A Comparative Study of Sentence-Final Particles Acquisition in Monolingual, Bilingual, and Trilingual Cantonese-Speaking Children: A Corpus-Driven Approach

by

Kassey Chang

A Thesis Submitted to
Columbia University
in Partial Fulfilment of the Requirements for
the degree of Bachelor of Arts in Linguistics

May 2024

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisors, Prof. Meredith Landman, Prof. John McWhorter, Prof. Ross Perlin, and Prof. William Foley, for their invaluable guidance and for pointing me in the right direction. A special thank you to Prof. Maggie Mai, who sparked the idea for this thesis during a lunch conversation.

I am also grateful to Prof. Alex Fang, who introduced me to the world of corpus linguistics, which has become my favorite subfield. The knowledge you imparted forms the fundamental methodology of this thesis.

I am also thankful to the Class of 2024 linguistics majors for providing valuable feedback.

I would like to thank my parents for introducing me to Cantonese and bilingualism.

A heartfelt thank you to Wael and my dear friends for the constant support during the thesis research and writing process.

Finally, this thesis would not have been possible without the incredible children participants and their parents who provided the data.

A Comparative Study of Sentence-Final Particles Acquisition in Monolingual, Bilingual, and Trilingual Cantonese-Speaking Children: A Corpus-Driven Approach

by Kassey Chang

Program of Linguistics

Columbia University

Abstract

Sentence-final particles (SFPs) play a crucial role in expressing speakers' attitudes and intentions in Cantonese, making them integral to the language's pragmatics. Despite their importance, the acquisition of SFPs by Cantonese-speaking children has not been extensively studied, particularly in multilingual contexts. This study compares SFP frequencies and usage patterns across monolingual, bilingual (Cantonese-English), and trilingual (Cantonese-English-Mandarin) children to identify early versus late-acquired SFPs and determine influential factors such as language exposure, dominance, and cross-linguistic effects. The findings reveal that assertive particles are among the first acquired SFPs across all language backgrounds. However, bilingual children do not acquire SFPs as well as their trilingual counterparts, showing a slower acquisition rate and the non-acquisition of certain particles. This difference can be attributed to the lack of direct SFP cognates between Cantonese and English and the limited cross-linguistic facilitation from English. In contrast, the presence of SFPs in Mandarin, despite the lack of direct cognates, appears to provide some facilitative effects for trilingual children acquiring Cantonese. These results demonstrate the importance of considering cross-language influence and typological distance when examining SFP acquisition in multilingual contexts. The absence of direct SFP cognates between Cantonese and English, as well as the structural differences between the two languages, may hinder the acquisition of certain particles in bilingual children. The outcomes of this research enhance our understanding of SFP acquisition in Cantonese and contribute to the broader knowledge of language development in multilingual settings. By providing quantitative insights into the developmental progression of SFP acquisition across different language backgrounds and examining the role of language distance in shaping acquisition patterns, this study sets a foundation for future research on the acquisition of pragmatic markers in multilingual contexts and informs pedagogical strategies to support Cantonese language learners.

CHAPTER 1

INTRODUCTION

Sentence-final particles (SFPs), also known as utterance-final particles, are integral to Cantonese conversation, as well as to other East and Southeast Asian languages. These monosyllabic morphemes, often positioned at the end of sentences or utterances, function as pragmatic markers. They are instrumental in conveying the speaker's attitudes, emotions, intentions, and other aspects of the speech act. The absence of SFPs in utterances can render them unnatural to native speakers, highlighting their importance in effective communication and natural speech (Leung, 2016). The complexity of particle systems in these languages constitutes one of the "hallmarks of natural conversation" (Luke, 1990). Moreover, second language (L2) learners often face considerable challenges in fully acquiring SFPs, owing to their intricate semantic-pragmatic nuances. Thus, the accurate use of SFPs is a crucial aspect of the competence of a Cantonese speaker and is often used as a marker of native-like proficiency.

The significance of SFPs has driven scholars to investigate their role across various linguistic domains, including pragmatics, discourse, syntax, semantics, and phonology, as evidenced by the works of Luke (1990), Matthews and Yip (1994; 2013), and Sybesma and Li (2007). These studies have sparked ongoing debates regarding the specific content and application of each particle. A major challenge in the linguistic description of SFPs involves addressing the functional diversity of individual particles, which complicates the identification of core functions for SFPs. The flexibility and subtlety in function illustrate the challenges in pinpointing whether specific meanings are semantically encoded within a SFP or are effects deriving from the utterance's context or SFPs' pragmatic meanings, such as changing illocutionary force, indicating an implicature or expectation, modifying the tone of an utterance, indicating epistemic modality, or expressing affection (Hancil et al., 2015).

Despite these extensive studies, there remains a notable gap in research concerning the acquisition of SFPs by Cantonese-speaking children. Previous research has extensively documented children's acquisition of content words, particularly nouns and verbs, as observed in Hao et al. (2008). These words are regarded as concrete and quantifiable elements in the lexicon. In contrast, the acquisition of SFPs has often been overlooked, primarily because they do not directly impact a speaker's fluency. Nonetheless, SFPs are crucial for semantic,

pragmatic, and communicative competence. Therefore, proficient use of SFPs is indicative of a child's comprehensive mastery of communicative Cantonese across various pragmatic contexts.

The current study is designed to address a series of research questions pertaining to the acquisition of Cantonese SFPs. The primary objective is to analyze the frequencies and usage patterns of SFPs among various groups of Cantonese-speaking children. These groups include monolingual, bilingual, and trilingual individuals.

In Hong Kong, where Cantonese is predominantly spoken, there is a notable presence of Cantonese-English bilinguals and Cantonese-Mandarin-English trilinguals. This linguistic diversity is a result of the socio-historical contexts, including the British occupation from 1828 to 1997 and the subsequent handover to Mainland China. Factors such as heavy language contact and immigration from China have contributed to the rise of Cantonese-Mandarin bilingualism. Consequently, apart from monolinguals, bilingual and trilingual populations are increasingly significant in Hong Kong. At the same time, bilingual and trilingual education is also encouraged by the "biliterate and trilingual"[1] policy implemented by the Hong Kong government since 1997(Wang & Kirkpatrick, 2013). By comparing SFP usage across these demographic and language groups, the study seeks to uncover developmental patterns in SFP acquisition and the potential impact of language background on the development of pragmatic competence. This comparative analysis contributes to our understanding of the influence of cross-linguistic factors on SFP acquisition.

Additionally, this study aims to investigate the acquisition timeline of various SFPs, assessing whether certain particles are acquired earlier or later during language development. This aspect of the research will provide insights into the developmental trajectory of SFPs in Cantonese-speaking children. Lastly, the study will explore factors influencing SFP acquisition, including language exposure, such as parental input, and cross-linguistic influence, as in how the linguistic features in the co-acquired languages facilitates or hinders the acquisition of Cantonese SFPs. Through a thorough analysis of these aspects, the study intends to enhance our understanding of how Cantonese-speaking children acquire SFPs, contributing valuable knowledge to the fields of language acquisition.

[1]The "biliterate and trilingual" policy has been adopted by the Hong Kong government since 1997. It acts as a benchmark in the curriculum design among elementary schools (grade 1 to 6). The aim of this language policy is to cultivate students to become biliterate in written English and written Traditional Chinese, and trilingual in English, Cantonese and Mandarin. However, every school implements trilingual education with different approaches.

CHAPTER 2

LITERATURE REVIEW

2.1 Sentence-Final Particles in Cantonese

Cantonese encompasses approximately 30 core sentence-final particles (SFPs), as identified by Kwok (1984). However, when considering the potential for combinations and clusters of particles, the actual count may exceed 100. Cantonese SFPs are generally categorized into six primary functional groups, as identified by Matthews and Yip (1994; 2013) (refer to Table 2.1). It is important to note that these categories are not strictly exclusive to their designated functions. Some SFPs carry multiple functions and can be categorized into more than one domain. Foundational studies by scholars have underscored the multifaceted nature of SFPs, elucidating their capacity to transform declaratives into interrogatives, moderate imperatives, facilitate reported speech, signal surprise, and denote statements as noteworthy.

Table 2.1: Cantonese SFPs Categorized by Function (Matthews & Yip, 1994; 2013)

Function	SFPs
Question	$aa4, ha2, ho2, le2, me1, le1/ne1, waa2$
Assertive	$aak3/aa3, ge3, gaa3, ge2, laa3, laak3$
Imperative and persuasive	$aa1, laa1$
Evidential	$aa1maa3, gwa3, lo1, lok5, wo3, wo5$
Exclamatory and affective	$bo3, ze1, zek1$
Adverbial	$sin1, tim1$

Further expanding upon this framework, Sybesma and Li (2007) introduced the concept of Minimal Meaningful Units (MMUs) in Cantonese (see Table 2.2). These MMUs pinpoint the minimal semantic units within the initials, rhymes, and tones of SFPs. For instance, the initial *l* in $laa1$ and $lo1$ signifies the realization of a state, whereas the rhyme *o* in $lo1$ and $wo3$ marks noteworthiness. Additionally, the first tone may indicate a "forward-looking" aspect and a "hearer-orientation." The combination and mixture of these MMUs creates a nuanced semantic profile for each SFP.

Based on the MMUs, Sybesma and Li (2007) propose the assignment of MMUs in a CP structure at different levels (see Figure 2.1).

Table 2.2: List of Cantonese MMUs in SFPs (Sybesma & Li, 2007)

Initials	$\varnothing, b, g, h, l, m, n, z$
Rhymes	aa, e, o
Tones	$1^{(55;(53))}, 2^{(35)}, 3^{(33)}, 4^{(21;11)}, 5^{(13)}$
Codas	$\varnothing, -k$

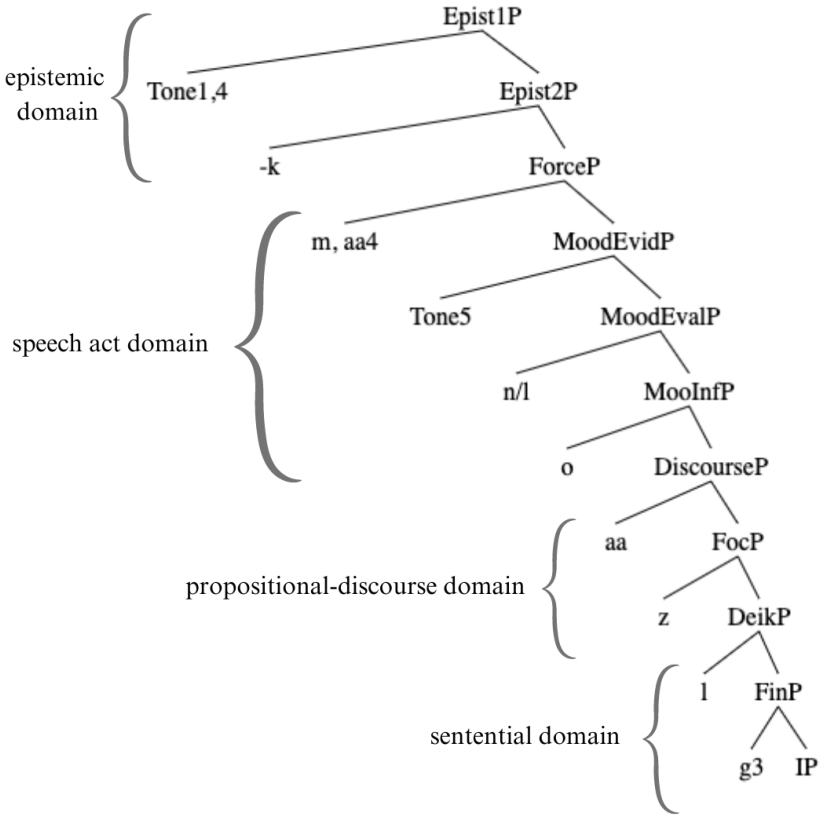


Figure 2.1: MMUs of SFPs in CP domain

The MMUs are ordered in the structure from bottom to top, with *g*3 occupying the lowest position in the C-domain, specifically in FinP. *aa* performs a discourse function and is placed in DiscourseP. The *o* MMU is associated with MoodInformativeP, *n*/*l* with MoodEvaluativeP, and tone 5 with MoodEvidentialityP. The *m* and *aa*4 MMUs, used for yes/no-questions, are assigned to ForceP. The tones 1 and 4, representing speaker/hearer orientation, are assigned to EpistemicP. The proposed structure provides evidence for distinguishing more than two layers, echoing Fung's (2000) comment that SFPs operate in different domains: the sentential, propositional, discourse, speech act, and epistemic domains. These domains are neatly separated in the structure, with FinP and DeikP representing the sentential domain, FocP and DiscourseP representing the propositional-discourse domain, MoodP and ForceP for the speech act domain, and the EpistemicPs for the epistemic domain.

Common SFPs in Cantonese

*lo*1 is a common SFP in Cantonese, typically categorized as an evidential and focal expression. This particle plays a multifaceted role in communication: it denotes obviousness or certainty, marks epistemic commitment to the truth of propositions, and can be used to invite agreement, cooperation, or sympathy (Fung, 2000; Kwok, 1984; Matthews & Yip, 1994; Tang, 2015; Chor, 2018). For instance, Luke (1990) demonstrates how *lo*1 assists in negotiating communication endings, while Law (1990) and Fung (2000) highlight its evaluative sense and association with negative attitudes such as reluctance, irony, and even irritation. The use of *lo*1 in a conversation from Hong Kong Cantonese Corpus (HKCanCor) (Luke & Wong, 2015) about a pet rabbit's behavior illustrates its occurrence in contexts involving intense epistemic processes:

(1) A: *nei5 mai6 bei2 zing3soeng4 ge3 je5 keoi5 mo4 ngaa4 lo1*
2SG ADV give normal POSS thing 3SG grind teeth SFP
"You could just give it something normal to grind its teeth, right?"

B: *hai6 lo1*
COP SFP
"Sure."

A: *hai6 lo1, mai6 ngo5 go2 bou6 daan1ce1 lo1*
COP SFP, ADV 1SG that CL bicycle SFP
"Yes, which is that bicycle of mine."

The first *lo*1 was used by speaker A to express obviousness and seek agreement. Speaker B responsed with *lo*1 to intensify her stance of agreement, while speaker A employed *lo*1 sarcastically, indicating irony.

*laa*3, another common SFP in Cantonese, functions as an assertive marker, emphasizing a point of current relevance or a change of state. Scholars like Tang (2015), Sybesma & Li (2007), and Matthews & Yip (2013) view *laa*3 as resembling the Mandarin *le*, marking a realization of physical or epistemic state, indicating that something is now the case that was not previously known or apparent. *laa*3 is used in sentences that are plain, neutral, and factual. For example:

(2) *aa3 cou4 laa3 , ngo5 fan3 m4 dou2*
too noisy SFP 1SG sleep not PRT

"I can't sleep, it's too noisy."

