a sincere attempt to assert oneself.

1.4.4 The sentence-final particle 欸/耶

The last particle examined in this study is 欸 *ei* / 耶 *ye*, two similar particles which are normally considered variants of the same form (Shie 1991:16 in Lin 2014:123-

124). This particle is likely a recent product of the influence on *Guoyu* of Taiwanese, the native language of a large segment of the Taiwanese population (Dabney 2014:30). The particle has been described as an “attention getter” as well as an indicator that the receiver or speaker’s understanding of a matter is insufficient or incorrect (Tsai 2008:1023). In (9), the speaker indicates his/her surprise that next week is not a full week of class, as had been his/her expectation.

- (9) 「我們下星期只上四天的課欸!」
 (Dabney 2014:28) *Women xia xingqi zhi shang si tian de ke ei!*
 We next week only attend four-CL day-M DE class EI
 “We only have four days of class next week!”

Importantly, this particle is functionally different from the unmarked and marked 喔.

While 欸 and the unmarked 喔 can both indicate the speaker’s realization of new information, 欸 implies that the realization goes against the speaker’s prior understanding of a matter, whereas 喔 does not carry such an implication. Furthermore, a speaker can use 欸 to call the receiver’s attention to new information, but this sense is not the same as the marked 喔. The marked 喔 as in (4) is an endearing way to inform and give options, whereas the addressee-oriented 欸 is used to correct an incorrect assumption or action of the receiver, as in (10), which resembles the English use of *hey!* in situations of injustice.

- (10) 「你插隊欸!」
Ni chadui ei!
 You cut-in-line EI!
 “(Hey!) You’re cutting in line!”

Another use of the same particle has been described as an effort to cooperate with a question despite the inability to provide an answer, a type of softener (Shie 1991:169 in Lin 2014:131).

(11) Speaker 1: 「為什麼他沒有來上課？」
Weisheme ta meiyou lai shangke?
 Why he not come attend-class?
 “Why didn’t he come to class?”

Speaker 2: 「我不知道耶！」
Wo buzhidao ye!
 I NEG know YE!
 “I don’t know!”

In (11), Speaker 2 is unable to answer Speaker 2’s question and adds the particle 耶 to her negative response. Note that the same response without the particle is fully acceptable, and absent negative tone-of-voice such a response would not necessarily be interpreted as aloof or standoffish. Therefore, the presence of the particle is merely a clarification of the speaker’s acceptance and cooperation with the question.

The four sentence-final particles investigated by this study encompass a diverse range of meanings and are difficult to define without the surrounding context. As such, one might object to the idea that a given particle carries the same associations and connotations in all instances. For example, suppose a particle had a set of “feminine senses” as well as a set of “masculine senses.” Then an investigation of the *overall* frequency of use of this particle by men or women is unlikely to reveal much about gendered use of language. Yet, in the case of sentence-final particles in Mandarin, very few “masculine senses” have been proposed. Indeed, all senses of all SFPs in Taiwanese Mandarin seem to be available to female speakers, whereas at least some particles have senses that are not typically available to males. Because it is assumed that there are no “masculine uses” in the data, then the “femininity” of a particle and frequency of its use by female speakers should be directly proportional. As such, the treatment of all instances of an SFP as uniform should still yield meaningful results.

2. Methodology

2.1 Corpus selection

As Taiwan's population is roughly sixty times smaller than that of mainland China, the comparatively small number of *Guoyu* corpora is understandable. The National Chengchi University Corpus of Spoken Taiwan Mandarin was chosen, among few alternatives, for the following reasons. First, it provides extensive, recent (2006-2017) data of conversations between native speakers of Taiwan Mandarin, at present comprising 43 conversations. Second, the conversations are annotated with basic demographic information about their speakers, including age, gender, and the speakers' relationship to each other. Third, both the written transcriptions and conversation audio are provided, facilitating clarification in certain cases of ambiguity. Fourth, the corpus includes conversations between individuals of diverse relationship types, including romantic relationships, friendships, coworkers, family, and classmate relationships. Because of these characteristics, the corpus is suitable for an inquiry of this kind, which seeks to isolate the variables of gender and conversational context.

2.2 Dimensions of interest

As this study primarily concerns the frequency of SFP use by males and females, speakers are grouped according to their gender. They are also classified by their relationship with the person they are speaking with. The relationship types examined by this study are opposite gender romantic partner, opposite gender friend or other non-romantic colleague, and same gender friend or other non-romantic colleague. In total, data from 40 speakers and 21 conversations were analyzed, with the following

distribution: five one-on-one conversations between male and female romantic partners, four one-on-one conversations between male and female non-romantic friends or other colleagues, with the addition of one female speaker from a conversation with two male colleagues and one male speaker from a conversation with two female colleagues; additionally, four one-on-one conversations between male non-romantic friends or other colleagues, with one additional conversation between three such males, and five one-on-one conversations between female non-romantic friends or other colleagues. The data from each speaker averaged 3391 Chinese characters, equivalent to 3391 syllables of speech. The speakers' conversation data was at a minimum 1229 characters and at a maximum 9816 characters long.

2.3 Challenges of data collection

The identification, classification, and counting of SFPs in the text of natural conversations is not without its challenges. First, the orthographic conventions used in the transcriptions of the NCCU corpus, while largely consistent, at times vary from conversation to conversation. For instance, while SFPs are normally represented by a Chinese character in the text, they are occasionally transcribed in a romanized form such as “ei” or “la.” The prevailing convention of the corpus is to reserve roman letters for actual English text and non-linguistic utterances such as filler noises or other vocalizations. While it is not challenging to convert these transcriptions to Chinese characters, care must be taken to prevent a false positive in the case that a romanized transcription does not actually represent an SFP.

