International Journal of Business, Human and Social Sciences
ISSN: 2517-9411
Vol:12, No:4, 2018

A Corpus-Based Analysis on Code-Mixing Features in Mandarin-English Bilingual Children in Singapore

Xunan Huang, Caicai Zhang

***Abstract*—**This paper investigated the code-mixing features in Mandarin-English bilingual children in Singapore. First, it examined whether the code-mixing rate was different in Mandarin Chinese and English contexts. Second, it explored the syntactic categories of code-mixing in Singapore bilingual children. Moreover, this study investigated whether morphological information was preserved when inserting syntactic components into the matrix language. Data are derived from the Singapore Bilingual Corpus, in which the recordings and transcriptions of sixty English-Mandarin 5-to-6-year-old children were preserved for analysis. Results indicated that the rate of code-mixing was asymmetrical in the two language contexts, with the rate being significantly higher in the Mandarin context than that in the English context. The asymmetry is related to language dominance in that children are more likely to code-mix when using their nondominant language. Concerning the syntactic categories of code-mixing words in the Singaporean bilingual children, we found that noun-mixing, verb-mixing, and adjective-mixing are the three most frequently used categories in code-mixing in the Mandarin context. This pattern mirrors the syntactic categories of code-mixing in the Cantonese context in Cantonese-English bilingual children, and the general trend observed in lexical borrowing. Third, our results also indicated that English vocabularies that carry morphological information are embedded in bare forms in the Mandarin context. These findings shed light upon how bilingual children take advantage of the two languages in mixed utterances in a bilingual environment.

***Keywords*—**Code-mixing, Mandarin Chinese, English, bilingual children.

I. Introduction

CODE-MIXING refers to the use of two or more languages in a single utterance [1], [2]. It is an intra-sentential mixing of different linguistic units which are relatively constrained by grammatical structures [3]. As a byproduct of globalization, it is easy to find code-mixing in bilinguals in recent decades. For example, in Hong Kong and Singapore, code-mixing is quite common in daily communications [4], [5]. Rather than being a sign of incomplete acquisition [6] or a hint of lack of differentiation between the different linguistic systems [7], it is widely accepted as an efficient social and communicative practice [8]. Understanding code-mixing features helps us to figure out how a bilingual child takes advantage of items from two languages in mixed utterances in a bilingual environment.

The code-mixing features in Cantonese-English bilingual children in Hong Kong have been systematically examined (see [2]). In the study, Yip and Matthews found that the code-mixing rates between the Cantonese and English contexts are significantly different in bilingual children in Hong Kong [2]. Children tend to mix more when interacting in the Cantonese context, regardless of whether their dominant language is Cantonese or English. This is inconsistent with the previously observed language dominance patterns [9], [10] that a child is more likely to use code-mixing when communicating in their weaker languages. The authors ascribed the inconsistency to the influence of adult input.

The current study aims to investigate the code-mixing features in Mandarin-English bilingual children in Singapore. The language background of Singapore is relatively similar to that of Hong Kong. First, both Singapore and Hong Kong children are in a multilingual environment. Second, both English and Chinese are official languages in the two cities, but English is the dominant language in Singapore, whereas, in Hong Kong, Cantonese is the dominant language. An investigation showed that in families with children aged between 5;0 and 9;0 in Singapore, English is the most frequently used spoken language (50.5%) at home, Mandarin Chinese is the second most popular language (28.3%), followed by the Malay language (13.1%), Indian languages (5.8%), and Others (2.2%) [11]. It is interesting to examine whether the mixing rate pattern in bilingual children in Singapore would be related to the language dominance when the children are at the children care center. In the childcare center, the children are expected to receive similar input from the teachers. If the children mix more when using their nondominant language, i.e. in the Mandarin context, the asymmetry might be accounted for by language dominance in the Mandarin-English bilingual children. Besides, concerning the syntactic categories involved in code-mixing, Sridhar [1] hypothesized that lexical items are not equally acceptable when they are embedded in the host languages, e.g. the Spanish-English code-mixing sentence “EI man viejo esta enojado.” “The old man is mad” is hardly accepted due to the grammatical constraints. In this paper, we aim to examine the features of syntactic categories under the constraints of English and Mandarin syntax. Among the code-mixing items in the Cantonese context in Cantonese-English bilingual children, noun-mixing, verb-mixing, and adjective-mixing are three primary categories involved in mixing in the Cantonese context [2]. This pattern is consistent with the general hierarchy of word borrowability in language contact [12]. In language contact, nouns are found to be most readily borrowed, followed by verbs and adjectives. However, the pattern is not always consistent. An earlier study analyzed the

Xunan Huang is with the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (corresponding author, phone: 852-5987-5282; e-mail: xunan.huang@connect.polyu.hk).