(3) *lok6 jyu5 laa3*
fall rain SFP
"It's raining now."

These sentences with *laa3* demonstrate its role in marking a change or realization, i.e., the realization of the noise or rain in the current environment, often used in straightforward factual statements.

Romanization of Cantonese

The Jyutping romanization system, developed by the Linguistic Society of Hong Kong (LSHK) in 1993 (https://lshk.org/jyutping-scheme/), is employed throughout this thesis for transcribing Cantonese. Although various romanization systems are currently used, Jyutping has gained wide recognition in academic circles for its accurate representation of Cantonese pronunciation, particularly its tones, without relying on the complex diacritics typical of IPA symbols.

As the most contemporary and widely used system, Jyutping ensures consistent and standardized transcription. It is the preferred romanization system for most Cantonese corpora and language data, including those utilized in this thesis. Moreover, Jyutping addresses some of the challenges encountered in other romanization schemes, such as Yale Romanization, commonly used in the United States, or the Sidney Lau system, by utilizing the Latin-based alphabet and morphemes. This relative simplicity makes Jyutping more accessible to a wider audience, including researchers and readers less familiar with Cantonese phonetics, without sacrificing linguistic precision.

2.2 Early acquisition of Cantonese

Language input, particularly parental or caregivers' input, plays a crucial role in child language acquisition and cannot be overlooked. Shek and Hui (2011) suggest that the type of language children regularly hear in their environment influences their daily speech and, consequently, the early lexicon, based on their study on Cantonese-speaking toddlers in Hong Kong. This factor is so significant that it has become an independent variable in various early language acquisition studies such as Mai and Yip's (2022). Others have looked into the acquisition in a

psycholinguistic perspective: The relationship between a child's pre-linguistic cognitive structures and the specific language they hear in the early stages of language acquisition is complex and bidirectional, as proposed by Choi and Gopnik (1993; 1995). This interaction highlights the importance of considering both cognitive development and linguistic input when studying early language acquisition in Cantonese.

2.2.1 Early acquisition of Particles

The early acquisition of SFPs reveals crucial developmental milestones in child language learning. According to Lee and Law (2001), Cantonese-speaking children as young as 1;8 to 1;9 have been observed using at least three different SFPs in their utterances. However, mastery of the nuanced meanings, especially those indicating epistemic modality, is typically not achieved until around the age of 6. This finding highlights a significant developmental trajectory in the understanding and usage of SFPs, underscoring the complexity involved in their full acquisition. In the broader context, studies in languages like German, Dutch, and Korean have extensively explored the early acquisition of modal and grammatical particles. The closest parallels to Cantonese are found in studies of Japanese SFPs. In these studies, Shirai et al. (2000) and Fujimoto (2008) utilized mother-child interaction data and MLU measurements to delineate the acquisition timeline of SFPs in Japanese. Shirai's (2000) research proposed a three-stage acquisition process:

1. comprehension of attention in the current space and time,
2. referencing the past, and
3. contrasting different situations, such as in expressions of complaint.

Conversely, Fujimoto's (2008) study, though focused on particles other than SFPs, outlined a syntactically inclined stage progression with overlaps, including

1. particles associated with illocutionary acts,
2. monosyllabic particles connected to noun phrases (NPs),
3. monosyllabic particles functioning as NPs, and
4. disyllabic semantically complex particles.

Both studies examined the correlation of SFP acquisition with

1. the parent's MLU, indicative of parental input;
2. the child's MLU, reflective of the child's linguistic and cognitive development; and

3. age.

Fujimoto (2008) observed that SFP acquisition began between an MLU of 1.00-1.49 and was considered "complete" at an MLU of 3.00.

These studies are crucial for comprehending how the early acquisition of SFPs can be segmented into distinct stages. Although Japanese SFPs possess a smaller inventory and display functions, syntactic positions, and semantic content that are distinct from those of Cantonese SFPs, they may still serve as a model for the current study, potentially revealing parallels in acquisition stages and sequences. Furthermore, these studies offer a methodological framework for analyzing Mean Length of Utterance (MLU) and parental input statistics, as well as their correlation with children's linguistic development.

2.3 Mean Length of Utterance (MLU) and Lexical Development

2.3.1 Mean Length of Utterance (MLU)

Mean Length of Utterance (MLU) is a widely used measure in child language research, serving as an indicator of linguistic development and proficiency. It is calculated by dividing the total number of morphemes in a sample of utterances by the total number of utterances. It indicates the grammatical complexity of one's speech.

2.3.2 MLU and Age Correlation

MLU has been found to correlate strongly with age, making it a valuable tool for assessing and predicting children's language development (Miller & Chapman, 1981). The strong correlation between age and MLU is particularly relevant to the current study. Kwong (1990) proposed that, on average, three-year-old Cantonese-speaking children have a vocabulary of 3.4 words. In a more recent study, Shek and Hui (2011) found that the MLU of three-year-old Cantonese-speaking children was 5.72. This correlation helps establish a developmental framework for SFP acquisition relative to age and linguistic maturity. As children grow older and their MLU increases, their usage of SFPs is expected to become more sophisticated and diverse. The trend of increasing MLU with age parallels the expected progression in SFP usage in children's speech, offering a measurable tool for assessing and predicting children's development in SFP acquisition. By examining the onset age and MLU associated with SFP

usage, researchers can explore the interplay between MLU and SFP proficiency, shedding light on the developmental patterns of SFP acquisition among Cantonese-speaking children.

2.4 Bilingual and Trilingual Acquisition

Multilingualism adds a layer of complexity to language acquisition, as bilingual and trilingual speakers often display patterns in their use of grammatical constructions and pragmatic markers that diverge from monolingual norms. Factors such as reduced or unbalanced language input, cross-linguistic interference, and the sociolinguistic environment surrounding the learner have been cited as contributing to these differences.

2.4.1 Language Input

Acquiring two or more languages simultaneously in bilingual or multilingual contexts presents distinct challenges compared to monolingual language acquisition. A crucial factor influencing language development is the quantity and quality of input each language receives. The child's linguistic environment, including language input (the speech a child hears during daily life, or so-called child-directed speech, see MacWhinney, 2000), plays a critical role in language development. In light of this theory, a variety of empirical research, including both longitudinal and cross-sectional studies, has been conducted to investigate the characteristics of children's early speech (language output), the linguistic input they receive, and the relationship between these two variables. The amount of language input a child receives has been positively correlated with their language development, including vocabulary size and grammatical complexity. In multilingual contexts, the role of language input becomes even more complex, as children are exposed to multiple languages with varying degrees of exposure and quality. In bilingual situations, children's input space is divided between two languages, resulting in reduced exposure to each language compared to their monolingual peers (Paradis & Genesee, 1996). This input reduction can significantly impact the rate and trajectory of language development. The effects of input reduction are even more pronounced in trilingual contexts, where the input directed to the child is split three ways (Mai & Yip, 2022). Children growing up with three languages simultaneously are likely to experience a decrease in both the quantity and quality of input in each language compared to monolingual children (Chevalier, 2015). Studies examining input variations in bilingual development have consistently found that the proportion of input in a language relative to the child's total input, also known as "input proportion," strongly predicts the developmental rate. Hoff et al. (2012) investigated

input-outcome relations among Spanish-English bilinguals and English monolingual children in the U.S. from a lexical and grammatical perspective. They discovered that the proportion of English exposure in the child's total input positively correlated with every outcome measure. Notably, even the group of bilinguals who received balanced input in both languages (50-60%) still differed from the monolingual group in both vocabulary scores and MLU in English.

Moreover, there is considerable variation in the language input young children receive from their caretakers (Hart & Risley, 1995). Child-directed speech, which is characterized by simplified vocabulary, exaggerated intonation, and repetition, has been found to facilitate language acquisition. The relative proportion of input in each language, as well as the consistency and richness of the input, can significantly influence the child's language outcomes (Hoff et al., 2012). As such, it is crucial to consider the characteristics of language input when examining language development in multilingual children, as it can provide valuable insights into the factors that shape their linguistic trajectories.

2.4.2 Cross-linguistic Influence in Bilingualism

Cross-linguistic influence, also known as language transfer, can be found in various multilingual contexts, such as bilingualism, diglossia, second language acquisition, and foreign language learning. The typological distance between the languages being acquired can significantly influence the acquisition process and outcomes. Blom et al. (2020) examined the impact of cross-language distance by comparing bilingual groups with a small and a large language distance (Bilingual Close group and Bilingual Distant group) with monolingual controls. The bilingual Distant group had lower receptive vocabulary outcomes than the bilingual Close and monolingual groups, while no difference emerged between the monolinguals and the bilingual Close group. This suggests that bilingual children whose languages provide ample opportunities for transfer and sharing knowledge do not experience receptive vocabulary delays.

Cross-linguistic transfer can manifest in various ways, such as the borrowing of grammatical structures, the overextension of linguistic rules, or the underuse of certain elements. For example, in the domain of vocabulary, bilingual children may take longer to name pictures when both languages are highly active, as illustrated by research showing longer response times in mixed compared to single language conditions (Kohnert et al., 1999). This demonstrates that interaction between bilingual children's two languages can lead to interference, causing naming delays in online performance. An instance of syntactic transfer is

that prenominal relative clauses, which are common in Cantonese, may be transferred to English in Cantonese-English bilingual children (Yip & Matthews, 2006).

However, cross-language interaction can also facilitate bilingual children's lexical retrieval if words in their two languages are cognates (Sheng et al., 2016). Cognates[2] are words in different languages that have a shared origin and resemble each other semantically and phonologically. The cognate facilitation effect refers to the phenomenon that bilingual performance on various vocabulary tasks is enhanced for cross-linguistic cognates as opposed to non-cognates (Sheng et al., 2016). Cross-linguistic cognates are typically semantically related but are not necessarily translation equivalents. The existence of cognates in two given languages usually indicates that these languages are historically related (Richards & Schmidt, 2002). Linguistically closer languages or languages within the same family have more phonologically form-similar and form-identical cognates, facilitating early lexical acquisition [3] (Bosch et al., 2014; Sheng et al., 2016).

Language Distance and Typological Differences

The direction and extent of transfer may depend on factors such as the typological distance between languages, the child's proficiency in each language, and the input patterns they are exposed to. The role of language distance in cross-linguistic influence can be demonstrated with Levenshtein distances, which is a measure of language distance based on two languages' phonological and lexical differences (Bakker et al., 2009; Wichmann et al., 2016). Table 2.3 presents the Levenshtein distances between four language pairs: English-Cantonese, English-Mandarin, Cantonese-Mandarin, and French-Spanish (as a control pair). Higher values indicate greater dissimilarity between the languages.

Table 2.3: Levenshtein Distances Between Language Pairs

Language Pair	Levenshtein Distance
English-Cantonese	97.35
English-Mandarin	102.30
Cantonese-Mandarin	81.00
French-Spanish	84.03

The Levenshtein distances reveal that English is more distant from both Cantonese and

[2] An example of a pair of cognates are *bed* in English and *Bett* in German.

[3] For the languages in this study, it is evident that there are more form-similar and form-identical cognates between Cantonese and Mandarin, as they are both Sinitic languages sharing a subset of common lexicon. Most form-similar cognates Between Cantonese and English, and Mandarin and English come from borrowing from English, e.g. $dik1si2$ in Cantonese and "taxi" in English.

Mandarin compared to the distance between Cantonese and Mandarin. This suggests that the typological differences between English and the two Chinese languages are more substantial than those between Cantonese and Mandarin. Brown and Iwasaki (2006) suggested that the Chinese language grouping can be compared to the Romance languages. Interestingly, the distance between French and Spanish (LD=84.03) is even higher than that between Cantonese and Mandarin (LD=81.00). However, it is important to note that Levenshtein distances are based on the Swadesh lists, which focus on the phonological and lexical aspects of a language. The perceived distance between languages can be very different due to sociolinguistic and cultural factors. In the case of Cantonese and Mandarin, the heavy contact, shared culture, and political unity may lead to a perception of closer proximity compared to French and Spanish, despite the slightly higher Levenshtein distance[4].

Still, it is crucial to consider the fundamental typological differences between the co-acquired languages to determine whether they can facilitate or hinder each other's development (Tardif et al., 2008). English is known for its noun bias, while Cantonese and Mandarin exhibit a verb bias in the early lexicon. Tardif and her colleagues insist that "Nouns are not always learned before verbs" in Chinese (Tardif, 1997). Sandhofer, Smith, and Luo (2000) also found that Mandarin-speaking parents tend to use more verb tokens than nouns, while English-speaking parents tend to use more nouns than verbs. Structurally, Cantonese and Mandarin are pro-drop languages in which verbs are obligatory and rarely omitted (Tardif et al., 1997). Tse (1993) found that among the 18 types of syntactic structures in Cantonese, 16 require verbs. In contrast, English is a non-pro-drop language that requires more noun phrases and common nouns in sentence formation. In formal contexts, overt subject and object noun phrases are usually obligatory, and nouns may occupy a perceptually salient position at the end of a sentence. Furthermore, English verbs may change in sound during declension, and irregular verbs can need more time to be fully acquired. In contrast, verbs are more constant without declension and can often occur alone as a complete sentence in Chinese languages (Tse, 1993).

Although Cantonese and Mandarin share many typological similarities, they are mutually unintelligible, with distinct phonological inventories and lexical differences. According to

[4]Learners evaluate linguistic proximity—and consequently decide on the application of L1 knowledge—guided by their perceptions of surface typology, a phenomenon Kellerman (1979) identifies as "psychotypology." In scenarios where L2 is perceived to be closely akin to L1, learners might deploy their L1 knowledge without restraint, which can lead to error patterns heavily influenced by L1 (manifesting as overgeneralization). On the contrary, perceiving a substantial linguistic gap may deter learners from drawing upon their L1 knowledge, even when the target and their native or previously learned languages bear resemblances. While the concept of psychotypology offers intriguing insights, the present study lacks the requisite data to explore this aspect further.

some estimates, vocabulary differences range from 10–50% depending on the source and style of material (Snow, 2004). Their grammatical structures are very similar, but with variations in obligatoriness of marking and the specific ways in which aspect, number, and negatives are marked, as well as word order differences in some syntactic constructions.