A further challenge is the widespread presence of code-switching within the corpus. Many speakers of Taiwan Mandarin freely switch to Southern Min or Hakka, two other Chinese languages, a phenomenon especially common in informal contexts such as those documented by this corpus. Japanese and English words and phrases also frequently appear in the corpus intermixed with the Taiwan Mandarin dialogue. The presence of other languages presents two distinct problems for data collection. One, because Min and Hakka transcriptions sometimes employ the same characters found in Mandarin as SFPs, the naïve solution of counting all instances of an SFP character will collect false positives. Two, word boundaries in non-Mandarin portions of the corpus are often poorly documented, inconsistent, or absent. This causes issues when calculating the total number of words spoken by a speaker for the purpose of an SFP frequency calculation. Furthermore, when counting speech with syllables or characters as a unit, the problem arises of how to account for Japanese or English glyphs, which represent a variable number of syllables. The solution adopted by this study is to fully neglect the portions of the conversations not in Taiwan Mandarin, encompassing all text marked as belonging to another language by the corpus transcription. As an effort to minimize the effect of this decision on the accuracy of the data, conversations with heavy use of Southern Min or other languages were not analyzed. It is acknowledged that this decision will lead to youth bias in the results; those who frequently code-switch and were therefore not analyzed tended to be the older speakers in the corpus. However, absent a better transcription of the Min data, the researcher was unable to obtain meaningful word-frequency figures for these speakers. The resulting bias of the data can be partially justified by the acknowledgment that young speakers are more likely to use SFPs in the

first place, and most of the SFPs studied in this inquiry are considered to belong to youth language (Dabney 2014:57).

Another significant issue that arose during the data collection was the resolution of ambiguities in the transcriptions. Not all instances of a character that can represent an SFP necessarily represent an SFP; in fact, the same characters are frequently used to transcribe exclamations, thinking noises, and filler words. Some of the characters used to transcribe SFPs are even present in more “normal” lexical items. For conciseness, in the remainder of the discussion, the term “SFP-character” will be used for any Chinese character that can be used to transcribe an SFP.

An additional obstacle is that the NCCU corpus does not segment the text into sentences, so the sentence-final position occupied by SFPs cannot be reliably located. Indeed, sentence segmentation in Chinese has proved to be a fundamental issue in Chinese language processing, and this study does not attempt to segment the text in pre-processing. Instead, this study takes advantage of the high prevalence of ellipsis (“...”) in the NCCU corpus, using this delimiter as an approximation for the sentence boundary. These delimiters will frequently correspond with the true sentence boundary and thus are a good starting point in the search for SFPs. However, it is posited that the ellipses in the corpus, which appear to correspond primarily with pauses, are likely to represent an under-segmentation of the text. In fluid speech, a speaker often crosses the sentence boundary with great speed, resulting in the inadequacy of pause-based segmentation. This issue must be accounted for. Consider the following corpus data.

- (12) [彭美花]..我常常回去看他的耶美花姊我們叫他美花姊
 “Peng Mei-Hua...I often go back to see her YE Sister Mei-Hua
 we call her Sister Mei-Hua”

In (12), we see an example of under-segmented text that contains an SFP-character (highlighted). A programmatic approach that only examines the segment-final position will miss this SFP-character and therefore be inadequate. Thus, we must expand the search to segment-internal contexts, lest we systematically undercount.

A less common scenario is that the pause-based segmentation results in an over-segmentation of the text. In this case, a programmatic approach that examines the segment-boundary position for SFP-characters will yield more candidate SFPs than the true amount.

- (13) 你們幹嘛...幹嘛去酒吧跟他見面
 “Why do you...Why [do you] go to bars to see him”

In (13), the pause-based segmentation would lead us to erroneously consider a segment-final context that does not correspond with a true sentence boundary. This specific problem is a special case of the broader issue that SFP-characters do not always represent SFPs, and thus need not be addressed directly.

The solution employed by this study to manage the ambiguities in the transcription can be summarized as follows: utilize various heuristics to reduce the ambiguities to an amount which can be feasibly examined manually. In the case of the occurrence of SFP-characters in non-SFP lexical items, taking no remedial action is normally justified because of the very low frequency of these lexical items. One exception, however, is the common word *ganma* 幹嘛, which contains the SFP-character 嘛. All occurrences of the string 幹嘛 are neglected, under the reasoning that the probability of an SFP 嘛 occurring after the word 幹 is negligible, especially in Taiwan Mandarin where in the researcher’s experience 幹 is a low frequency character. In the

cases where exclamations or non-linguistic utterances are transcribed with SFP-characters, this study makes use of heuristics that rely on the rough segmentation present in the corpus. By definition, an SFP cannot occur at the beginning of a sentence, or at the beginning of any clause for that matter. As such, it is assumed that any instance of corpus data where a delimiter is followed immediately by an SFP-character need not be considered. It would be highly unlikely for a speaker to break from fluid speech and then add on an SFP after the pause. In instances where the same SFP-character appears repeatedly, as is common in exclamations such as 喔喔喔 (“Oh oh oh”), the whole cluster is disregarded. The segmentation issue illustrated in (12) was empirically found to be infrequent and thus feasibly handled by manual examination. However, it should be noted that the manual judgements required to resolve this problem still made up the bulk of the data collection time, and a rigorously segmented corpus would still be preferable for future inquiries.

2.4 The algorithm for data collection

In summary, the program designed for the corpus search works as follows:

1. Identify all instances of an SFP-character.
2. Disregard those that occur in impossible segmentation contexts.
3. Disregard those that occur consecutively with other instances of the same SFP-character.
4. Disregard 嘛 after 幹.
5. Mark those that appear segment-internally for further review.
6. Accept the remaining SFP-characters.

7. Review the SFP-characters tagged in step 5 manually.

While this algorithm has limitations, which have already been discussed, an empirical test of the algorithm on one of the corpus conversations yielded the same set of SFPs as a completely manual approach. It is believed that the greatest source of error, if any, will result from erroneous manual judgments made in step 7, and not as a result of any of the heuristics employed.