Caicai Zhang is with the Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong. She is also with Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China (e-mail: caicai.zhang@polyu.edu.hk).

Spanish-English code-mixing by syntactic categories and found that nouns are most likely to be mixed [13], consistent with the general pattern. But, the difference starts to appear in the following rankings. In that study, the adjective-mixing ranks the second, the adverb-mixing the third, and the verb-mixing ranks the fourth. Considering Cantonese mixing in the English context in the Cantonese-English bilingual children, another study found that sentence-final particles, such as "laa1" and "aa3", are the third frequently inserted Cantonese items, preceded by noun-mixing and verb-mixing [2]. This study examines the syntactic categories in English-Mandarin speaking children in Singapore.

Another factor is whether the inserted items are lexical items or functional morphemes in code-mixing. According to the System Morpheme Principle, lexical items but not functional morphemes can be embedded in the matrix language [7], [14]. According to the theory, the morphosyntactic frame for the sentence is set by the matrix language. On the other hand, the Ivy-Hypothesis [15] maintains that the grammatical morphemes of the dominant language may be retained when speaking the weaker language. This hypothesis was proposed on the basis of mixing features in Swedish-French/Italian bilingual children. It is possible that when inserting language items in a language without inflectional features, the picture may be different. When mixing the English components in the Cantonese context, both nouns and verbs are found to be inserted in their bare forms [2]. For example, in the sentence below (Example 1), the plural form of "apple" is required, but the child used the bare form "apple."

Example 1. sik6 di1 apple
'Eat some apples.' (Alicia)

To explore whether Mandarin-English bilingual children would follow the same code-mixing pattern as the Cantonese-English bilingual children, morphological features involved in code-mixing in Mandarin-English bilingual children in Singapore will be analyzed. It is particularly interesting to analyze whether the English words would carry morphological information in the Mandarin context. Because in this condition, Mandarin is the matrix language, and English is the embedded language. Meanwhile, it is also true that English is the dominant language for the children, whereas Mandarin is their nondominant language. In Mandarin Chinese, no English-style inflectional information is required, but it is necessary to add some other syntactic information to make the utterance grammatical. For example, no morphological information is required for nouns or verbs in Mandarin, but it is necessary for speakers to add qualifiers to indicate quantity whereas in English, inflectional information is required. If the English words are embedded in bare forms in the Mandarin context, and syntactic information is provided by Mandarin Chinese, it would support the System Morpheme Principle. If the embedded English words carry morphological information, it possibly maps the Ivy-Hypothesis; because the functional morphemes of the dominant language are preserved when they are inserted in the matrix language.

To summarize, we aim to address the following three research questions:

- First, to examine whether the code-mixing rates in the Mandarin Chinese and English contexts are significantly different in Singapore bilingual children.
- Second, to explore the syntactic categories of code-mixing in Singapore bilingual children, and determine whether the pattern is consistent with the hierarchy of borrowability trend observed in language contact.
- Last, to investigate whether morphological information is preserved when inserting syntactic components into the context language.

II. Methodology

In this study, the data are derived from the Singapore Bilingual Corpus [16], in which speech utterances of a total of sixty English-Mandarin 5-to-6-year-old children were recorded and transcribed. In addition to the children's productions, nine teachers, and one caretaker were also recorded, but very few of their productions were transcribed for analysis. The corpus did not provide detailed individual information of the children, but they gave a summary of their background information (see Table I). As we can see from the table, English is the children's dominant language, while comparatively Mandarin Chinese is their nondominant language. The average socioeconomic status (SES) was measured by rating the parents' educational status in which 0 represents no formal education and 5 represents postgraduate degree. All the recordings were from children's spontaneous soliloquies and conversations at different circumstances in two private childcare centers in Singapore (Center M and Center E). For example, recordings were collected in their free play time, meal time, and their group activity. The duration of recording was 51:26:31 hours in total.

TABLE I
Background Information of the Children

Variables	Mean(SD)
age	6.06 (0.34)
Average of parents' education	3.98 (0.54)
Parental report of English exposure	55.30 (19.93)
Parental report of Mandarin exposure	41.8(20.15)

Items marked with the postcode '@s' in CHAT (Codes for the Human Analysis of Transcripts) format were used for analysis. These items refer to the utterances that contain elements from both languages [17]. The code-mixing in the English context (EC) refers to the Mandarin elements that are inserted into the English frames. For example, the nouns "蛇" ("snake") and "鱼" (fish") in Example 2, which are marked by the postcode '@s' are the Chinese mixings in the English context. Accordingly, code-mixing in the Mandarin context refers to English elements that children inserted into the Mandarin frames, such as the English noun "bookmark" in the Chinese context in Example 3.