These distances and dissimilarities have implications for cross-linguistic influence and acquisition of particles. Brown and Iwasaki's study (2006) examined the acquisition of Korean particles by Japanese-Korean speakers (Bilingual Close) and English-Korean speakers (Bilingual Distant). They found that the Japanese learners possessed an advantage in some areas, particularly accurate use of delimiters, as their native language also overtly marks categories such as nominative and accusative case. This provides them with a conceptual head-start in determining the grammatical role of sentence constituents. However, in the use of postpositions, the English speakers performed just as well as the Japanese speakers. Here, Japanese-Korean surface similarity (i.e., use of postpositions) turned out to be no more of an advantage than having an L1 (i.e., English) where the same grammatical relationships are marked in a different way. This study has shown that linguistic proximity (and learner perceptions of it) plays an important role in the acquisition process and that a learner's dominant language can play both an impeding and facilitative role, rather than primarily being an impediment.

In another early acquisition study on particles, Yan et al. (2023) suggests that cross-linguistic influence has led to an advantage of co-activation of the Mandarin SFP *le* and the English adverb "already" in bilingual representation. It is highly possible that the co-activation of L1 when processing L2 in bilinguals can either boost or compete with L2. This finding has implications for bilinguals and trilinguals in this thesis. If their languages have similar representations of SFPs, it is possible that the co-activation processing effect may also apply. In other words, the presence of similar SFP structures across the languages of a multilingual individual could potentially facilitate the acquisition and processing of these particles.

Potential Form-Similar Cognate SFPs in Cantonese and Mandarin

While Cantonese and Mandarin are mutually unintelligible, they share many typological similarities due to their common ancestral roots. As a result, these two languages may have SFPs that are form-similar cognates, having similar phonological and/or orthographic forms. The presence of these SFPs in Cantonese and Mandarin may have significant implications for cross-linguistic influence in trilingual children. Trilinguals may benefit from the presence of

these shared linguistic elements, as they can leverage their knowledge of SFPs in one language to support their understanding and use of similar particles in the other language, but may also lead to negative transfer or interference, particularly when the particles have subtle differences in meaning or usage between the two languages.

Table 2.4: Potential Form-Similar Cognates of SFPs in Cantonese and Mandarin (Tang, 2015)

Cantonese SFPs	**Mandarin SFPs**	**Category**	**Note**
*sin*1	*xian*1	Adverbial	Mandarin: Before verbs
*laa*3	*le*	Assertive	Mandarin: Perfective
*ne*1	*ne*	Question	Mandarin: focal
*gwaa*3	*ba*	Question (speculative)	Mandarin: more functions

Note: *sin*1 and *xian*1, as well as *ne*1 and *ne* have same orthographical forms, while the other two pairs do not. Although these SFPs share phonological similarities and some overlapping functions, they are not necessarily direct translation equivalents. The exact usage and pragmatic nuances may differ between Cantonese and Mandarin.

CHAPTER 3

METHODOLOGY

3.1 Corpora

Three longitudinal spoken corpora from the CHILDES (Child Language Data Exchange System) TalkBank database (MacWhinney, 2000) will be utilized in this study, each offering distinct perspectives on adult-child interactions, such as playing with toys and reading books. The large-scale monolingual Lee-Wong-Leung Corpus provides comprehensive insights into the linguistic development of Cantonese-speaking children in a singular language environment, making it essential for understanding the natural trajectory of SFP acquisition without the influence of additional languages.

In contrast, the Yip-Matthews Corpus, another extensive corpus, captures the linguistic interplay between Cantonese and English, presenting a bilingual environment that could potentially influence the use and comprehension of SFPs. This corpus employed the one parent-one language approach, where each parent consistently communicates with the child in their respective language. This approach aims to provide a balanced exposure to both languages. Similarly, the trilingual Leo Corpus, while focused longitudinally on a single child, is the only existing early trilingual corpus involving Cantonese, Mandarin, and English. This corpus not only utilized the one caregiver-one language approach but also incorporated a one day-one language practice, where each day was dedicated to a specific language.

Table 3.1: Overview of the Corpora Used in the Study

Corpora	Lee-Wong-Leung Corpus	Yip-Matthews Corpus	Leo Corpus
Production years	1991-94	2000-04	2016-18
Total no. of words	963,578	655,923	83,153
Total no. of words produced by children	221,753	133,473	15,507
No. of participant(s)	8	9	1
No. of samples	137	156	18
Age range	1;05;22-3;08;09	1;03;10-4;06;07	1;06;21-2;11;21

All of the CHILDES TalkBank data are in the CHAT format. The CLAN (Computerized

Language Analysis) program (MacWhinney, 2000) is used throughout the study to view, extract, annotate and obtain the speech data and SFPs instances.

It is important to note that the samples in the Yip-Matthews Corpus and Leo Corpus have shown imbalanced acquisition in children, with Cantonese being the dominant language for most participants[5]. In the Yip-Matthews Corpus, although the dominant language was not static, the majority of children in the study had higher proficiency in Cantonese by the age of three. Only two children deviated from this pattern: one appeared to have balanced proficiency, while the other was English-dominant. Similarly, in the Leo Corpus, the input from caregivers was imbalanced, with Mandarin accounting for 54%, Cantonese 26%, and English 20%. Consequently, Leo's language outcomes reflected this imbalance. While he performed on par with monolingual Cantonese children, his Mandarin and English proficiency were not at the same level as their monolingual counterparts, lacking grammatical complexity and lexical diversity. Even though the children in these samples are classified as bilingual or trilingual, they have clearly developed a dominant language, likely influenced by factors such as input proportion, geographical location, and overall language exposure.

A key methodological consideration for this study is the age range of participants across the corpora. To ensure a fair and meaningful comparison, the study focuses on the age range common to all three corpora, spanning from 1 year, 10 months to 3 years. This age range aligns with the definition of early child acquisition, particularly in the context of trilingualism. As noted by Unsworth (2013), infant or early trilingualism refers to a child's acquisition of three languages before the age of three. Fujimoto (2008) also observed that SFPs are fully acquired by L1 Japanese children within this age range. By concentrating on this specific developmental stage, the study aims to capture the critical period of SFP acquisition and minimize the influence of later linguistic experiences.

3.2 Data Analysis

This study adopts a nuanced analytical approach to investigate the acquisition of SFPs in Cantonese-speaking children, including monolingual, bilingual, and trilingual speakers. The

[5]In bilingual acquisition literature, the concept of language dominance is used to describe situations where one of a child's languages is more advanced or developing faster than the other. This dominance can influence the acquisition process, with the dominant language potentially influencing the non-dominant language through cross-linguistic transfer (Yip & Matthews, 2007). Yip and Matthews (2009) demonstrated that MLU differentials can indicate language dominance. This finding suggests that Cantonese-dominant children may exhibit more advanced SFP usage in Cantonese compared to their other languages. The tendency for the language with a higher MLU to influence the one with a lower MLU has important implications for understanding the acquisition of SFPs in multilingual Cantonese-speaking children.

analysis begins by calculating the frequencies of all SFPs spoken by children in the three groups. Ten SFPs are selected based on three criteria: 1) most frequently used across all three corpora, 2) expected to be acquired at an early stage, and 3) representing all six categories of SFPs identified by Yip and Matthews (1994; 2013). These targeted SFPs form the basis for the subsequent analyses.

The data analysis includes the following steps:

1. Comparing the frequencies of SFPs in conversations between adults and child-directed speech by parents across all three corpora to understand the differences in SFP usage patterns in child-directed speech.
2. Examining the MLU of children across the three corpora to assess their lexical and language development.
3. Identifying the first SFPs used by children from the three groups to determine which SFPs are acquired earliest.
4. Analyzing the frequencies of SFPs spoken by adults and children in adult-child conversations. One child from each corpus with similar MLU onsets is selected for a detailed examination of their SFP frequencies at different ages.
5. Determining the onset of using each targeted SFP in each group.

3.2.1 Statistical Analysis

Following the data analysis, correlation analyses are conducted for each group to investigate the relationships between age and MLU, age and SFPs used, MLU and SFPs, parental MLU and SFPs, and input SFPs and child SFPs. Linear regression is also performed to assess the correlations within each group.

Statistical analyses are done on Python, with the special help from PyCantonese (Lee et al., 2022) and PyLangAcq (Lee et al, 2016) libraries. These libraries were designed specially for assisting in the analysis of language data in a Python environment and contain built-in Cantonese parsing as well as calculations for developmental measures like MLU and TTR. Correlation analyses are done on the JASP software interface.

The findings from these analyses will provide insights into the developmental trajectories of SFP acquisition among monolingual, bilingual, and trilingual Cantonese-speaking children, as well as the factors influencing their acquisition, such as age, MLU, and parental input.

It is expected that children from all three language backgrounds (monolingual, bilingual,

and trilingual) will acquire most of the frequently used SFPs by the age of three. However, variations in the rate and pattern of acquisition are expected due to differences in language input and individual progress. While the overall language development of the children is expected to be on track, it is highly possible that SFP acquisition may not necessarily follow the same developmental trajectory as other lexical categories, such as nouns and verbs. SFPs may show a latency in acquisition compared to content words, as they serve more complex pragmatic functions and require a higher level of communicative competence.

Regarding the influence of language background, monolingual children are expected to be the least affected by cross-linguistic influence, as they are exposed to a singular linguistic system. In contrast, bilingual and trilingual children may exhibit different patterns of SFP acquisition due to the potential influence of their other co-acquired languages.

Lastly, parental input is hypothesized to be the most significant factor correlating with children's use of SFPs. The frequency, consistency, and quality of SFP input in child-directed speech are expected to have a direct impact on children's acquisition and production of these particles.

CHAPTER 4

FINDINGS

This chapter presents the findings of the study, focusing on the acquisition and usage of SFPs in Cantonese-speaking children from monolingual, bilingual, and trilingual backgrounds. The analysis begins by examining the frequency of SFPs in child speech across the three language groups, and the lexical development of the children, as measured by MLU, and the initial use of SFPs by children in each group. Furthermore, the chapter explores the relationships between various factors influencing SFP acquisition, such as age, MLU, parental input, and SFP usage, through correlation analyses. Linear regression models are also employed to assess the predictive power of these factors in explaining SFP acquisition and usage.

4.1 Frequencies of SFPs in Child Speech

The following graphs illustrates the SFP frequencies in child speech across the three corpora:

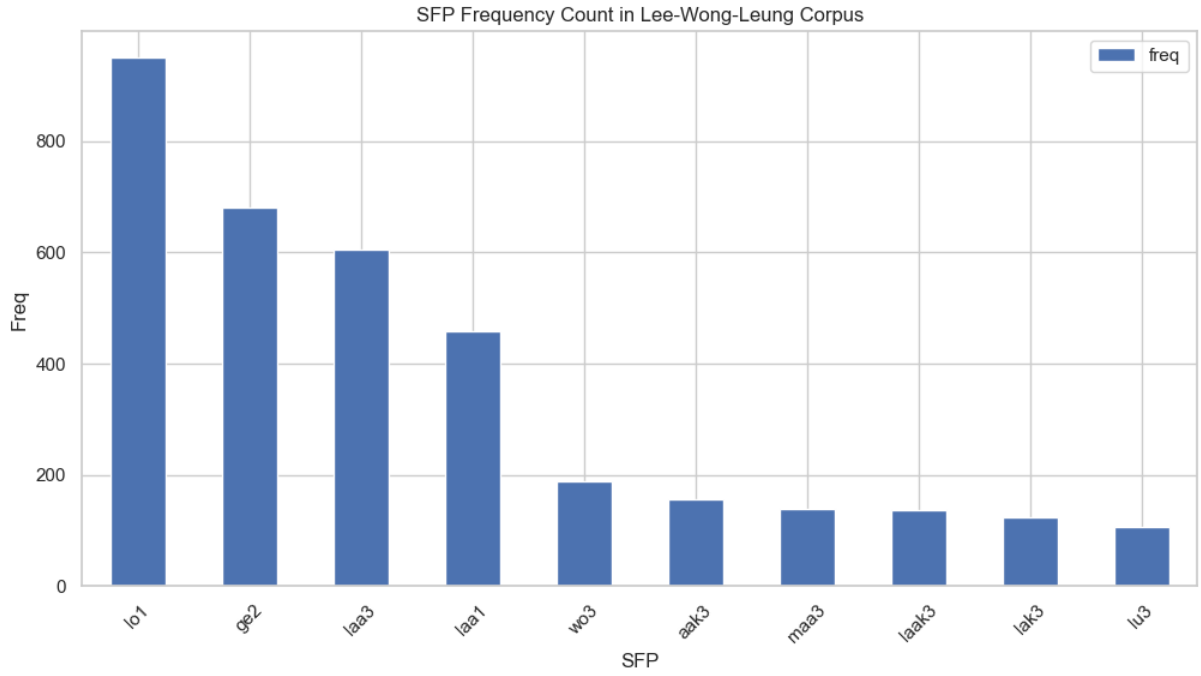


Figure 4.1: SFP frequency Count in Lee-Wong-Leung Corpus

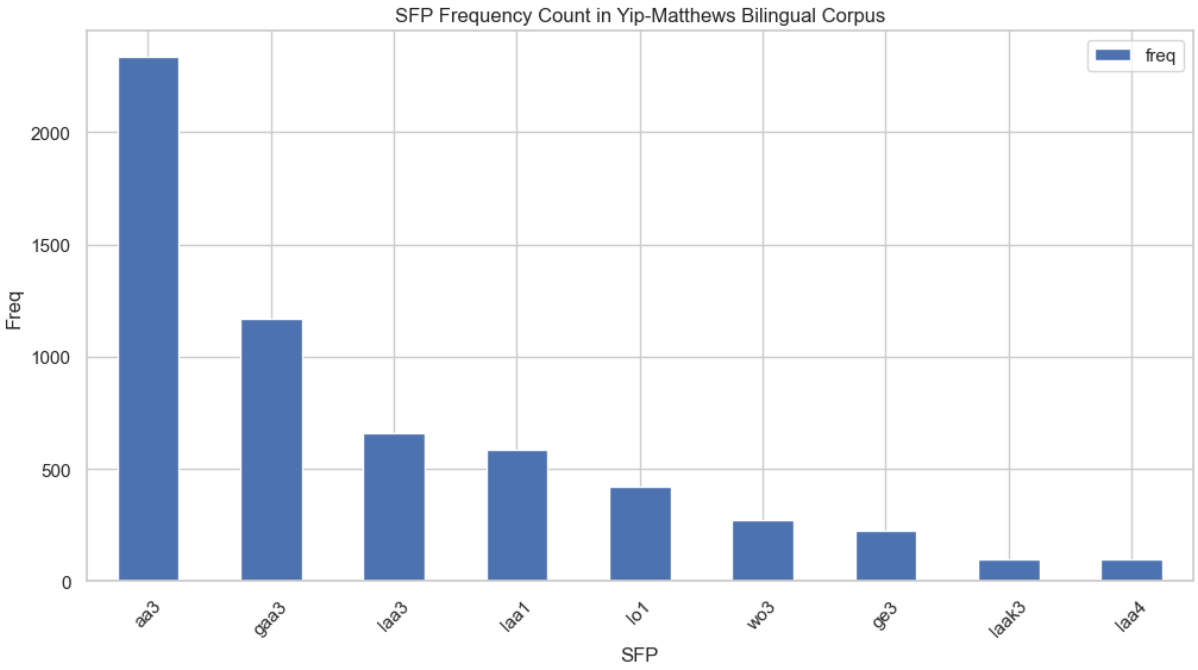


Figure 4.2: SFP frequency Count in Yip-Matthews Bilingual Corpus

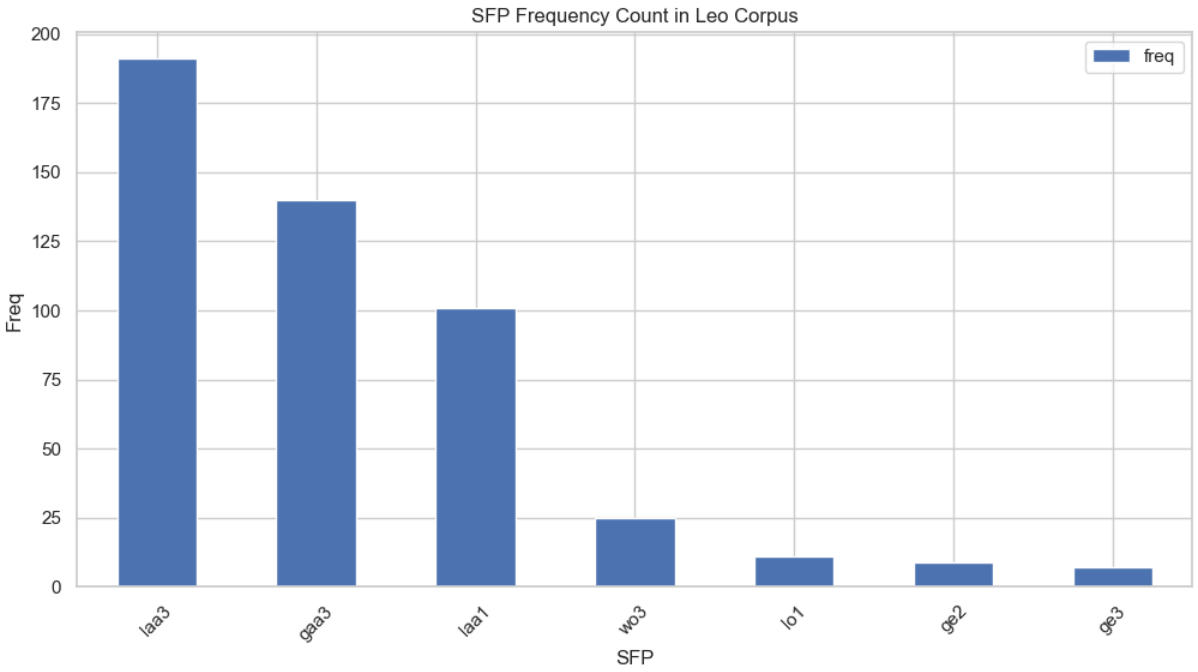


Figure 4.3: SFP frequency Count in Leo Corpus

Figures 4.1, 4.2, and 4.3 illustrate the frequencies of each SFP used by children in the monolingual, bilingual, and trilingual corpora, respectively. The graphs reveal that assertive particles, such as *laa*3, *aa*3, and *gaa*3, consistently appear among the most frequently used SFPs across all three groups.

To further investigate the acquisition and usage patterns of SFPs, ten particles that are most frequent across the three corpora, are acquired in an early stage (refer to Table 4.4-4.6) and span the six categories identified by Yip & Matthews (1994; 2013) have been selected for more detailed analysis. These targeted SFPs are as follows:

- Assertive: *aa*3, *laa*3, *ge*2, *gaa*3
- Imperative: *laa*1
- Question: *me*1
- Adverbial: *sin*1
- Evidential: *lo*1, *wo*3
- Exclamatory: *bo*3

Table 4.1 presents the frequency of SFPs in child speech as a percentage of total utterances across monolingual, bilingual, and trilingual children at different age points. In the monolingual group, there is a consistent increase in the frequency of SFPs used by children as they grow older. Similarly, the bilingual group demonstrates an increase in SFP frequency over time, although at a slightly slower pace compared to their monolingual counterparts. The trilingual child exhibits a unique pattern of SFP use. There is a substantial increase in the frequency of SFPs from 1.28% at 1;10;00 to 6.84% at 2;05;00, indicating a rapid acquisition of these particles in the early stages of language development.

Table 4.1: Frequency of SFPs in child speech (%)

	Age (Year; Month; Day)			
	1;10;00	2;01;00	2;05;00	2;09;00
Monolingual (LWL)	3.04	3.95	5.68	6.39
Bilingual (YM)	1.37	3.13	4.13	4.32
Trilingual (Leo)	1.28	1.85	6.84	5.78

4.2 Comparison with Adult Speech

Table 4.2 shows that in adult-to-adult conversations from the Hong Kong Cantonese Corpus (HKCanCor) data (Luke & Wong, 2015), the frequency of SFPs is approximately 11.4%. In

Table 4.2: Frequency of SFPs in conversation between adults (%)

Monolingual (HKCanCor)	Adult
Frequency (%)	11.4

Table 4.3: Frequency of SFPs in child-directed speech (%)

	Age (Year; Month; Day)			
	1;10;00	2;01;00	2;05;00	2;09;00
Monolingual (LWL)	1.12	4.92	4.06	5.06
Bilingual (YM)	3.04	3.95	5.68	6.39
Trilingual (Leo)	4.92	6.16	6.52	6.05

contrast, Table 4.3 presents the frequency of SFPs in child-directed speech across different age groups for monolingual, bilingual, and trilingual children. It is evident that in child-directed speech, adults tend to use more SFPs as children get older. This trend is observed across all three language backgrounds. However, the frequency of SFPs in child-directed speech varies among the different language groups. In the samples of bilingual and trilingual children, adults consistently used more SFPs compared to the monolingual group. This suggests that, in general, bilingual and trilingual children in the studied samples received more SFP input from adults compared to their monolingual counterparts.

4.3 MLU and Lexical Development

4.3.1 MLU of Children

Figure 4.4 illustrates the MLU of monolingual children (on average), bilingual children (on average), and a trilingual child, which serves as a measure of their language development. At the initial stage, bilingual children exhibited a slightly higher MLU compared to their monolingual and trilingual counterparts. However, all three groups demonstrated a gradual progression in their MLU as they aged. By the age of 3, all groups had reached an MLU greater than 3.5.

4.3.2 Initial SFP Use in Children

Tables 4.4, 4.5, and 4.6 present the first SFPs used by each child in the monolingual, bilingual, and trilingual groups, respectively, along with the age and MLU at which these SFPs were observed. One notable finding is that the assertive particle *aa*3 is among the first acquired SFPs in all three groups. This suggests that *aa*3 may be one of the easiest and most

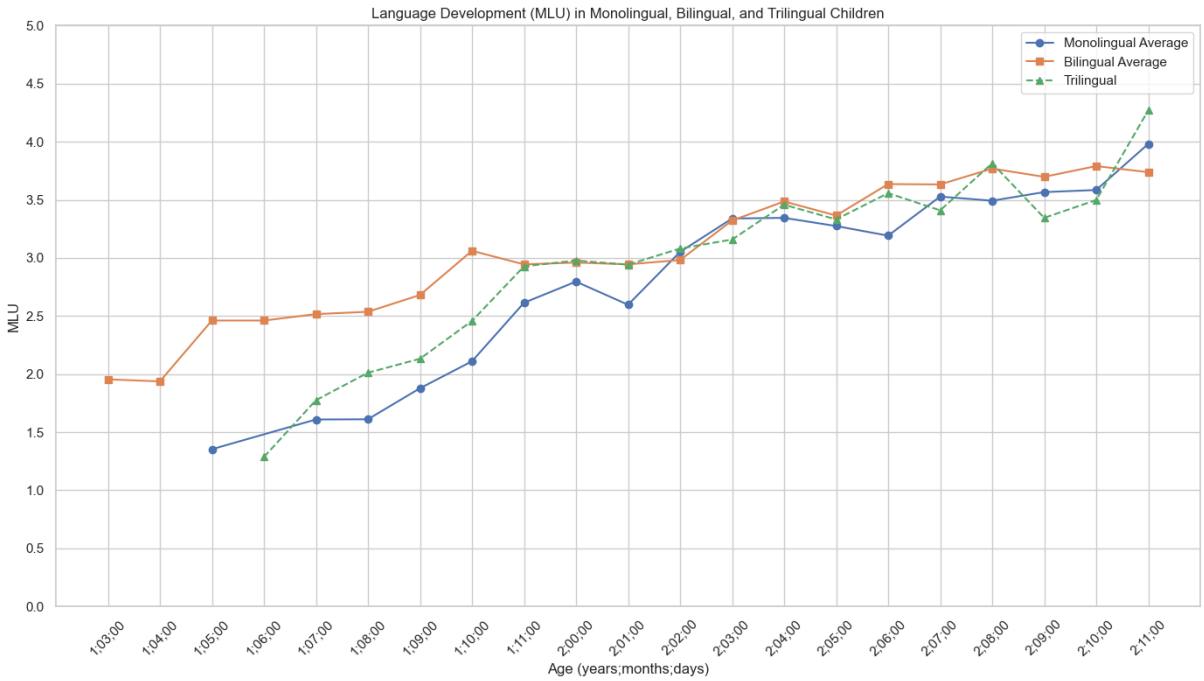


Figure 4.4: Language Development (MLU) in Monolingual, Bilingual and Trilingual Children

fundamental SFPs.

Interestingly, the bilingual group appears to start using SFPs earlier compared to their monolingual counterparts. The trilingual child, Leo, used his first SFP earlier than the average age of first SFP use in the monolingual group.

Table 4.4: Initial SFP Use in Monolingual Children

Child	**Age**	**MLU**	**SFPs**
ccc	1;11;00	1.39	aa3, aa1
cgk	1;11;01	2.00	laa3, aa3, sin1
ckt	1;07;03	1.586	aa3
mhz	1;08;00	1.602	aa3
hhc	N/A	N/A	N/A
lly	N/A	N/A	N/A
ltf	N/A	N/A	N/A
wbh	N/A	N/A	N/A

Note: Data for certain children have been excluded from the tables, as indicated by the "N/A" entries. These instances represent cases where the first speech data were collected after the age of two. By this stage, these children had likely acquired a broader range of SFPs, which could potentially misrepresent the initial acquisition phase targeted by this study. To ensure the fairness and accuracy of the data presented, only children whose first recorded use of SFPs occurred before the age of two are included.

Table 4.5: Initial SFP Use in Bilingual Children

Child	**Age**	**MLU**	**SFPs**
Alicia	1;03;10	1.34	bo3, aa3, lo1, ge3, gaa3
Charlotte	1;08;28	1.857	aa3, gaa3
Darren	1;07;23	1.542	aa3
Sophie	1;06;01	1.865	aa3, gaa3, lo1, laa3, laa1
Janet	N/A	N/A	N/A
Kasen	N/A	N/A	N/A
Llywelyn	N/A	N/A	N/A
Tim	N/A	N/A	N/A

Table 4.6: Initial SFP Use in Trilingual Children

Child	**Age**	**MLU**	**SFPs**
Leo	1;06;21	1.04	aa3

4.4 SFP Usage in Adult-Child Conversations

4.4.1 SFP acquisition timelines

To track the individual progress of SFP parental input and child output, one child from the monolingual (ckt) and bilingual (Darren) data was selected for further analysis. Tables 4.6, 4.7, and 4.8 present the frequency of ten SFPs used by the child and their parents, along with the correlation coefficients. The correlation is calculated for data from 1;10 through 2;10 (2;07 in monolingual due to end of data).

The results reveal inconsistencies in the correlation across all groups. However, a general trend emerges: children tend to acquire frequently used SFPs very early on. Interestingly, parental input of SFPs seems to have the largest effect on trilingual children, as evidenced by the higher correlation coefficients in Table 4.9 compared to the monolingual and bilingual data. In contrast, the bilingual child (Darren) shows lower correlations between parental input and child output of SFPs. It is important to note that some SFPs are not fully acquired by the children, even when their mothers frequently use those particles. For example, the exclamatory particle *bo*3 is rarely produced by the children across all groups, despite its presence in the parental input. Similarly, the particles $lo1$ and $laa1$ are not fully acquired by the bilingual child, despite their occurrence in the parental speech. The onset of SFP usage by the children does not appear to be solely dependent on the frequency of parental input.