3. Predictions

Using Robin Lakoff's description of women's language in English as a starting point, this study extrapolates from her observations and hypothesizes that distinct speech behaviors in Mandarin that accomplish the same goals as she describes in English will similarly be more prevalent among female speakers. Furthermore, it is hypothesized that a particle's prevalence in *sajiao* will be correlated with its greater use among female speakers.

3.1 The sentence-final particle 喔

It is predicted that 喔 will be used by female speakers at a higher frequency. In the case of 喔-sentences like (4a), the speaker shares new information with the receiver and calls attention to the novelty of the information. The role of the SFP in this example can be analyzed as either sympathetic or deferential. As the utterance without the SFP would be a direct claim regarding a state of affairs that affects the receiver, adding the SFP converts such a direct assertion into an expression of concern or warmth towards the receiver. The speaker's concern for the receiver is implicit in the speaker's belief that the receiver ought to be aware of the information shared. In a fully impartial context, in which the speaker has not considered how the information he/she shares might affect the receiver, the speaker would have little pragmatic need to communicate the information's importance to the receiver. The presence of 喔 indicates the speaker has indeed considered the receiver's situation and furthermore believes that it is in the receiver's interest to be aware of the information. As such, this 喔 is an expression of sympathy, one that softens the blow of a receiver-oriented assertion by tying it to the self-interest of

the receiver. Lakoff claims that interest in the mental states of others is typically reflected in female speech (2004:101). It follows that this “sympathetic 喔” could likewise be a female-gendered usage. A separate pragmatic analysis of 喔 for the same data interprets it as a way to give the receiver options. A speaker may wish to communicate the information in (4c) without imposing it on the receiver. With 喔 present, the receiver is licensed to respond to the assertion as he/she sees fit, for instance, by acceptance or rejection. With 喔 not present, the speaker makes more of a commitment to the information and leaves the matter less open for debate. Lakoff, who claims that women are supposed to be more polite than men, proposes deference to the receiver as a maxim for politeness (2004:88). In this light, a female association for this sense of 喔 emerges. As these uses of 喔 and many others can be assigned to the pragmatic functions of women’s language, it is predicted that 喔 will be used more by female speakers.

3.2 The sentence-final particle 啦

In the discussion of 啦 in 1.2.2, various conversational functions were proposed. Among them, many can be linked to *sajiao* or to the goals of women’s language as described by Lakoff. For instance, in (5d) and (5f), the softening function of 啦 reduces the perception of negative emotionality and thereby preserves harmony and avoids conflict. Recall that the same functions were present in English question intonation as discussed earlier. The annoyance expressed in (5b) and the coyness of (5e) are common features of *sajiao*, where they are tactics in persuasion. Because of the particle’s

application in *sajiao* contexts and its softening and harmony-preserving functions, it is predicted that 啦 will be used more by female speakers.

3.3 The sentence-final particle 嘛

The previous discussion of 嘛 has noted its occasional function of insistence.

Given the association between assertive speech and masculinity, some have proposed 嘛 as a rare “masculine” SFP. This is the view taken by Wamsley (2019), who found a greater use of 嘛 by men, albeit with a small sample size (n=8) and in a highly particular context of one Chinese talk show (2019:9). However, other research shows that 嘛 is more common among speakers in the south of China, where women are considered more traditionally feminine, and suggests that the particle retains a feminine association (Dabney 2014:24, 49). Dabney offers an explanation for this pattern, claiming that while 嘛 does convey insistence, males, as authority figures, might “feel less of a need to be insistent” (2014:24). As such, 嘛 “has evolved into a platform for a female to be insistent despite her status as a woman” (2014:24). Furthermore, the insistence expressed in data like (7) and (8) is distinctly whiny, akin to how a child might make a demand of its parent. As such, the proposal of 嘛 as a tool of masculine discourse falls short in at least this case. Finally, the “as you know” function of 嘛 (6) is a polite acknowledgement of the receiver’s knowledge. According to Lakoff’s theory, women use polite forms more frequently (2004:94-102). As such, we would expect 嘛 in these uses to be more common among female speakers.

3.4 The sentence-final particle 欸/耶

Like the others, a gendered usage of this particle derives from its presence in *sajiao* and its pragmatic relationship to the goals of feminine speech. As per the prior discussion, 欸 can indicate a realization, or more specifically, the negation of a prior assumption on the part of the speaker or receiver. In the context of *sajiao*, this particle often appears after statements about the receiver, such as ‘you’re cute’ or ‘you’re annoying,’ with the implication that the speaker has just realized the receiver’s possession of this trait. Of course, *sajiao* is most frequently directed at a familiar person, so the sincerity of such realizations is typically limited. Instead of legitimate realizations that occur organically, 欸-sentences in *sajiao* are better thought of as verbal rewards or punishments in response to the receiver’s behavior. As such, they become tools of persuasion. Additionally, in non-*sajiao* contexts like (11), 耶 can be an expression of cooperation when unable to answer a question. The particle is a friendly attempt to avoid discomfort or awkwardness after failing to provide the interlocutor with the information they seek (see 1.4.4). Lakoff describes friendliness as a form of politeness, which is a characteristic of women’s language (2004:90). This SFP serves to maintain an air of friendliness in question contexts. Because of its gendered use after questions and in *sajiao*, it is unsurprising that Dabney found this particle to have the strongest association with femininity of all the particles examined (2014:49). It is predicted that 欸/耶 will similarly exhibit a strong female usage trend in this study.

3.5 The conversational context

In the analysis of SFPs so far, it has been established that the particles often soften statements, express emotion, and suggest concern or attention for the receiver. As such, it is predicted that SFPs will be present in higher frequency in conversations between romantic partners, where they function as indications of affection. This prediction extends to men, who are not expected to use many SFPs, and are predicted to use them more with their romantic partners than in other contexts. Furthermore, in romantic contexts, one would expect to find relatively more *sajiao*. As such, it is predicted that SFPs will be used the highest among women talking to their romantic partner, more than in any other conversational context tested.