Example 2. Is it 蛇@s or 鱼@s?
Is it a snake or a fish? (JAR)
Example 3. 是 bookmark@s 吗?
Is it a bookmark? (TAL)

The rate of code-mixing was calculated in each language

context as the number of mixed utterances in each language context divided by the total number of utterances produced by the children. Repetitions of the same utterance and incomplete utterances were excluded from analysis. There were a total of 31134 child utterances in the corpus.

III. Results

A. Rates of Code-Mixing in Mandarin and English Contexts

By calculating the code-mixing rates in the Mandarin Chinese context and English context, we found that children mixed more in the Mandarin context than in the English context. Table II shows the overall code-mixing rate in Mandarin and English context in the Singapore bilingual children. As can be seen from the table, there were more English utterances, but code-mixing was rarely used in the English context, whereas code-mixing was more frequently used in the Mandarin context. We further examined individual code-mixing patterns and found large individual differences. Among the 60 children, 24 of them never showed any code-mixing utterances, whereas some of them had a very high rate of code-mixing. More specifically, the top three code-mixing rates in the Mandarin context were 30.357%, 16.299%, and 14.906%, respectively.

TABLE II
THE OVERALL CODE-MIXING RATE IN MANDARIN AND ENGLISH CONTEXT

	Mixed	Total utterances	Ratio of mixing
Mandarin context	433	6672	6.49%
English context	12	24462	0.05%

B. Syntactic Categories in Code-Mixing

In the corpus, the syntactic categories of code-mixed words in code-mixing utterances included nouns, verbs, adjectives, pronouns, adverbs and so on. We categorized them into four groups, including nouns, verbs, adjectives, and others; all the syntactic categories except for the three major ones were counted as "others." The results showed that there were 462 nouns, 73 verbs, 67 adjectives, and 189 other items. Proportionally, nouns made up 58.41% of the total number of code-mixed words, followed by verbs and adjectives which occupied 9.23% and 8.47% respectively. Fig. 1 shows the breakdown of the main syntactic categories inserted in code-mixed utterances in the Chinese context. For the code-mixing in the English context, there were only twelve mixed utterances in the Chinese context, and all of them were nouns.

We then examined whether the English nouns and verbs carrying the morphological information occurred in their bare forms or inflected forms when inserted in the Mandarin Chinese context. We found that most of them were inserted in their bare forms. Most of them were syntactically acceptable, for the matrix language has already provided sufficient syntactic environment to make the bare form grammatical (see Example 4). However, some components were still inserted in their bare forms even when they were not acceptable according to English syntactic rules (see Example 6).

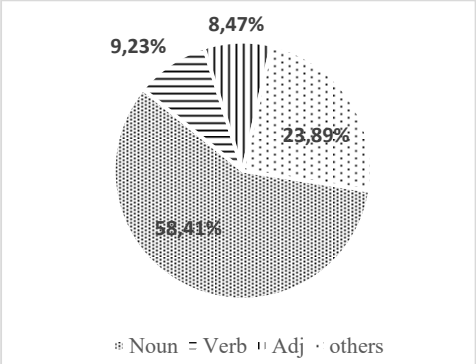

Fig. 1 The main syntactic categories inserted in code-mixed utterances in Chinese context

Example 4. 有一个 seesaw@s.
"There is a seesaw" (JON)
Example 5. 两个 soldier@s 和两个 maid@s 会帮 princess@s 的, 对吗?
"The two soldiers and the two maids would help the princess, right?" (EUG)

In Example 5, the plural forms of "soldier" and "maid" are required, but the child used the bare form in the sentence. In the corpus, there were also a few children who sometimes added morphological information along with the lexical items, such as Examples 6 and 7 that were collected from TON. Influenced by the boy, another child (Jon) also added the morphological information once in a turn taking with TON (see Example 8).