[6]PAR: Parent/caregiver; CHI: Child; r: correlation coefficient

Table 4.7: Frequency of Adult Input and Child Output of SFPs in a Monolingual Context
Child: ckt

Category	**SFP**		**1;10**	**2;00**	**2;02**	**2;04**	**2;06**	**2;07**	**Total**	**r**[6]
Adverbial	sin1	PAR	39	96	80	61	42	55	373	-
		CHI	0	1	2	0	4	2	9	
Assertive	aa3	PAR	332	469	378	367	307	217	2070	0.75
		CHI	148	196	206	240	174	93	1057	
	laa3	PAR	134	94	129	97	126	123	703	0.20
		CHI	0	2	0	1	10	16	29	
	ge2	PAR	15	19	20	14	25	9	102	0.10
		CHI	0	0	0	0	6	5	11	
	gaa3	PAR	162	168	177	169	211	165	1052	0.51
		CHI	0	0	1	0	43	46	90	
Evidential	lo1	PAR	14	16	31	31	31	32	155	0.65
		CHI	0	0	16	0	35	24	75	
	wo3	PAR	61	74	69	35	41	59	339	-
		CHI	0	1	0	1	2	1	5	
Exclamatory	bo3	PAR	7	1	4	0	0	2	14	-
		CHI	0	0	0	0	1	0	1	
Imperative	laa1	PAR	97	125	126	65	85	95	593	-
		CHI	0	3	1	2	26	34	66	
Question	me1	PAR	79	103	60	95	80	84	501	0.03
		CHI	3	5	1	2	20	9	40	

Table 4.8: Frequency of Adult Input and Child Output of SFPs in a Bilingual Context
Child: Darren

Category	**SFP**		**1;10**	**2;00**	**2;03**	**2;04**	**2;06**	**2;07**	**2;10**	**Total**	**r**
Adverbial	sin1	PAR	42	42	31	23	35	42	52	267	0.18
		CHI	0	2	0	0	0	0	0	2	
Assertive	aa3	PAR	273	304	137	140	286	246	273	1659	0.12
		CHI	0	4	3	4	0	6	19	36	
	laa3	PAR	137	76	148	136	132	125	82	836	-
		CHI	0	0	1	0	0	0	1	2	
	ge2	PAR	14	20	14	13	23	14	11	109	-
		CHI	0	1	0	0	0	0	1	2	
	gaa3	PAR	70	57	73	39	75	78	64	456	0.51
		CHI	0	0	1	0	0	2	1	4	
Evidential	lo1	PAR	29	23	14	7	13	25	40	151	-
		CHI	0	0	0	0	0	0	0	0	
	wo3	PAR	78	73	159	111	149	85	48	703	-
		CHI	0	0	0	0	0	0	1	1	
Exclamatory	bo3	PAR	7	1	7	2	0	2	2	21	-
		CHI	0	0	0	0	0	0	0	0	
Imperative	laa1	PAR	54	50	34	22	30	56	44	290	-
		CHI	0	0	0	0	0	0	0	0	
Question	me1	PAR	84	64	66	94	126	130	83	647	0.33
		CHI	0	0	0	4	0	2	0	6	

Table 4.9: Frequency of Adult Input and Child Output of SFPs in a Trilingual Context
Child: Leo

Category	SFP		1;10	2;00	2;02	2;03	2;06	2;08	2;10	Total	r
Adverbial	sin1	PAR	22	9	7	15	27	18	19	117	0.76
		CHI	1	0	0	1	2	0	2	6	
Assertive	aa3	PAR	245	127	158	134	161	99	131	1055	-
		CHI	1	8	11	13	39	48	11	131	
	laa3	PAR	24	32	43	24	81	33	54	291	0.62
		CHI	3	8	6	24	30	17	27	115	
	ge2	PAR	3	7	5	8	7	6	7	43	0.10
		CHI	3	2	2	1	6	5	7	26	
	gaa3	PAR	68	42	69	16	21	20	14	250	-
		CHI	8	7	12	19	6	12	3	67	
Evidential	lo1	PAR	3	8	11	10	4	9	1	46	0.82
		CHI	1	1	3	2	0	1	0	8	
	wo3	PAR	18	19	17	22	24	23	18	141	0.75
		CHI	0	2	0	4	5	2	3	16	
Exclamatory	bo3	PAR	0	0	1	0	2	2	0	5	-
		CHI	0	0	0	0	0	0	0	0	
Imperative	laa1	PAR	16	14	29	50	48	63	64	284	0.75
		CHI	0	0	4	11	17	8	38	78	
Question	me1	PAR	44	33	35	12	4	5	12	145	0.86
		CHI	3	5	3	0	0	0	1	12	

4.4.2 Onset of SFP Usage

Table 4.10: Onset Age of SFPs Across Different Linguistic Groups

Category	SFP	Monolingual	Bilingual	Trilingual
Adverbial	sin1	2;00	2;00	1;10
Assertive	aa3	1;10	2;00	1;10
	laa3	2;00	2;03	1;10
	ge2	2;06	2;00	1;10
	gaa3	2;02	2;03	1;10
Evidential	lo1	2;02	N/A	1;10
	wo3	2;00	2;10	2;00
Exclamatory	bo3	2;06	N/A	N/A
Imperative	laa1	2;00	N/A	2;02
Question	me1	1;10	2;04	1;10

Table 4.10 presents the onset age of each of the ten selected SFPs across the monolingual, bilingual, and trilingual children. In the monolingual child (ckt), all ten SFPs were acquired by the age of 2;06. The assertive particle *aa*3 and the question particle *me*1 were the earliest to emerge, with an onset age of 1;10. The other SFPs were acquired gradually over the following months, with the exclamatory particle *bo*3 and the assertive particle *ge*2 being the last to

appear at 2;06. The bilingual child (Darren) exhibited a different pattern of SFP acquisition. Three of the ten SFPs (*lo*1, *bo*3, and *laa*1) were not acquired by the end of the observation period. Moreover, Darren had a later onset age for all SFPs compared to the monolingual child. In contrast, the trilingual child (Leo) demonstrated a remarkably early acquisition of seven out of ten SFPs by the age of 1;10.

4.5 Correlation Analysis

The correlation analysis and linear regression results provide valuable insights into the relationships between age, MLU, and SFP usage in monolingual, bilingual, and trilingual children.

For monolingual children, the correlation table and plots reveal strong positive correlations between age of child, MLU of child, and SFP percentage in child speech. Parent MLU and adult SFP percentage also show significant positive correlations with MLU of child and SFP usage. The linear regression model explains 46.8% of the variance in SFP percentage in child speech, with age, child MLU, and SFP percentage in child-directed speech (adult SFP percentage)[7] being significant predictors.

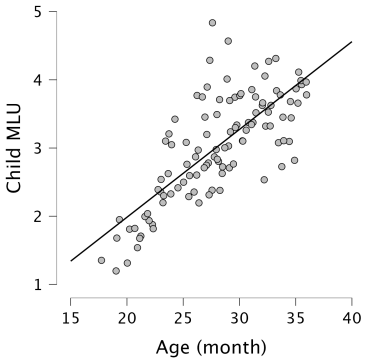
Table 4.11: Correlations Between Age, MLU, and SFP Usage in Monolingual Children

			Pearson		Spearman	
			r	p	rho	p
Age	-	Parent MLU	0.230**	0.008	0.236**	0.007
	-	Child MLU	0.752***	< .001	0.733***	< .001
	-	Child SFP Percentage	0.565***	< .001	0.571***	< .001
	-	Adult SFP Percentage	0.184*	0.029	0.087	0.186
Parent MLU	-	Child MLU	0.427***	< .001	0.457***	< .001
	-	Child SFP Percentage	0.211*	0.014	0.273**	0.002
	-	Adult SFP Percentage	0.258**	0.003	0.217*	0.012
Child MLU	-	Child SFP Percentage	0.599***	< .001	0.640***	< .001
	-	Adult SFP Percentage	0.303***	< .001	0.146	0.065
Child SFP Percentage	-	Adult SFP Percentage	0.430***	< .001	0.336***	< .001

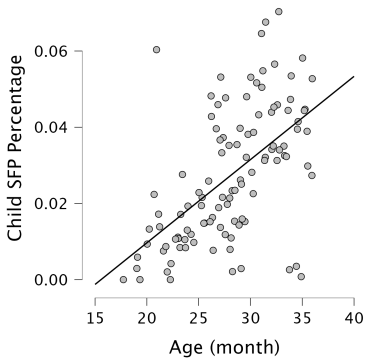
* p < .05, ** p < .01, *** p < .001, one-tailed

In bilingual children, age and MLU of child exhibit strong positive correlations with SFP percentage in child speech. However, MLU of parent shows no significant correlation with child language measures. The linear regression model accounts for 62.8% of the variance in

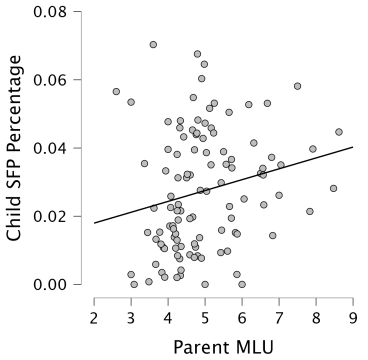
[7]The factor is named "adult SFP percentage/SFP percentage in child-directed speech" instead of parent SFP input because the analysis needed to account for adult input from investigators who were sources of language input (SFP input) as well.



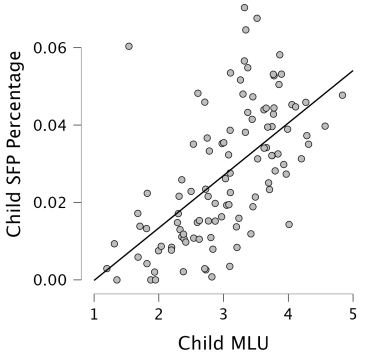
(a) Correlation between MLU and Age of Children



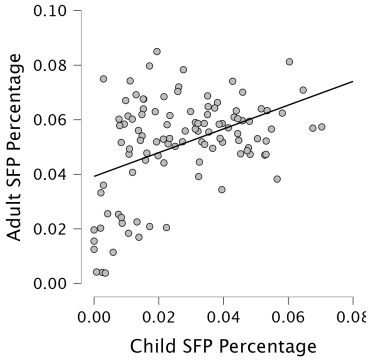
(b) Correlation between Age and SFP usage



(c) Correlation between Parent MLU and SFP usage by Children



(d) Correlation between MLU and SFP Usage by Children

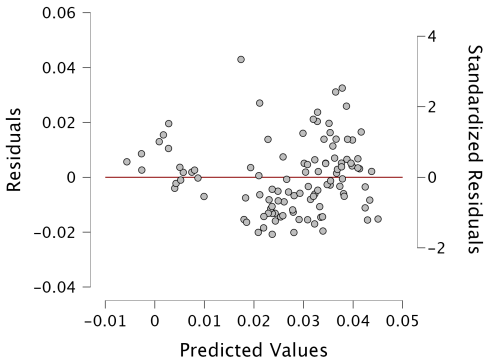


(e) Correlation between SFP input and output

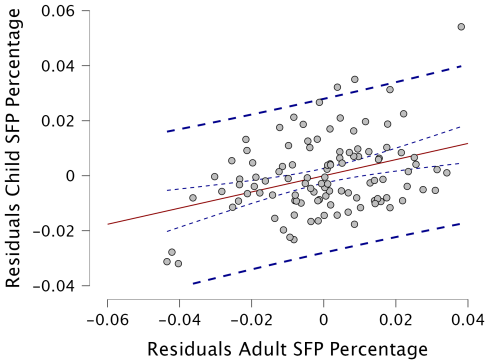
Figure 4.5: Correlation Plots for Monolingual Children

Table 4.12: Monolingual Model Summary - Child SFP Percentage

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.018
H_1	0.684	0.468	0.447	0.013



(a) Residuals vs. Predicted



(b) Partial Regression (Input and Output)

Figure 4.6: Linear Regression Plots for Monolingual Children

SFP percentage in child speech, with MLU of child and SFP percentage in child-directed speech being the most significant predictors.

Table 4.13: Correlations Between Age, MLU, and SFP Usage in Bilingual Children

			Pearson		Spearman	
			r	p	rho	p
Age	-	Parent MLU	−0.033	0.608	−0.036	0.617
	-	Child MLU	0.724***	< .001	0.721****	< .001
	-	Child SFP Percentage	0.322**	0.003	0.288**	0.008
	-	Adult SFP Percentage	0.275*	0.011	0.365***	< .001
Parent MLU	-	Child MLU	0.009	0.472	−0.002	0.505
	-	Child SFP Percentage	0.082	0.251	0.048	0.347
	-	Adult SFP Percentage	−0.053	0.669	−0.102	0.800
Child MLU	-	Child SFP Percentage	0.596***	< .001	0.482***	< .001
	-	Adult SFP Percentage	0.330**	0.003	0.207*	0.043
Child SFP Percentage	-	Adult SFP Percentage	0.649***	< .001	0.130	0.141

* p < .05, p < .01, * p < .001, one-tailed

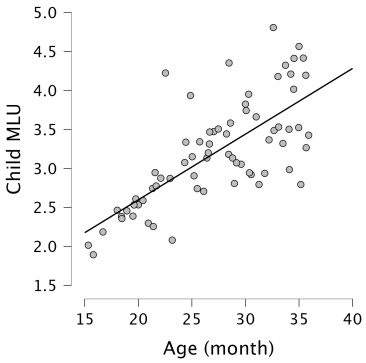
Table 4.14: Bilingual Model Summary - Child SFP Percentage

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.034
H_1	0.792	0.628	0.605	0.022

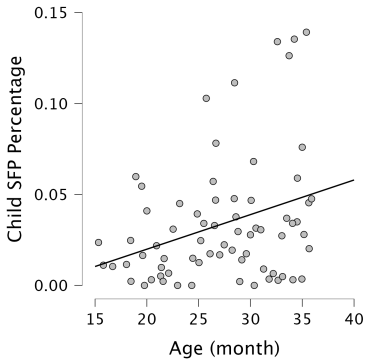
For the trilingual child, age, MLU of parent, and MLU of child demonstrate strong positive correlations with SFP output. The linear regression model explains 73.2% of the variance in SFP output, although individual predictors do not reach statistical significance, likely due to the small sample size.

The residuals vs. predicted plots for all three groups suggest that the linear regression models fit the data well, with no apparent patterns in the residuals. The partial regression plots highlight the positive relationships between adult SFP input and child SFP output, particularly in the monolingual and bilingual groups.

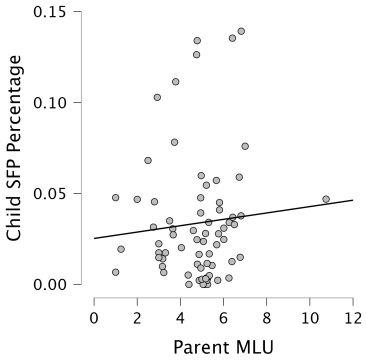
Overall, these findings underscore the importance of age, child language development (MLU), and adult SFP input in predicting child SFP usage across different language backgrounds. The results also highlight some differences between the groups, such as the varying influence of parent MLU on child language measures.