On the other hand, in non-romantic contexts in which *sajiao* and displays of affection are less common, it is predicted that SFPs will be less frequent. In the opposite-gender non-romantic partner category, men will use few SFPs, yet women will still exhibit high SFP use. The societal forces and expectations that result in women's higher usage rates should apply in all contexts, so it is expected that women will continue to employ many SFPs in non-romantic contexts.

If SFPs are indeed a part of women's language, it is reasonable that male speakers wishing to project a masculine personality would benefit from reduced use of SFPs. In conversations with their male peers, this goal may become even more prioritized, as neither party wishes to come off as feminine. As such, in same-gender non-romantic contexts, men are expected to use SFPs even less than they do in opposite-gender non-romantic contexts. As SFPs are thought to suggest a feminine personality in all contexts,

it is expected that female speakers will use SFPs at the same rate in same-gender as in opposite-gender non-romantic contexts.

4. Results

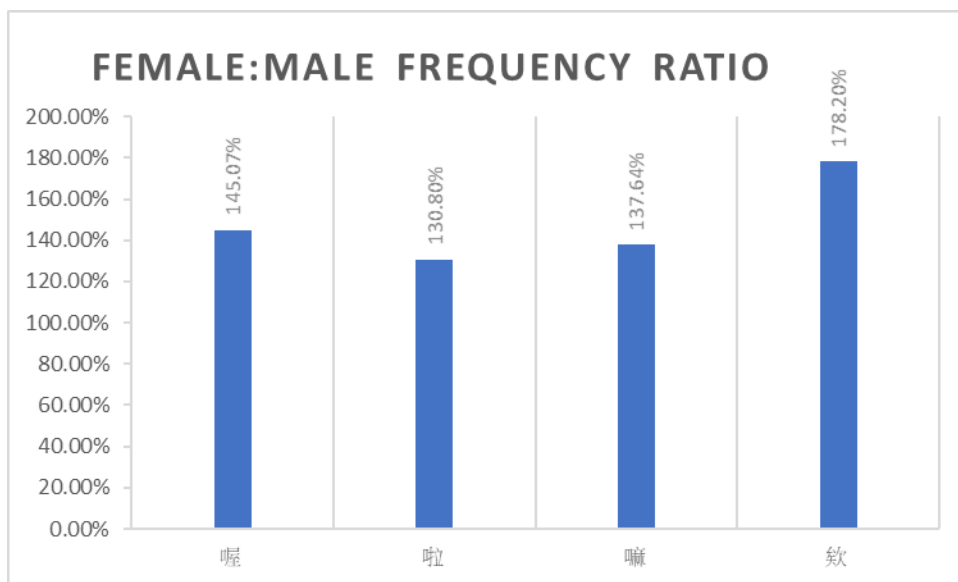
The results are presented as frequency figures for the SFPs out of all characters spoken by the set of speakers, converted to a percentage. The figures are not word frequencies, but rather character frequencies i.e. each syllable contributes equal weight to the frequency figure. As such, the values are strictly less than their corresponding word frequency values.

4.1 Cross-context

Table 1: SFP frequency for male and female speakers across all contexts

	喔 <i>o</i>	啦 <i>la</i>	嘛 <i>ma</i>	欸/耶 <i>ei/ye</i>
Male	0.29%	0.27%	0.06%	0.15%
Female	0.42%	0.36%	0.09%	0.26%
F:M ratio	1.45	1.31	1.38	1.78

Graph 1: Cross-context female:male usage frequency ratio



For both genders, the frequency ranking of the four particles was identical. The most frequent particles for both men and women were 喔 and 啦, which are relatively versatile and can be used by speakers of both genders to draw attention or add emphasis to a wide range of utterances. 啦 can also add a tone of annoyance to a variety of statements, and this sense of the particle is available to both genders. 欸/耶 is less frequent, likely because of its recent introduction to the language and more limited range of meanings. 嘛 is the least frequent particle, as it is only felicitous in a few specific contexts. Although the ranking was the same, male and female speakers exhibited large differences in the amount of SFPs used. As Table 1 demonstrates, females used all the SFPs tested at least 30% more frequently than males. Graph 1 is a graphical representation of the last row of Table 1. 欸/耶 exhibited the greatest gender disparity in usage, with females using the particle 78% more than males. 喔 was the most frequently used particle for both genders, almost five times more common in the data than the least frequent particle, 嘛. The particle that exhibited the greatest gender parity, 啦, was still used 31% more by female speakers. The results provide clear support for the hypothesis that these SFPs are used most heavily by female speakers in Taiwanese Mandarin.

What accounts for the unequal gender disparity between particles? It is taken as an assumption that the *sajiao* uses of an SFP are unavailable to most male speakers. Considering the remaining uses of the particles, it is thought that 啦 and 嘛 have senses that are not necessarily gendered, such as those in (5c) and (6), respectively. The prevalence of these uses ought to at least partially account for the relative popularity of these particles among males. 喔 in its unstressed form (3) may be another example of a

relatively gender-neutral particle. However, these more neutral senses are less clear for 欸/耶, which seems more categorically female.

4.2 Romantic opposite gender

Table 2: SFP frequency for male speakers in opposite-gender romantic context

	喔 <i>o</i>	啦 <i>la</i>	嘛 <i>ma</i>	欸/耶 <i>ei/ye</i>
M-Rom-Opp	0.28%	0.37%	0.07%	0.23%

Table 3: SFP frequency for female speakers in opposite-gender romantic context

	喔 <i>o</i>	啦 <i>la</i>	嘛 <i>ma</i>	欸/耶 <i>ei/ye</i>
F-Rom-Opp	0.46%	0.30%	0.02%	0.30%

In romantic contexts, male speakers used the SFPs 啦 and 欸/耶 more than they did on average. Perhaps most significant is the high frequency of 欸/耶, which in romantic contexts men used almost as much as females used the particle on average. The rate at which male speakers used 欸/耶 in romantic contexts is almost double that found in other contexts. This result, coupled with men's relatively high use of 啦 in the romantic-partner conversations, lends support to the claim that there are at least some SFPs that men are more likely to use when talking to a romantic partner. However, men did not use 喔 or 嘛 more frequently in these contexts. Contrary to prediction, female speakers did not exhibit much of an increase in SFP use in romantic contexts. While they

used 喔 and 欸/耶 at slightly higher rates, they used 嘛 only a quarter as frequently as they did on average.