Example 6. 我以为你说 a@s clay@s.
"I thought you mean a clay" (TON)
Example 7. 那 blocks@s 呢?
"Then what about blocks" (TON)
Example 8. 我哪里讨厌, blocks@s
"I do not hate blocks at all" (JON)

IV. Discussion

A. Rates of Code-Mixing in Mandarin and English Contexts

As shown in Section III.A, the code-mixing rates of the Mandarin-English bilingual children were asymmetrical in different settings. That is, children were more likely to code-mix when using Mandarin Chinese, which is their nondominant language but tended to mix less when interacting in English, which is their dominant language. The result is consistent with the pattern of language dominance found in Danish-English and German-English bilingual children [9], [10]. When the bilingual children were not able to find proper expressions immediately, or they were sure of the meaning of a weaker language utterance, they can take advantage of their dominant language to achieve the goal. Where communicating with their dominant language, it is less necessary to adopt code-mixing to express themselves. However, the code-mixing rate in English context was significantly lower in Singapore bilingual children compared with that in the Hong Kong Cantonese-English bilingual children, although the Hong Kong bilingual children also mixed less in the English context. The low code-mixing

rate may result from the bilingualism policy in Singapore [18]. According to this policy, English is the shared official language among all Singaporean children at school. Since the data were collected from the service center, children are more likely to mix their shared language when talking in their mother tongues. On the other hand, in the current study, the children showed high individual variability even though they were in the same childcare centers. We have limited data on the language input of teachers and caregivers to the children in childcare centers. Thus, it is difficult to examine this kind of input influence. Scholars have reached a consensus that the ambient input plays a vital role in children's code-mixing [2], [19]-[21]. For example, both experimental and corpus-based studies found that the rate of mixing produced by adults is highly correlated with the children's overall rates of code mixing [2], [20]. Significant correlations can also be found between children's code-mixing rate and parents' discourse strategies. For example, Min [21] examined the relationship between parental discourse strategies and the code-mixing rate of a 2-year-old Mandarin-English bilingual girl in interactions. In that case study, the father was more tolerant of her code mixing, but her mother required the child to use English only. As a result, the child was found to use mixed Mandarin and English in interactions with her father more frequently than with her mother. Considering that children were receiving similar language inputs in the childcare centers, it is possible that the individual variability results from the influence of their parent's input and their discourse strategies.

Example 9. Syntactic Categories in Code-Mixing

For syntactic categories involved in code-mixing, we found that nouns were the major mixing category in the Singapore bilingual children, composing more than half of the total code-mixing components. And among other syntactic categories, verb-mixing and adjective-mixing were also frequently used. The "nouns > verbs > adjectives" pattern was in line with the trend observed in the Cantonese-English bilingual children in Cantonese context, and with the general hierarchy of borrowability in language contact [12]. That is, nouns are most readily borrowed, followed by verbs and adjectives. However, the pattern is not entirely consistent, for example, with the earlier study conducted by Poplack [9]. Analyzing the Spanish to English code-mixing by syntactic categories, he found that nouns are most likely to be mixed, whereas the second and the third frequently mixed types are adjectives and adverbs; verbs rank the fourth in mixing rate. The consistent results that nouns are the most likely to be mixed may be attributed to several reasons. First, nouns possibly make up the largest percentage of the syntactic categories in these languages. Second, nouns usually help to provide references to the new referent, whereas verbs rarely have a similar function. However, to address the question why the syntactic divisions involved in the code-mixing show different patterns in different languages, further investigations are needed.

Our results also indicated that English words carrying morphological information are more likely to be embedded in bare forms in the Mandarin context. And on most occasions, the context language could provide sufficient syntactic environment to make the bare form acceptable. Nonetheless, the components were still inserted in their bare forms even when they were not acceptable according to English syntactic rules. The English words are embedded in bare forms most of the time in the Mandarin context. This may mirror the System Morpheme Principle. Influenced by the syntax structure of the matrix language---Mandarin Chinese---where no inflectional information is required, the words are embedded in the bare form. However, it is also possible that inserting the words in their bare forms corresponds to the effort-saving strategy. When making use of the two languages at the same time, it could be more economical to neglect the morphological information of the inserted words. To explore the underlying reasons for this question, further studies will be conducted in the future.

Besides, there were some exceptions in that a few children added the morphological information several times when it was required. Since the corpus did not provide detailed individual information of the children, nor the information of their parental input, limited analysis can be conducted to dig into the reasons.

V. Conclusion

This study explored the code-mixing features in Mandarin-English bilingual children in Singapore. First, results indicated that the code-mixing rates in the Mandarin and English contexts are not balanced. The mixing is more prevalent in the Mandarin context than the English settings. This phenomenon can be explained by the language dominance condition, consistent with previous findings that children have a stronger tendency to switch when speaking their weaker language. When communicating with their nondominant language, they can take advantage of their dominant language to better express themselves.