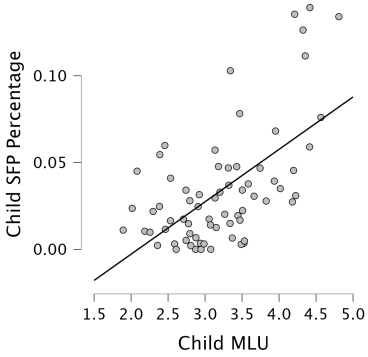
(a) Correlation between MLU and Age of Children



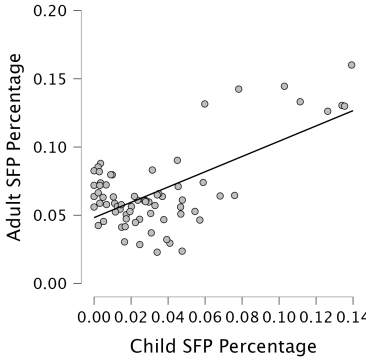
(b) Correlation between Age and SFP usage



(c) Correlation between Parent MLU and SFP usage by Children

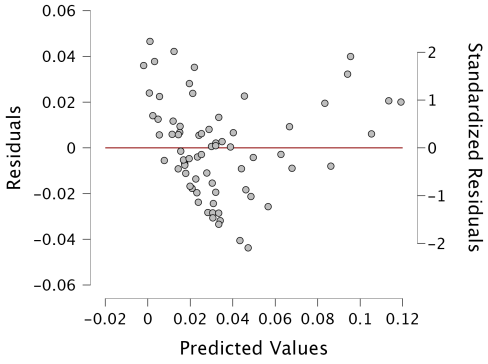


(d) Correlation between MLU and SFP Usage by Children

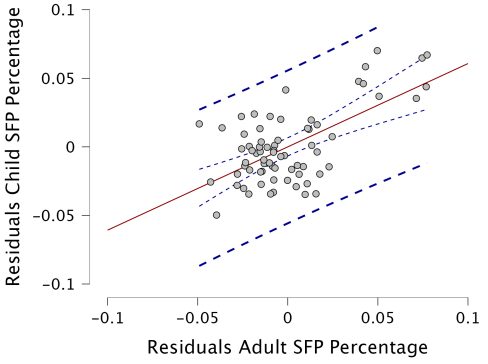


(e) Correlation between SFP input and output

Figure 4.7: Correlation Plots for Bilingual Children



(a) Residuals vs. Predicted



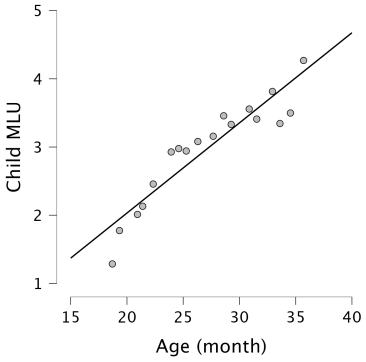
(b) Partial Regression (Input and Output)

Figure 4.8: Linear Regression Plots for Bilingual Children

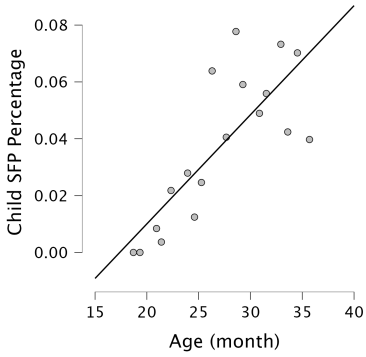
Table 4.15: Correlations Between Age, MLU, and SFP Usage in Trilingual Children

			Pearson		Spearman	
			r	p	rho	p
Age	-	Child MLU	0.929***	< .001	0.953***	< .001
	-	Parent MLU	0.764***	< .001	0.773***	< .001
	-	Child SFP Percentage	0.791***	< .001	0.779***	< .001
	-	Adult SFP Percentage	0.157	0.266	0.183	0.233
Child MLU	-	Parent MLU	0.897***	< .001	0.862***	< .001
	-	Child SFP Percentage	0.799***	< .001	0.818***	< .001
	-	Adult SFP Percentage	0.262	0.147	0.348	0.079
Parent MLU	-	Child SFP Percentage	0.778***	< .001	0.791***	< .001
	-	Adult SFP Percentage	0.374	0.063	0.515*	0.015
Child SFP Percentage	-	Adult SFP Percentage	0.389	0.056	0.399	0.051

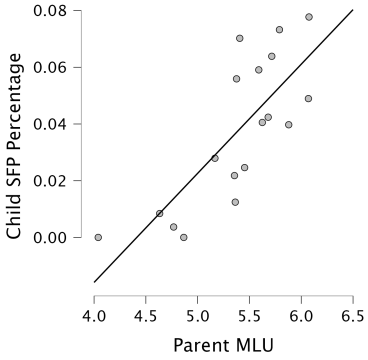
* p < .05, p < .01, * p < .001, one-tailed



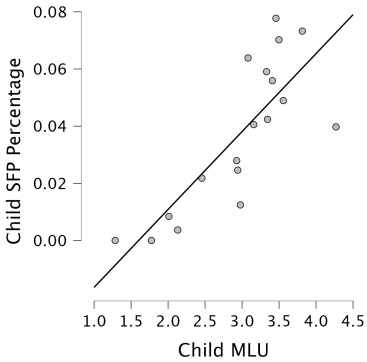
(a) Correlation between MLU and Age of Children



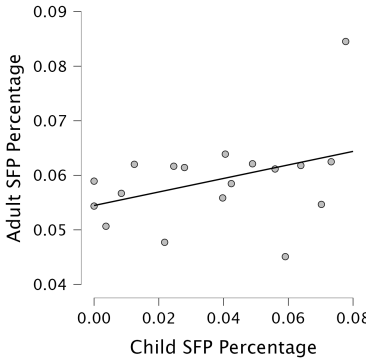
(b) Correlation between Age and SFP usage



(c) Correlation between Parent MLU and SFP usage by Children



(d) Correlation between MLU and SFP Usage by Children

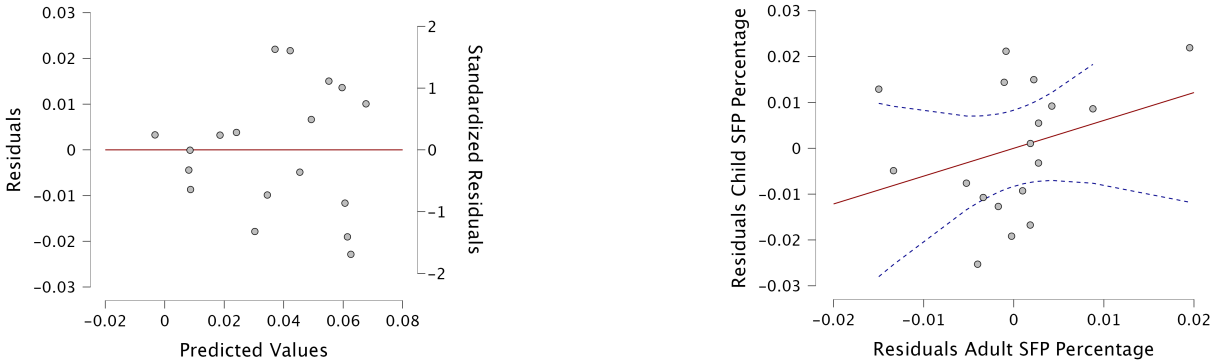


(e) Correlation between SFP input and output

Figure 4.9: Correlation Plots for Trilingual Children

Table 4.16: Trilingual Model Summary - Child SFP Percentage

Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	0.026
H_1	0.855	0.732	0.649	0.015



(a) Residuals vs. Predicted

(b) Partial Regression (Input and Output)

Figure 4.10: Linear Regression Plots for Trilingual Children

4.6 SFP Usage

In this section, we delve into the specific usage patterns of the ten SFPs studied in this research. By examining concrete examples from the corpus data, we can gain a more nuanced understanding of how monolingual, bilingual, and trilingual children employ these particles in their everyday speech.

4.6.1 Assertive SFPs (*aa*3, *laa*3, *ge*2, *gaa*3)

*aa*3 is the first SFP acquired in all language groups. It conveys a sense of finality (Matthews & Yip, 2013) and softens an utterance (Sybesma & Li, 2007). It can occur with various sentence types, such as declaratives (example 1), imperatives, exclamatives, and questions (example 2). The early acquisition of *aa*3 can be attributed to its commonness, flexibility, and phonetic easiness to pronounce (without a consonant onset). In both examples 1 and 2, *aa*3 softens the utterances, making them sound less abrupt and harsh.

(1) INV[8]: *nei5 zaa1 m4 zaa1 dou2 gaa3 loeng5 go3*
2SG hold NEG hold ASP SFP two CL
"Can you hold two of them?"

CHI: *zaa1 dou2 aa3*
hold ASP SFP
"I can hold them!"

(2) The investigator slowly pushes up the puppet's head .
CHI: *keoi5 hai2 bin1dou6 aa3*
3SG LOC where SFP
"Where is it?"

INV: *it1 hai2 dou6 aa3*
INTJ here SFP
"Here it is."

*laa*3 emphasizes a point of current relevance, marking a change of state or realization of a state. Sentences with laa3 are plain, neutral, and factual. In example 3, the child uses *laa*3 to assert the current state of the car being here and how it is now coming here, implying that it might not have been here previously.

(3) CHI: *ce1ce1 lei4 laa3*
vehicle come SFP
"The car is here!"

*ge*2 occurs in both interrogative and declarative sentences (example 4). In declaratives, *ge*2 is used for assertions of facts with a sense of uncertainty and reservation (Matthews & Yip, 2013), often followed by an overt or implicit "but"-sentence. In interrogatives, *ge*2 is interpreted as a "why"-question, even without an explicit "why" element. In example 4, the child's question with *ge*2 expresses puzzlement and seeks an explanation for the absence of a fireman among the toys, equivalent to "why is there no fireman?" in English.

(4) INV: *tai2 haa5 nei1dou6 jau5 mou5 siu1fong4jyun4*
look INTJ here have not-have fireman
"Look if there's fireman here."

[8]INV: Investigator

CHI: *mou5 siu1fong4jyun4 ge2*
not-have fireman SFP
"There's no fireman?"

*gaa*3 is an assertion of relevance marker (Sybesma & Li, 2007), conveying the meaning "it is a relevant fact that ... but I don't mind that you don't know or forgot" or serving as a reminder. In example 5, both the investigator's question and the child's answer use *gaa*3 to convey a sense of relevance. The child's statement with *gaa*3 marks the relevance of the current situation to him, stating that he was the one who pushed the toy.

(5) INV: *bin1go3 ngung2 dit3 gaa3*
who push fall SFP
"Who pushed it?"

CHI: *zeon3zeon3 ngung2 dit3 gaa3*
PROPN push fall SFP
"(Name of child) pushed it."

4.6.2 Question SFP (*me*1)

*me*1 is used for confirmation-seeking questions and questions expressing disbelief. In example 6, the child uses *me*1 after a copula, forming a question similar to "is it?" or "really?" in English, seeking confirmation from the investigator. When children are forming interrogative sentences at en early stage, it is found to be "easier" to acquire the interrogative construction of a tag question that adds interrogative SFPs than other constructions like the A-not-A structure[9] (Matthews and Yip, 1994; Tse & Li, 2011).

(6) INV: *hai2 hok6haau6 aa3*
LOC school SFP
"At school."

CHI: *hai6 me1*
COP SFP
"Really?"

INV: *hai6 aa3*

COP SFP

"Yeah."

4.6.3 Evidential SFPs (*lo*1, *wo*3)

*lo*1 suggests that what is said is self-evident and is used to invite agreement, cooperation, or sympathy. It can also be used to negotiate a settlement (Luke, 1990). In answers, *lo*1 is less factual and more evasive, giving an excuse rather than an explanation. In example 7, the child changes the final SFP from *gaa*3 to *lo*1 when repeating the character's name, expressing that it is self-evident that the character is Mr. Fan and possibly conveying a sense of impatience and sarcasm.

(7) CHI: *xxx sin1saang1 lai4 gaa3*

xxx[10] Mister SFP SFP

"That's Mister xxx."

INV: *me1 waa2*

what SFP

"What?"

CHI: *faan6 sin1saang1 lai4 lo1*

PROPN Mister SFP SFP

"That's Mister Fan!"

*wo*3 emphasizes a noteworthy or newsworthy piece of information, such as a surprising discovery or a reminder. In example 8, after counting the cups, the child restates "three cups *wo*3," using the SFP to mark the new information and recent discovery.

(8) The child is counting toy cups.

CHI: *jat1 ji6 saam1 . saam1 bui1 wo3*

one two three. three cups SFP

"One, two, three. There're three!"

[9] A-not-A question can be comparable with asking "Is A the case or not?" The construction is formed by repeating/reduplicating A, the verb or adjective, with the negative marker *m*4 inbetween For example, instead of *hai*6*me*1 in example 6, the child could say *hai*6*m*4*hai*6*gaa*3 (COP NEG COP SFP) to form a confirmation-seeking question and express a similar pragmatic content of doubt.

[10] Unintelligible words

4.6.4 Adverbial SFP (*sin*1)

*sin*1 behaves more like an adverb than other SFPs and has a clear meaning of "first." In example 9, the child responds to the father's question about the toy's location with *sin*1, implying "I don't know where it is right now, so let me first find it before I get back to you." Here, *sin*1 conveys a sense of ordering events: first finding the toy, then providing the location.

(9) FAT: *hai2 bin1dou6 aa3*
LOC where SFP
"Where is it?"

CHI: *m4 gin3 . ngo5 wan2 haa5 sin1*
NEG see . 1SG find INTJ SFP
"I don't know. Let me find it."

4.6.5 Imperative SFP (*laa*1)

*laa*1 is often used with invitations or requests and is more tentative and less forceful compared to *laa*3 (Fung, 2000). The difference lies in the assumed knowledge of the hearer: *laa*3 assumes no knowledge, while *laa*1 assumes the hearer should know. In example 10, the child uses *laa*1 to request the investigator to open the toy.

(10) INV: *ji3 , ngo5 m4 gau3 lik6 wo3. dim2syun3 aa1*
INTJ 1SG NEG enough strength SFP . What SFP ?
"I don't have enough strength! What can I do?"

CHI: *aa3 aa3 hoi1 laa1*
INTJ INTJ open SFP
"Open it!"

4.6.6 Exclamatory SFP (*bo*3)

*bo*3, which is not very frequent (Leung, 2005), is the last SFP acquired by monolingual children and is not acquired by bilingual and trilingual children in this study. While this might suggest that *bo*3 is more complex than other SFPs, it is important to consider that its complexity may stem from its association with the epistemic domain. However, the late or non-acquisition of *bo*3 could also be attributed to the rarity of situations where it is used,

leading to lower input and output frequencies in child-directed speech and child production. *bo*3 primarily expresses appreciation or approval in exclamations, marking sentences as exclamatives. Some consider *bo*3 and *wo*3 as free variants or completely overlapping elements, while others, such as Matthews and Yip (1994; 2013), treat them as different SFPs with distinct functions. In example 11, the child uses *bo*3 to emphasize that there is still one car left, adding emotion and exclamation to the utterance. This usage could also be interpreted as conveying "noteworthiness," similar to *wo*3, as the child discovers that he hadn't finished putting all the cars inside the box.

(11) The child is putting the toy cars in the box.
CHI: *zung6 jau5 jat1 gaa3 bo3*
still have one CL SFP
"There's one left."

CHAPTER 5

DISCUSSION

The present study aimed to investigate the acquisition and usage of SFPs in Cantonese-speaking children from monolingual, bilingual, and trilingual backgrounds. By analyzing data from multiple corpora, patterns are uncovered in SFP development, compare SFP usage across different language backgrounds, and identify factors influencing SFP acquisition and usage. The findings revealed both similarities and differences in SFP acquisition and usage among monolingual, bilingual, and trilingual children. Correlation and regression analyses showed strong positive relationships between age, MLU of child, and SFP percentage in child speech across all language backgrounds, suggesting that SFP acquisition is closely tied to overall language development. However, the influence of quantity and quality of parental input on child SFP usage varied, with stronger correlations observed in monolingual and trilingual children compared to bilingual children.