4.3 Non-romantic opposite gender

Table 4: SFP frequency for male speakers in opposite-gender non-romantic context

	喔 <i>o</i>	啦 <i>la</i>	嘛 <i>ma</i>	欸/耶 <i>ei/ye</i>
M-Non-Opp	0.32%	0.14%	0.05%	0.14%

Table 5: SFP frequency for female speakers in opposite-gender non-romantic context

	喔 <i>o</i>	啦 <i>la</i>	嘛 <i>ma</i>	欸/耶 <i>ei/ye</i>
F-Non-Opp	0.39%	0.34%	0.08%	0.24%

In non-romantic opposite gender contexts, men used SFPs at about the rate that they did on average, with one exception: 啦 was used much less frequently by men in this context than elsewhere. Women used SFPs slightly less than they did on average across all the particles tested. These results generally conform with the predictions made, as fewer particles were used in these contexts than in romantic contexts. The exception is 嘛 for females, which did not decrease in frequency because of the unusually low frequency of this particle in female speech in romantic contexts.

4.4 Non-romantic same gender

Table 6: SFP frequency for male speakers in same-gender non-romantic context

	喔 <i>o</i>	啦 <i>la</i>	嘛 <i>ma</i>	欸/耶 <i>ei/ye</i>
M-Non-Same	0.27%	0.33%	0.07%	0.12%

Table 7: SFP frequency for female speakers in same-gender non-romantic context

	喔 <i>o</i>	啦 <i>la</i>	嘛 <i>ma</i>	欸/耶 <i>ei/ye</i>
F-Non-Same	0.42%	0.39%	0.12%	0.26%

In non-romantic same-gender contexts, men's use of SFPs resembled their average usage rates. The prediction that men would use especially few SFPs in male-male conversations was not strongly supported by the findings. Females used SFPs at rates similar to their average. This finding supports the prediction that the gender of the receiver would not affect females' use of SFPs in non-romantic contexts.

5. Discussion

5.1 Main takeaways

The results of the study strongly support a gender divide for the particles 喔, 啦, 嘛, and 欸/耶 in Taiwanese Mandarin. Each of these particles was found to be used 30-80% more by females than by males. The large gender gap in usage frequency suggests that these particles are indeed part of “women’s language” in Taiwan Mandarin.

Furthermore, there is evidence that male speakers are more comfortable with 啦 and 欸/耶 in conversations with their romantic partners than in other contexts. In this study, men speaking to their girlfriends or wives used 欸/耶 almost twice as much as they did in other contexts. Other particles did not see such a large increase among male speakers in this context. This finding is especially notable given that other research such as Dabney (2014) has found this particle to be the most feminine of all SFPs in Taiwan Mandarin. Various factors might explain the contextual sensitivity of 欸/耶 among male speakers. The particle as used in (9) suggests the inaccuracy of a prior assumption on the speaker’s part. In this sense, to use the particle is to suggest that one has been mistaken. Perhaps men are more willing to admit they have been mistaken with romantic partners, whom they presumably trust and receive validation from. That said, uses like (10) rather correct the assumption or action of the receiver. If instances like (10) can explain the increased frequency of 欸 among males in romantic contexts, perhaps it is that men are using the particle to correct their partners’ false assumptions.

5.2 Limitations of this study

5.2.1 Representativeness of the corpus

A natural question to ask in a corpus study is how representative the corpus is of the population from which it is collected. Detailed information on the collection process of the NCCU corpus is not readily available (Chui 2008). For instance, it is unclear whether the speakers hail from different parts of Taiwan, or mostly from Taipei where NCCU is located. Without this information it is difficult to make conclusions about the entire community of Taiwan Mandarin speakers. However, the title (“NCCU Corpus of Spoken Taiwan Mandarin”) and the stated goals of “research and teaching” suggest that some effort has been made to make the data representative of the linguistic community (Chui 2018). The wide range of conversational contexts and 11-year timespan also suggest a degree of participant diversity in the corpus.

As discussed in section 3, the results of this study have greatest bearing on the habits of young speakers of the language. Because of the difficulty of processing the Taiwanese Southern Min data in the corpus, the conversations of many speakers who frequently code-switched could not be included in the analysis. As older speakers are more likely to use Taiwanese (Ministry of Education 2010), the removal of these conversations from consideration results in a youth bias in the data. However, many of the particles tested in this inquiry have strong associations with youth to begin with (Dabney 2014). Furthermore, 欸/耶 is a relatively new particle and as such especially likely to be used by young speakers (Starr 2011). While a study of SFPs need not be limited to their use by young speakers, it is thought that focusing on young people to the exclusion of other ages will not obscure the fundamental truths about SFP use in Taiwan.

5.2.2 Individual variation

While a corpus study has the potential to reveal meaningful information about a linguistic community, it is important to consider the potential impact of individual variation and confounding variables. Personal experience has revealed that use of SFPs varies widely between speakers of the same gender. A large enough and sufficiently representative sample should minimize the effect of this variance. Yet this study, when isolating the variable of conversational context, was unfortunately unable to consider more than five to ten speakers for each gender and contextual category. The NCCU corpus does not include enough conversations to support a substantially larger sample size for the contexts tested in this study (Chiu 2008). Small sample size may explain the data's failure to support some of the predictions made about the effect of conversational context. Furthermore, it has occurred to the researcher that the use of a corpus study to isolate the effect of conversational context may encounter a more fundamental challenge. Namely, corpuses that collect "natural" conversation data may tend to log conversations between individuals who know each other. This is understandable, as it would be difficult for such lengthy and natural conversations to arise between strangers. But using typical corpora to draw conclusions about how SFP use, or any linguistic feature, varies with context may therefore result in selection bias. For instance, the speech of an individual who frequently talks at length with an opposite-gender friend may not be representative of other speakers of their gender.