Second, the observed "nouns > verbs > adjectives" code-mixing rates in Mandarin Chinese background completely echo the trend in Cantonese-English bilingual children in the Cantonese context, as well as the pattern observed in language contact situations. But it is unknown yet whether it is because Cantonese and Mandarin are similar in many ways, or due to some other reasons. Besides, children are more likely to embed the bare forms in the context language although there exist individual differences.

Acknowledgment

We thank Professor Virginia Yip for her valuable suggestions on this study.

References

[1] S. N. Sridhar, K. K. Sridhar, "The syntax and psycholinguistics of bilingual code mixing," *Canadian Journal of Psychology Revue Canadienne De Psychologie,* 1980, 34(34):407-416.

[2] V. Yip, and S. Matthews, "Code-mixing and mixed verbs in Cantonese-English bilingual children: input and innovation," Languages, 2016, 1(1): 4.

[3] T. K. Bhatia, and W. C. Ritchie, "Social and Psychological Factors in Language Mixing," In W. C. Ritchie and T. K. Bhatia (eds.), *Handbook of*

Bilingualism, 2004, pp.336-352.
[4] K. Kohnert, D. Yim, K. Nett, P. F. Kan, and L. Duran, "Intervention with linguistically diverse preschool children: a focus on developing home language(s)," *Language Speech & Hearing Services in Schools*, 2005, 36(3), 251.
[5] L. Comeau, F. Genesee, and L. Lapaquette, "The modeling hypothesis and child bilingual codemixing," *International Journal of Bilingualism*, 2003, 7(2), pp. 113-126.
[6] J. F. Hamers, and M. H. A. Blanc, *Bilinguality and bilingualism*. Cambridge University Press: Cambridge, UK, 2000, pp.241-271.
[7] V. Yip, "Simultaneous Language Acquisition," In François G. & Ping. *The Psycholinguistics of Bilingualism*. Malden, MA & Oxford: Wiley-Blackwell, 2013, pp.117-137.
[8] F. Grosjean, *Life with two languages. An introduction to bilingualism Cambridge*, MA: Harvard University Press，1982.
[9] J. Petersen, "Word-internal code-switching constraints in a bilingual child's grammar," *Linguistics,* 1988, 26(3), pp. 479-494.
[10] U. Lanvers, "Language alternation in infant bilinguals: a developmental approach to code switching," *International Journal of Bilingualism*, 2001, 5(4), 437-464.
[11] Singapore Department of Statistics (DOS) Singapore Census of Population 2010: Statistical Release 1- Demographic Characteristics, Education, Language and Religion, 2010, Retrieved from http://www.singstat.gov.sg.
[12] R. Lass, *Historical Linguistics and Language Change.* Cambridge University Press: Cambridge, UK, 1997, pp.189.
[13] S. Poplack "'sometimes I'll start a sentence in spanish y termino en espan': toward a typology of code-switching," *CENTRO Working Papers,* no. 4. Adults, 1979, 18(7-8), 83.
[14] M-S. Carol, *Dueling languages: grammatical structure in codeswitching.* Oxford University Press, 1997, 75-119.
[15] P. Bernardini, and S. Schlyter, "Growing syntactic structure and code-mixing in the weaker language: the ivy hypothesis," *Bilingualism*, 2004, 7(1), 49-69.
[16] W. Y. Quin and F. Patrycia, "Challenging the 'Language Incompetency Hypothesis': Language Competency Predicts Code-Switching," In Elizabeth Grillo, Kyle Jepson, & Maria LaMendola (Eds.), *BUCLD 39 Online Proceedings Supplement.* Massachusetts: Cascadilla Press. 2015.
[17] K. F. Cantone, *Code-Switching in Bilingual Children*; Springer: Dordrecht, The Netherlands, 2007.
[18] S. Gopinathan, "Preparing for the next rung: Economic restructuring and educational reform in Singapore," *Journal of Education and Work,* 1999, 12(3), pp. 295-308.
[19] F. Genesee, "Early bilingual development: One language or two?" *Journal of Child Language*, 1989, 16(1):161-79.
[20] L. Comeau, F. Genesee, and L. Lapaquette, "The modeling hypothesis and child bilingual codemixing," *International Journal of Bilingualism*, 2003, 7(2), pp.113-126.
[21] H. Min, "A case study on parental discourse strategies and a bilingual child's code-mixing," *Bulletin of Educational Psychology*, 2011, 43(1), pp.175-202.