The following section will delve deeper into the implications of these findings for our understanding of SFP acquisition patterns. By interpreting our results in light of the existing literature and examples of SFP usage from our data, we aim to contribute new insights into the fascinating world of Cantonese SFP acquisition and usage in multilingual children.

5.1 Implications

5.1.1 Acquisition patterns

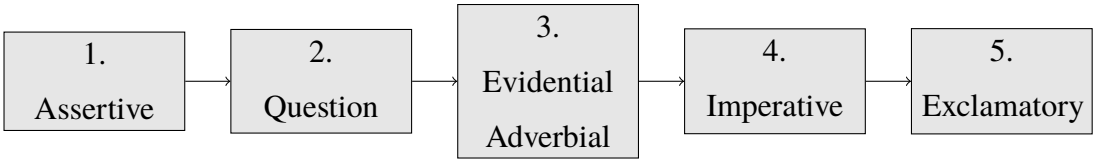
All children in the three groups (monolingual, bilingual, and trilingual) demonstrated positive and on-track lexical development, as evidenced by their MLU scores, which were even greater than the standard average MLU proposed by Kwong (1990), indicating robust language growth across all language backgrounds.

Similar to the findings of Lee and Law (2001), most children in our study started using three or more SFPs at the age of 1;08 or 1;09, regardless of their language background. However, our results also confirmed Lee and Law's (2001) observation that not all SFPs are fully acquired by the end of the observational stage at three years old. The acquisition process

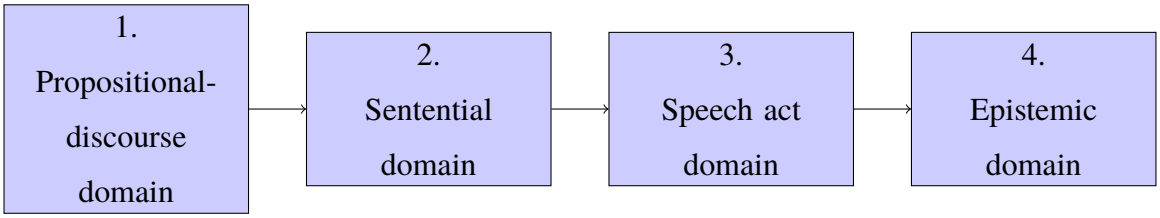
of Cantonese SFPs appears to be motivated more by semantic factors than syntactic ones, which is similar to the findings of Shirai et al. (2000) in their study of Japanese particles. While this contrasts with Fujimoto's (2008) emphasis on the syntactic aspects of particle acquisition, our results do agree with some aspects of Fujimoto's proposed process. For example, disyllabic particles are not acquired during the observational period because they are acquired last, and particles associated with illocutionary acts are acquired early, though not first.

Unlike Japanese particles, which are considered to be completely acquired by the age of three (Fujimoto, 2008), Cantonese SFPs take longer to acquire. The onset of SFP usage in Cantonese-speaking children (Overall: 1;03;10-1;11;00, Monolingual: 1;07;03-1;11;00, Bilingual: 1;03;10-1;08;28, Trilingual: 1;06;21) is also generally later than in Japanese (1;00-1;49). Furthermore, because Cantonese SFPs tend not to be restricted in referencing temporal aspects, the acquisition process may not have a distinctive stage of referencing the past and present, as observed in Shirai's study of Japanese particles.

The acquisition process of Cantonese SFPs can be analyzed based on the six categories identified by Yip and Matthews (1994; 2013):



Furthermore, the acquisition process can be analyzed with operating levels of the MMUs in SFPs (Sybesma & Li, 2007):



The acquisition timeline observed in this study aligns with Lee and Law's (2001) findings, which suggested that the mastery of more complex meanings, especially those involving epistemic modality, is usually not achieved until around the age of 6. In the current study, SFPs related to the epistemic domain were not among the earliest acquired ones. SFPs with $-k$ and SFPs in tone 4 were not found at all, and while SFPs in tone 1 ($me1, laa1, sin1, lo1$) were acquired, they were not the earliest or most frequent ones. Their acquisition seemed to be primarily based on their function categories (Timeline 1). The exclamatory SFP $bo3$, known to be associated with the epistemic domain, was not acquired by monolingual, bilingual, or trilingual children. These findings suggest that among the two proposed acquisition processes

(function category-based and operating level-based), the function category-based process has a more prominent effect on SFP acquisition. Within this process, operating levels might play a role, but given the limited research on the use of MMUs, this remains a theoretical proposition.

5.1.2 Cross-language Influence

The effect of cross-language influence is evident in SFP acquisition of the bilingual distant group (Cantonese-English). The bilingual children in our study not only failed to acquire some of the target SFPs but also used a lower percentage of SFPs in their speech compared to their monolingual and trilingual peers. Table 4.1 presents the frequency of SFPs in child speech as a percentage of total utterances across monolingual, bilingual, and trilingual children at different age points. While the monolingual group shows a consistent increase in the frequency of SFPs used by children as they grow older, the bilingual group demonstrates a slower pace of increase in SFP frequency over time. This suggests that bilingual children may face more challenges in incorporating SFPs into their language production, possibly due to the limited cross-linguistic influence from English.

The bilingual child Darren in our study failed to acquire 3 out of 10 SFPs, despite having parental input, parent MLU, and child lexical development that were on track and comparable to the monolingual and trilingual children. In contrast, the trilingual child Leo only failed to acquire one out of ten SFPs, while the monolingual children acquired all ten, given adequate input and normal language development. These differences may be attributed to the higher cross-language distance between the languages acquired by the bilingual child. Although Cantonese and Mandarin do not have many SFP cognates, and none of the ten SFPs in this study have directly transferrable cognates in Mandarin, the presence and usage of SFPs in Mandarin can facilitate the acquisition of Cantonese SFPs. English, on the other hand, has some similar representations of SFPs, but they are not directly comparable due to differences in word class, syntactic position, and function, making them less transferable. As a result, the co-activation effect (Yan et al., 2013) may not be as prominent in bilingual Chinese-English children as it is in L1-L2 speakers.

Table 5.1 shows the post-hoc analysis of an additional child (Sophie) from the bilingual corpus. Among all four children studied, Sophie had the highest initial MLU and started using 5 different SFPs in her first recorded speech sample. However, as evident from the table, she did not acquire all ten SFPs by the end of the observation period, specifically lacking $ge2$ and $bo3$, with limited acquisition of $sin1$. This post-hoc example demonstrates that Darren, the

Table 5.1: Frequency of Adult Input and Child Output of SFPs in a Bilingual Context
Child: Sophie

Category	SFP		1;10	2;00	2;03	2;04	2;06	2;07	2;10	Total	r
Adverbial	sin1	PAR	21	6	15	3	7	8	15	75	0.23
		CHI	0	0	3	0	0	1	0	4	
Assertive	aa3	PAR	281	388	316	182	320	244	268	1999	0.60
		CHI	147	174	123	76	108	152	152	932	
	laa3	PAR	50	17	18	22	49	28	28	212	-
		CHI	1	10	2	5	5	6	10	39	
	ge2	PAR	4	12	9	0	6	9	8	48	-
		CHI	0	0	0	0	0	0	0	0	
	gaa3	PAR	78	67	115	51	111	162	107	691	0.72
		CHI	1	4	20	24	14	55	17	135	
Evidential	lo1	PAR	2	6	5	2	5	8	17	45	0.84
		CHI	0	1	1	1	6	2	10	21	
	wo3	PAR	27	29	24	11	21	22	45	179	0.83
		CHI	0	0	1	0	0	1	16	18	
Exclamatory	bo3	PAR	3	6	1	0	0	2	0	12	-
		CHI	0	0	0	0	0	0	0	0	
Imperative	laa1	PAR	30	25	16	17	28	13	40	169	-
		CHI	0	1	2	2	10	10	5	30	
Question	me1	PAR	42	62	89	14	61	68	67	403	-
		CHI	1	0	1	4	0	4	1	11	

other bilingual child in our main analysis, was not an isolated case. Although bilingual children generally perform well in overall lexical development, in terms of SFP acquisition, the lack of cross-linguistic facilitation from English may have led them to perform slightly worse than the trilingual child and show latency compared to monolingual development.

These results have demonstrated the importance of considering cross-language influence and typological distance when examining SFP acquisition, or in general, language acquisition in multilingual contexts. The absence of direct SFP cognates between Chinese and English, as well as the structural differences between the two languages, may hinder the acquisition of certain particles in bilingual children. In contrast, the presence of SFPs in Mandarin, despite the lack of direct cognates, appears to provide some facilitative effects for trilingual children acquiring Cantonese.

Possible Transfers from Mandarin

In the trilingual child Leo's use of SFPs, there are a few instances that potentially demonstrate the influence of his knowledge of Mandarin SFPs. As discussed in section 2.4.2, there are several potential form-similar cognates of SFPs between Cantonese and Mandarin. The way

Leo uses Cantonese SFPs suggests that he may be employing them in a manner similar to their Mandarin counterparts, particularly in terms of syntax and semantics.

In example 12, the child uses *sin*1 in the expected way in Cantonese: at the utterance-final position, with the meaning of "first" and conveying a request to the mother not to touch.

(1) CHI: *m4hou2 gaau2 sin1 . ngo5 gaa3*
NEG do SFP 1SG SFP
"Don't touch. It's mine."

However, in example 13, after the mother asks the child to command her in a toy play for fireman rescue, he used *sin*1 before the verb phrase "save DET CL kid," which is how the Mandarin counterpart of *sin*1 is used (pre-VP) instead of the Cantonese syntactic position (post-VP). A natural Cantonese expression would be *gau3 ji1 go3 siu2pang4jau5 sin1* (save DET CL kid first), possibly with the addition of the SFP *laa*1 utterance-finally to convey a sense of request or command.

(2) MOT: *nei5 zi2fai1 ngo5 aa3*
2SG command 1SG SFP
"You should command me."

CHI: *nei5 sin1 gau3 ji1 go3 siu2pang4jau5*
2SG ADV save DET CL kid
"You should save this kid first."

In this case, the child seems to understand the connotation of adding *sin*1 into the sentence but adopts the structure and placement from its Mandarin counterpart.

In most instances, Leo uses *laa*3 accurately, as shown in example 14, where he responds to the mother's question with *laa*3, indicating that he has finished.

(3) MOT: *gaau2dim6 mei6*
finish have-not
"Have you finished?"

CHI: *gaau2dim6 laa3*
finish SFP
"Finished!"

However, in example 15, it appears that the child has used Cantonese *laa*3 in a way similar to how Mandarin *le* is used. Mandarin *le* is a "change of state" marker or a symbol of a "currently relevant state," tied to evidentiality aspects (Li & Thompson, 1981). It is also the perfective marker in Mandarin, as in *wo3 che1 le fan4* ("I have eaten"). Note that Cantonese *laa*3 and Mandarin *le* do not map perfectly: for the post-verbal perfective marker, Cantonese uses *zo*2, while Mandarin still uses *le*.

In this example, the child's first use of *laa*3 seems unnatural. A natural Cantonese expression would use *laa*1 instead of *laa*3, preceding *maa*3. This usage is possibly influenced by and transferred from the use of Mandarin *le*. To express the same semantic content in Mandarin, one could say *zhe4li3 (de) wei4zi gou4 le ma* (here (POSS) space enough SFP SFP). The mother's prompts helped the child use the SFP more accurately: she dropped the SFP *maa*3 and reconfirmed that there are enough people, not enough space. This example also demonstrates how high-quality parental input can help refine a child's language acquisition process.

(4) CHI: *jan1wai6 ji1dou6 gau3 wai2 laa3 maa3*
because here enough position SFP SFP
"Because there's enough space here."

MOT: *gau3 wai2*
enough position
"Enough space?"

CHI: *hai6 jaa3*
COP SFP
"Yup!"

MOT: *o4 , gau3 jan4 laa3*
INTJ enough people SFP
"Oh, there's enough people."

CHI: *hai6 jaa3*
COP SFP
"Yup!"

MOT: *mou5 wai2 laa3 , hai6mai6 aa3*
not-have position SFP is-not SFP
"Not enough space , right?"

CHI: *mou5 wai2 laa3*
not-have position SFP
"Not enough space."

While these two cases are observable from the data, there might be more instances of transfer from Mandarin in the child's Cantonese speech. It is possible that the frequencies of SFPs used and the SFPs the child acquired are influenced by such uses. Therefore, even though the trilingual child has acquired more SFPs, or acquired them more advanced or faster than the bilingual children, it does not necessarily mean that their usage is comparable to or as accurate as that of monolingual children.

Relevance of Tone 3 in SFP Acquisition

The investigation into Cantonese SFPs reveals a particular prominence of tone 3. As indicated by Yip (2002) and Law (1990), tone 3 functions as the default and unmarked tone within the spectrum of Cantonese SFPs. This default status implies that tone 3, by itself, does not infuse additional meaning into the SFPs; rather, it fulfills the phonological requirement that each syllable in Cantonese must carry a tone.

In this study, among the first acquired SFPs across all three language groups (monolingual, bilingual, and trilingual), SFPs with tone 3 (*aa*3, *laa*3, *gaa*3, *wo*3) stand out conspicuously. This finding serves as evidence supporting the idea that tone 3 is the basic default tone for SFPs in Cantonese. Sybesma and Li (2007) observed that in both the *l*-family and the *g*-family of SFPs, the particle identified as the "base particle" carries tone 3. Furthermore, Law (1990) suggests that SFPs with a high tone differ from their tone 3 counterparts in being slightly softer and less committed. When examining the complete array of SFPs, it becomes apparent that in all families, the tone 3 particle is the semantically barest and most neutral member. Thus, it is natural and plausible for children to acquire SFPs with tone 3 first. However, it is important to note that this area has not been extensively studied, and the current results, while interesting, cannot serve as definitive evidence.

5.2 Further Research

The importance of this study lies in three key aspects: understanding the acquisition processes and timelines of Cantonese SFPs, identifying the factors influencing their acquisition in a multilingual context, and revealing the cross-linguistic influence in bilingual and trilingual acquisition through the lens of SFP acquisition. The acquisition process provides insights into the early acquisition of semantic elements, which complements the typical morpho-syntactic focus in language acquisition research. By examining how children prioritize assertion and acquire default and simpler forms first, we can better understand the essence of early discourse in Cantonese. Moreover, this study demonstrates that SFPs are still acquired under the influence of bilingualism and trilingualism, albeit with some differences compared to monolingual acquisition. The findings reveal that cross-linguistic influence plays a significant role in shaping the acquisition and usage of SFPs in bilingual and trilingual children. The study highlights the impact of language distance, with Cantonese-English bilingual children showing evidence of cross-language distance effects, such as the non-acquisition of certain SFPs and lower frequencies of SFP usage compared to their monolingual and trilingual counterparts. However, the Cantonese dominance of these bilingual children suggests that the influence of English may not pose a long-term problem for their overall language development. On the other hand, the trilingual child, exposed to both Cantonese and Mandarin, exhibited a more rapid acquisition of SFPs, possibly benefiting from the typological proximity and the presence of similar particles in Mandarin. This finding underscores the importance of considering the specific language pairs and their typological relationships when examining cross-linguistic influence in multilingual acquisition.