Specifically, it was noted by the researcher during data collection that some male speakers used speech patterns associated with the gay community. While the corpus did not provide information on the sexual orientation of speakers, the researcher and native-

speaker informants shared a clear intuition for certain speakers. Amid the widespread acknowledgment that gay speech differs in significant ways from hetero speech (Bouavichith 2017, Davenport 2018), it is important to recognize the confounding effect that this factor could have on the data collected, especially for research that directly concerns topics such as gender, masculinity, and femininity.

5.2.3 The ideal corpus

As expressed in the previous section, various qualities of the NCCU corpus presented challenges for the research conducted in this study. As such, it is valuable to describe the characteristics of a corpus that would better lend itself to research of this kind. Perhaps the most important change that could be made is in the segmentation scheme present in the transcription. For research on discourse particles, which typically act as punctuation between utterances, the ability to locate sentence boundaries is crucially important. As a solution, conversation transcribers could simply provide their own sentence segmentation based on intuition as native speakers. If such judgments proved inconsistent between transcribers for the same data, then segmentation could be reviewed by those who produced the data themselves after an initial transcription is drafted. It is thought that such a process would ultimately yield an accurate and consistent segmentation scheme and would avoid the problem of retroactive computational segmentation, which is an open problem in Chinese natural language processing. Another design decision that would yield an improved corpus for the purposes of this study is the standardization of orthographic conventions. As sizeable corpora such as the NCCU corpus likely rely on multiple individuals to transcribe the full extent of oral conversation

data, it is important to ensure consistency among different transcribers. If the same word or particle has a different orthographic representation in different conversations, researchers must thoroughly account for such variation or else suffer losses in accuracy. Additionally, to ensure the adequate representation of individuals who code-switch, data in languages other than the primary language of study must be transcribed consistently and with clear word or syllable boundaries. Otherwise, researchers who are not speakers of the secondary language may find it prohibitively difficult to properly account for its presence.

5.3 Gender and the SFP-intonation spectrum

In section 1, differences between Taiwan Mandarin and Standard Mandarin in SFP inventory and prosody were explained by their differing positions on a typological spectrum from intonation-rich languages to SFP-rich languages. Cantonese, a language extremely rich in SFPs, represents one end of the spectrum, with English and other SFP-poor languages at the opposite end. It is proposed that Standard Mandarin, with a semi-developed stress system and high frequency of syllables in the neutral tone, as well as a handful of sentence-final particles, falls closer to the middle of the spectrum. Taiwan Mandarin, with stricter syllable-timing and a high use of SFPs, lies closer to the Cantonese end.

Linguistic choices reflect gender identity, that is, the speech habits of men typically diverge from the speech habits of women (Benwell 2006, Lakoff 2004). This corpus inquiry establishes SFPs in Taiwan Mandarin as an example of a gendered speech habit; women use the four particles examined more often than men. If we assume that

gendered language is present cross-linguistically, then the question arises of how languages that lack SFPs compensate for their absence. Considering that intonation seems to perform a function similar to that of SFPs in languages that lack them, a natural hypothesis is that gender may manifest itself in intonation as well. Sure enough, studies have found that English intonation varies systematically by gender (McConnell-Ginet 1978) (Jiang 2011). Speakers who use more upward intonation and a “larger utterance semitone range” are more likely to be perceived as female (Cartei 2012). Information found online counseling those undergoing a transition from male to female emphasizes the importance of dynamic, feminine intonation when aspiring to pass as female ([The Voice Stylist 2019]). And though biological processes cause male and female voices to diverge acoustically during adolescence, consistent differences have been found in male and female voices even before such divergences occur, a finding that suggests that biology alone cannot explain the broad differences in male-female intonation (Cartei 2013). Instead, cultural factors influence the way men and women manifest gender with their voices.

Furthermore, it seems the gender of the receiver could affect a speaker’s intonation, drawing a parallel with the similar trait of SFPs as proposed by this study. Recall uptalk, rising intonation in non-question contexts, a pattern associated with female speech. A study of 300 *Jeopardy!* contestants found that men on the show were 29% likely to use uptalk when playing against two female contestants, yet only 14% likely to do so when facing two male opponents (Linneman 2013). With one male and one female contestant, men were slightly less likely to use uptalk than when playing against two women. The strength of this study is compounded by the arbitrary nature of the

conversational context: a contestant's opponents are determined by the producers of the show. As such, the selection bias resulting from prior familiarity with the recipient is not present (see 5.2.2). However, Linneman suggests that the relative absence of upspeak in all-male contexts is due to the heightened competitive atmosphere that emerged in these contexts, a phenomenon that could be especially salient in a game show setting.

Moreover, upspeak is just one aspect of gendered intonation in English. Further research is needed to determine whether male speech in the general case becomes less intonational when speaking with other men.

Additional evidence for a fundamental relationship between gendered SFPs and gendered intonation can be found in Wakefield's 2010 study of Cantonese SFPs and English intonation. Wakefield proposes that certain SFPs in Cantonese can be consistently matched to English intonational "equivalents" (2010:2). He found that bilingual native speakers of Cantonese and English consistently convert Cantonese SFPs into the same linguistically meaningful pitch contours in English. Additionally, the study found that the SFP-pitch contour pairs were acceptable in similar contexts in their respective languages. Wakefield argues that the pitch contours are "floating tone morphemes," as they "consistently represent a particular meaning and function" (2010:246) (2016:1). As such, they can be thought of as a suprasegmental equivalent to SFPs, their segmental counterpart in Cantonese.

In summary, the gender-intonation link in intonation-rich languages and gender-SFP link in SFP-rich languages, coupled with the observation that that intonation and SFPs play similar pragmatic roles and respond to similar contextual factors suggests a deep connection between the two systems, the linguistic system of SFPs and intonation,

and the sociocultural system of gender expression in language. It seems the way speakers project gender in a language depends on the expressive avenues present in that language.

5.4 Cross-linguistic expectations

If the theory presented in 5.3 is correct, then we would expect to find examples of other intonation-rich languages like English where intonation is a strongly gendered feature, and other final-particle-rich languages like Cantonese where particles are a strongly gendered feature. Furthermore, in languages where both features are present, such as Standard Mandarin, we would perhaps expect a distributed reliance on both features for gender expression. Such a distributed reliance might manifest itself in a reduced male-female disparity in the frequency of either gendered intonation or SFP-use. Alternatively, in such a language, the range of strongly gendered SFPs and intonation contours may be relatively limited, and a wider range of SFPs and intonation may be available to both genders. However, it is not expected that languages with both SFPs and intonation will have a reduced reliance on these linguistic features to express gender. Rather, these features will still be powerful tools of gender expression, but neither feature will be dominant over the other in gender-expressive capacity. While it is outside the scope of this study, the accuracy of these predictions could be tested in further research.

6. Conclusion

Beginning in 1975 with the publication of Robin Lakoff's classic book *Language and Woman's Place*, the relationship between language and gender has been an important topic in sociolinguistics. This thesis contributes to the study of language and gender in a few main areas. First, the pragmatic functions of SFPs *o* 喔, *la* 啦, *ma* 嘛, *ei/ye* 欸/耶 in Taiwan Mandarin have been analyzed in detail. This analysis enabled a discussion of the gendered usage of these particles. Next, a corpus analysis has yielded empirical data on the usage rates of these SFPs by male and female speakers of Taiwan Mandarin across different confrontational contexts. Finally, a relationship has been proposed between the parameters of intonation and SFPs and the expression of gender in language. Let us review the key findings in each of these areas.

This paper proposes that SFPs *o* 喔, *la* 啦, *ma* 嘛, *ei/ye* 欸/耶 all have gendered uses i.e. senses that are favored by speakers of a certain gender, in this case female speakers. 喔 can express the speaker's interest in the receiver's well-being and give the receiver options in response to an assertion. 啦 can express playful annoyance or have a softening and harmony-preserving function. 嘛 can lend an air of whininess or acknowledge the receiver's prior knowledge as a form of respect. 欸/耶 can express playful surprise in flirtatious contexts and also indicate cooperation with a question that cannot be answered. The diversity of these gendered functions is a testament to the expressive power of SFPs and their importance to gender expression in Taiwan Mandarin.

This paper's study of the NCCU corpus has calculated the frequency of SFPs across the variables of speaker gender, receiver gender, and nature of speaker-receiver

relationship for four SFPs in Taiwan Mandarin. The results have provided clear evidence in support of gendered usage for each of the particles tested: female speakers used the particles at 30 to 80 percent higher rates than male speakers. It was found that male speakers used 欸/耶 with a female romantic partner at a rate about twice as high as with others. This finding supports the hypothesis that SFP use is in part determined by the relationship of the speaker to the receiver and the expectations for men and women in different social contexts.

This study upholds that SFPs and intonation are different values of a cross-linguistic parameter, and that the prominence of one feature in a language is inversely related to the prominence of the other. Based on the similarity of gendered use of intonation in English and gendered use of SFPs in Cantonese and other SFP-languages, it is proposed that linguistic expressions of gender have a deep relationship with the SFP-intonation spectrum. The ways in which gender is expressed in a language depend on how the language utilizes these linguistic systems. It has been proposed that languages that primarily rely on intonation will not exhibit gendered SFPs and languages that primarily rely on SFPs will not exhibit gendered intonation. Languages that possess both features may exhibit gendered usages of both.

This thesis also contributes to corpus linguistics research by providing a discussion of the challenges of data collection from transcriptions of oral language data. Many of these issues are especially relevant to Chinese and other languages for which sentence segmentation presents a particular challenge. But the majority of the discussion has applications to corpus studies based on conversation data in any language. Furthermore, the thesis discusses the characteristics of an ideal corpus for this kind of

research and the ways in which corpus formatting can improve the accuracy of data collection.

The challenges present in the data collection of this study included orthographic irregularity, which often results from the inconsistency in orthographic conventions of different conversation transcribers. The presence of code switching in the corpus also hindered the accuracy of character-frequency calculations and led to the removal from consideration of conversations that contained much data in a language other than Taiwan Mandarin. A youth bias of the analyzed conversations resulted from the neglect of these data. The problem of homograph disambiguation limited the efficacy of automated techniques that are unable to perform comprehensive syntactic analysis. And finally, the lack of sentential segmentation limited the ability to reliably locate the sentence-final position.

Further research can proceed in a variety of directions. Within the scope of Taiwan Mandarin, subsequent studies can analyze more SFPs from the language, such as *hou* 齣, *nie* 捏, *lei* 咧 etc. Studies with a larger sample size can draw more meaningful conclusions about the effect of conversational context on SFP use. Fieldwork that can manipulate the conversational context while keeping the speaker constant can also mitigate the confounding effect of individual variation. Corpora with formatting more conducive to a study of SFPs can be analyzed to test the results of this inquiry. There are also multiple cross-linguistic lines of research that could deepen understanding of the SFP-intonation spectrum and gendered use of these expressive features. Additional languages could be analyzed to see whether gendered use in such languages conforms with the typological expectations proposed in this study. A comparative analysis of

Taiwan Mandarin and Standard Mandarin may prove especially enlightening, as the languages are closely related yet differ in a few important ways. Further research could also identify “pairings” between intonational patterns and SFPs in the vein of Wakefield (2010) and (2016).