By assessing the overall language development and speech fluency of bilingual and trilingual children, this study demonstrates that being multilingual does not significantly hinder children's language development, and they do not lag far behind their monolingual peers. SFPs, like many other linguistic features, are acquired over time. Parental input, particularly SFP input in child-directed speech, emerges as a crucial factor in facilitating earlier exposure to SFPs. It is essential for parents to engage in frequent and contextually diverse communication with their children to increase input and exposure to various daily scenarios. The same principle applies to L2 and Ln learners, highlighting the importance of input and exposure to different contexts in achieving fluency.

However, this study has several limitations. Firstly, the corpora used are not entirely balanced in terms of the number of samples and child participants, with the trilingual corpus

including only one child. Additionally, other factors in the corpora, such as parents' educational backgrounds, which can influence the quality of child-directed speech, are not well-controlled. This is evident in the bilingual data, where children generally exhibit earlier language development. Furthermore, the data are drawn from different time periods: monolingual data from the 1990s, bilingual data from the 2000s, and trilingual data from the 2010s. Although a few decades may not significantly influence SFP usage, the approach to prompting child speech may have differed, leading to different results and potentially not accurately assessing children's real-time language capabilities. Moreover, speech data from different corpora are obtained under different stimuli, with some situations (e.g., playing with toys) creating more opportunities for SFP use compared to others (e.g., reading books). It is also important to acknowledge the possibility of human errors in the transcription of audio data and calculations, as well as potential errors in computer-assisted annotations, such as miscategorizing phonetic or orthographic equivalents of SFPs into the SFP word class. Although the software used in this study (Python libraries and CLAN) is advanced, such errors may still occur.

Given the small-scale nature of this study and its limitations in representativeness, future research should focus on obtaining large-scale, balanced, and well-controlled data. Employing more advanced correlation analyses, such as machine learning and other regression models, can provide deeper insights. Additionally, considering a wider range of potential factors and using more metrics, such as the Type-Token Ratio for assessing lexical development, can further enhance our understanding of SFP acquisition. To gain a more comprehensive understanding of cross-linguistic influence in bilingual and trilingual acquisition, future studies should aim to identify and analyze all accurate uses and errors occurring in child speech. By examining these instances in detail, researchers can determine whether they present clear evidence of language transfer. This approach would involve a thorough qualitative analysis of the children's utterances, focusing on the specific contexts and functions of SFPs in their speech. Researchers could compare the use of SFPs by bilingual and trilingual children to that of monolingual children, as well as to adult native speakers, to identify any deviations or non-target-like usage patterns. They could also investigate the potential influence of the children's other languages on their Cantonese SFP usage, considering factors such as language dominance, proficiency, and the typological similarities or differences between the languages.

Furthermore, longitudinal studies that track the development of SFP usage in bilingual and trilingual children over an extended period could provide valuable insights into the trajectory of cross-linguistic influence and its potential long-term effects on pragmatic competence. By comparing the children's SFP usage at different stages of development,

researchers could determine whether any observed transfer effects are temporary or persistent and how they may evolve as the children's language proficiency increases. As SFPs are indicators of fluency and natural speech, they can be used to assess language development in Cantonese for both L1 toddlers and L2/Ln learners. This study lays the groundwork for future research in this area, highlighting the importance of understanding the acquisition of these vital pragmatic markers in various linguistic contexts.

CHAPTER 6

CONCLUSION

This study aimed to investigate the acquisition of sentence-final particles (SFPs) in Cantonese-speaking children from monolingual, bilingual, and trilingual backgrounds. By analyzing data from multiple corpora, the thesis sought to uncover patterns in SFP development, compare SFP usage across different language backgrounds, and identify factors influencing SFP acquisition and usage.

The findings revealed both similarities and differences in SFP acquisition and usage among monolingual, bilingual, and trilingual children. The acquisition process of Cantonese SFPs appears to be motivated more by semantic factors than syntactic ones, with children acquiring the most frequent and functionally simple SFPs earlier. However, the rate of acquisition and the specific SFPs acquired varied across language backgrounds, with bilingual children showing a slightly slower acquisition rate compared to their monolingual and trilingual peers.

Importantly, the study shed light on the significant role of cross-linguistic influence in shaping the acquisition and usage of SFPs in bilingual and trilingual children. The findings demonstrated that language distance and typological similarities between the children's languages can have a notable impact on their SFP development. While trilingual children, who were exposed to both Cantonese and Mandarin, exhibited a relatively quick acquisition of SFPs, possibly benefiting from the presence of similar particles in Mandarin, bilingual children, whose other language was English, showed a slower acquisition rate. This difference may be attributed to the typological differences between Cantonese and English and the lack of direct equivalents for SFPs in English.

These findings contribute to our understanding of the complex interplay between language-specific and universal factors in shaping children's pragmatic competence in multilingual contexts. They highlight the importance of considering cross-linguistic influence and the specific language combinations when examining the acquisition of pragmatic markers like SFPs in bilingual and trilingual children.

The study also emphasized the crucial role of parental input in shaping children's SFP usage, with a strong positive correlation observed between adult SFP input and child SFP

output. This finding underscores the importance of language exposure and interaction in facilitating the acquisition of pragmatic markers like SFPs.

The insights gained from this study have practical implications for language education and assessment. They can inform pedagogical approaches and materials design for teaching Cantonese as a first or additional language, emphasizing the importance of providing rich input and opportunities for interaction to support the acquisition of SFPs. Additionally, the study's findings can guide the development of assessment tools for evaluating pragmatic competence in Cantonese-speaking children, taking into account the developmental trajectories, factors influencing SFP acquisition, and the potential impact of cross-linguistic influence.

However, it is important to acknowledge the limitations of the study, such as the relatively small sample size, especially for the trilingual group, and the potential influence of individual differences and socioeconomic factors on language acquisition. Future research should aim to replicate and extend these findings with larger and more diverse samples, considering a wider range of factors that may impact SFP acquisition and usage.

In conclusion, this study provides a valuable contribution to the understanding of SFP acquisition in Cantonese-speaking children, highlighting the importance of language background, input factors, cross-linguistic influence, and the semantic-pragmatic nature of these particles. As a preliminary study to systematically compare SFP acquisition across monolingual, bilingual, and trilingual children, with a specific focus on cross-linguistic influence, this work lays the foundation for future research in this area, exploring the world of pragmatic development in multilingual contexts.

REFERENCES

Bakker, D., Müller, A., Velupillai, V., Wichmann, S., Brown, C. H., Brown, P., Egorov, D., Mailhammer, R., Grant, A., & Holman, E. W. (2009). Adding typology to lexicostatistics: A combined approach to language classification. *Linguistic Typology*, *13*(1), 169–181.

Blom, E., et al. (2020). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language*, *40*(2), 151–171.

Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, *38*(4), 317–322.

Brown, L., & Iwasaki, N. (2013). Cross-linguistic influence in the l2 acquisition of korean case particles by japanese-speaking and english-speaking learners: L1-l2 proximity and learner perceptions. *Electronic Journal of Foreign Language Teaching*, *10*(2).

Chevalier, S. (2015). *Trilingual language acquisition: Contextual factors influencing active trilingualism in early childhood* (Vol. 16). John Benjamins.

Chor, W. (2018). Sentence final particles as epistemic modulators in cantonese conversations: A discourse-pragmatic perspective. *Journal of Pragmatics*, *128*, 34–47.

Fung, R. S. Y. (2000). *Final particles in standard cantonese: Semantic extension and pragmatic inference* [Doctoral dissertation, The Ohio State University].

Fujimoto, M. (2008). *L1 acquisition of japanese particles: A corpus-based study* [Doctoral dissertation, City University of New York].

Hancil, S., Haselow, A., & Post, M. (Eds.). (2015). *Final particles*. De Gruyter Mouton.

Hao, M., Shu, H., & Xing, A. (2008). Early vocabulary inventory for mandarin chinese. *Behavior Research Method*, *40*, 728–733.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children*. Paul H. Brookes.

Hoff, E., Core, C., Place, S., Rumiche, R., Senor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, *39*(1), 1–27.

JASP Team. (2024). JASP (Version 0.18.3)[Computer software]. https://jasp-stats.org/

Kellerman, E. (1979). Transfer and non-transfer: Where we are now. *Studies in Second Language Acquisition*, *2*(1), 37–57.

Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, *12*(2), 251–269.

Kohnert, K. J., Bates, E., & Hernandez, A. E. (1999). Balancing bilinguals lexical-semantic production and cognitive processing in children learning spanish and english. *Journal of Speech, Language, and Hearing Research*, *42*(6), 1400–1413.

Kwok, H. (1984). *Sentence particles in cantonese* (Vol. 56). Centre of Asian Studies, University of Hong Kong.

Kwong, S.-M. (1990). *The syntactic development of cantonese-speaking preschool children* [M.Ed. dissertation]. The University of Hong Kong.

Laleko, O., & Polinsky, M. (2013). Marking topic or marking case: A comparative investigation of heritage japanese and heritage korean. *Heritage Language Journal*, *10*(2), 178–202.

Law, S. B. (1990). *The syntax and phonology of cantonese sentence-final particles* [Doctoral dissertation, Boston University].

Lee, J. L., Burkholder, R., Flinn, G. B., & Coppess, E. R. (2016). *Working with chat transcripts in python* (Technical Report No. TR-2016-02). Department of Computer Science, University of Chicago.

Lee, J. L., Chen, L., Lam, C., Lau, C. M., & Tsui, T.-H. (2022). Pycantonese: Cantonese linguistics and nlp in python. *Proceedings of the 13th Language Resources and Evaluation Conference*.

Lee, T. H., Wong, C. H., Leung, S., Man, P., Cheung, A., Szeto, K., & Wong, C. S. P. (1991-94). The development of grammatical competence in cantonese-speaking children [Report of RGC earmarked grant].

Lee, T. H.-T., & Law, A. (2001). Epistemic modality and the acquisition of cantonese final particles. In M. Nakayama (Ed.), *Issues in east asian language acquisition* (pp. 67–128). Kurosio.

Leung, H. H. L. (2016). *The semantics of utterance particles in informal hong kong cantonese (natural semantic metalanguage approach)* [Doctoral dissertation, Griffith University].

Leung, W. Two historical sources of the final particle wo in cantonese [Paper presented at the 13th Annual Meeting of the International Association of Chinese Linguistics]. In: In *Proceedings of the 13th annual meeting of the international association of chinese linguistics*. Paper presented at the 13th Annual Meeting of the International Association of Chinese Linguistics. Leiden, The Netherlands, 2005.

Li, C., & Thompson, S. (1981). *A functional reference grammar of mandarin chinese*. University of California Press.

Luke, K. K. (1990). *Utterance particles in cantonese conversation*. John Benjamins.

Luke, K. K., & Wong, M. L. (2015). The hong kong cantonese corpus: Design and uses. *Journal of Chinese Linguistics*, *25*, 309–330.

MacWhinney, B. (2000). *The childes project: Tools for analyzing talk* (3rd). Lawrence Erlbaum Associates.

Mai, Z., & Yip, V. (2022). Caretaker input and trilingual development of mandarin, cantonese and english in early childhood (1;6-2;11). *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2022.2060037

Mai, Z., Zhao, L., & Yip, V. (2022). The mandarin ba-construction in school-age heritage speakers and their parental input. *Linguistic Approaches to Bilingualism*, *12*(3), 377–405.

Matthews, S., & Yip, V. (1994). *Cantonese: A comprehensive grammar* (1st). Routledge.

Matthews, S., & Yip, V. (2013). *Cantonese: A comprehensive grammar* (2nd). Routledge.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech, Language, and Hearing Research*, *24*(2), 154–161.

Paradis, J., & Genesee, F. (1996). Syntactic development in bilingual children: Autonomous or independent? *Studies in Second Language Acquisition*, *18*, 1–25.

Richards, J., et al. (1986). Longman dictionary of applied linguistics. *RELC Journal*, *17*(2), 105–110.

Shirai, J., Shirai, H., & Furuta, Y. (2000). Acquisition of sentence-final particles in japanese. In *New directions in language development and disorders* (pp. 243–250).

Sheng, L., et al. (2016). A robust demonstration of the cognate facilitation effect in first-language and second-language naming. *Journal of Experimental Child Psychology*, *141*, 229–238.

Snow, D. (2004). *Cantonese as written language: The growth of a written chinese vernacular*. University of Hong Kong Press.

Sybesma, R., & Li, B. (2007). The dissection and structural mapping of cantonese sentence final particles. *Lingua*, *117*, 1739–1783.

Tang, S. W. (2015). *Yueyu yufa jiangyi [lectures on cantonese grammar]*. The Commercial Press.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of english, italian, and mandarin. *Journal of Child Language*, *24*, 535–565.

Tardif, T., Fletcher, P., Liang, W., Zhang, Z., Kaciroti, N., & Marchman, V. A. (2008). Baby's first 10 words. *Developmental Psychology*, *44*(4), 929.

Tse, S.-K. (1993). *The composing process of hong kong children in primary schools* [Ph.D. dissertation]. University of Nottingham.

Tse, S., & Li, H. (2011). *Early child cantonese: Facts and implications* (Vol. 42). Walter de Gruyter.

Unsworth, S. (2013). Current issues in multilingual first language acquisition. *Annual Review of Applied Linguistics*, *33*, 21–50.

Wang, L., & Kirkpatrick, A. (2013). Trilingual education in hong kong primary schools: A case study. *International Journal of Bilingual Education and Bilingualism*, *16*(1), 100–116. https://doi.org/10.1080/13670050.2012.689479

Wichmann, S., Holman, E. W., & Brown, C. H. (2016). The asjp database (version 17).

Yan, S., Mai, Z., & Zhao, Y. (2023). Positive cross-linguistic influence in the representation and processing of sentence-final particle le by l2 and heritage learners of chinese. *Frontiers in Psychology*, *14*.

Yip, V., & Matthews, S. (2006). Assessing language dominance in bilingual acquisition: A case for mean length utterance differentials. *Language Assessment Quarterly: An International Journal*, *3*(2), 97–116.

Yip, V., & Matthews, S. (2007). *The bilingual child: Early development and language contact*. Cambridge University Press.