As gender increasingly becomes a subject of discussion throughout society, the ways in which gender manifests itself in language cannot be ignored. While most of us may not consider it on a daily basis, the patterns of our speech are clear reflections of our personal identities and the societal pressures we face. And while commonalities can be found, the elements of speech that carry gendered meaning can also differ dramatically across languages. Certainly, an understanding of gendered speech can teach us not only about those we speak with, but also the societies we live in.

References

- Ainsworth, J. E. (1993). In different register: The pragmatics of powerlessness in police interrogation. *Yale Law Journal*, 103(2), 259-322.
- Benwell, B., & Stokoe, E. (2006). *Discourse and identity*. Edinburgh: Edinburgh University Press.
- Bouavichith, D. (2017). Cross-contextual consistency of /s/ length and spectral quality in gay men's speech. *The Journal of the Acoustical Society of America*, 142(4), 2583.
- Cartei, V., Cowles, H., Reby, D., & Mathevon, N. (2012). Spontaneous Voice Gender Imitation Abilities in Adult Speakers. *PLoS ONE* 7(2): E31353.
- Cartei V., Reby D. (2013). Effect of Formant Frequency Spacing on Perceived Gender in Pre-Pubertal Children's Voices. *PLoS ONE*, 8(12): E81022.
- Chan, M. K-M. (1999). Sentence-final particles in Cantonese: A gender-linked survey and study. *Eleventh North American Conference on Chinese Linguistics (NACCL-11)*, 87-101.
- Chao, Y-R. (1932). A preliminary study of English intonation (with American variants) and its Chinese equivalents. *Bulletin of the Institute of History and Philology of the Academia Sinica, (The Ts'ai Yuan P'ei Anniversary Volume: Supplementary Volume I)*, 104-156.
- Cheung, K. (1986). The phonology of present day Cantonese. Doctoral dissertation, University College, London.
- Chui, K. (2018). The NCCU Corpus of Spoken Taiwan Mandarin. Retrieved from <https://spokentaiwanmandarin.nccu.edu.tw>.

- Chui, K., & Lai, H.- L. (2008). The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics*, 6(2), 119-144.
- Dabney, N. (2014). The influence of discourse particles on listeners' perception of speakers in Mandarin. Master's thesis, Indiana University.
- Davenport, E. (2018). A sociophonetic analysis of gay male speech stereotypes in Buenos Aires, Argentina. *The Journal of the Acoustical Society of America*, 143(3), 1970.
- Huang, K. (2013). A Study of Neutral-tone Syllables in Taiwan Mandarin. *Dissertation Abstracts International, A: The Humanities and Social Sciences*, 74(4).
- Jiang, H-L. (2011). Gender difference in English intonation. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 974-977.
- Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, 132(2), 1050-1060.
- Lakoff, R. (1972). Language in Context. *Language*, 48(4), 907-927.
- Lakoff, R., & Bucholtz, M. (2004). *Language and woman's place: text and commentaries* (2nd ed., Studies in language and gender). New York: Oxford University Press.
- Lin, C-H. (2014). Utterance-final particles in Taiwan Mandarin: Contact, context and core functions. Doctoral dissertation, Leiden University.
- Linneman, T. (2013). Gender in Jeopardy: Intonation Variation on a Television Game Show. *Gender & Society*, 27(1), 82-105.
- McConnell-Ginet, S. (1978). Intonation in a Man's World. *Signs*, 3(3), 541-559.

- Ministry of Education, Republic of China (2010). 各年齡層之閩南語使用率 [Usage rate of Min Nan by age group]. 本土語言調查報告 [*Homeland language survey report*]. Retrieved from <https://mhi.moe.edu.tw/newsList.jsp?ID=5&ID2=33>
- Shih, Y. H. (1984). 從社會語言學觀點探討中文男女兩性語言的差異 [A sociolinguistic study of gender differences in male and female's language in Chinese]. *教學與研究* [*Teaching and Research*], 6, 207-229.
- Starr, R. (2011). Variation in Affective Sentence-Final Particle Use and Transcription on Taiwanese Mandarin Television Dramas. *Texas Linguistics Forum vol. 54*, 196-197.
- Tsai, I-N. (2008). Projecting the Unanticipatory: The Mandarin Particle Ei and its Projectability in Daily Conversation. *Twentieth North American Conference on Chinese Linguistics (NACCL-20)*, 1023–1040.
- Tseng, C-C. (2004). Prosodic Properties of Intonation in Two Major Varieties of Mandarin Chinese: Mainland China vs. Taiwan. *International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages*. Beijing.
- Uyeno, T. (1971). A study of Japanese modality: a performative analysis of sentence particles. Doctoral dissertation, University of Michigan.
- [The Voice Stylist]. (2019). Video Six - Intonation Work for Male to Female Transgender Voice [Video File]. Retrieved from <https://www.youtube.com/watch?v=xw84tbVURRU>
- Wakefield, J. (2010). The English equivalents of Cantonese sentence-final particles: A contrastive analysis. Doctoral dissertation, The Hong Kong Polytechnic University.

- Wakefield, J. (2016). Sentence-final Particles and Intonation: Two Forms of the Same Thing. *Speech Prosody 2016*.
- Wamsley, J. (2019). Gendered Usage of Sentence-Final Particles in Mandarin Chinese. *IULC Working Papers Online*, 19(1).
- Wu, W. (2013). Cantonese Prosody: Sentence-final Particles and Prosodic Focus. Doctoral dissertation, University College London.
- Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. *Proc. 17th International Conference of Phonetic Sciences (ICPhS)*, Hong Kong, 152-155.
- Yueh, H-I. (2013). The Tactic of the Weak: A Critical Analysis of Feminine Persuasion in Taiwan. *Dissertation Abstracts International, A: The Humanities and Social Sciences*, 74(1).