Conversion of Chinese and English Bilingual Children's Ethnographic Codes Based on Web Bilingual Corpus

Wenjuan He

1College of Foreign Languages, Hengyang Normal University, Hengyang, China, 421002

*Corresponding author e-mail: 714822382@qq.com

Abstract. Parallel corpus is widely used in machine translation, lexicography, bilingual teaching. Parallel corpus has the following characteristics: 1) multiple fields. Web The fields covered by the information are very broad, such as bilingual on open curriculum plans. Resources contain resources for the content of each subject. 2) Regularity of page layout. Through observation of a large amount of bilingual data on Wb, I found that the vast majority of sentence pairs are corresponding and logical locations are adjacent. 3) Noise. Web bilingual pages usually contain large Text content that is not related to bilingual, such as comments, comments, descriptions, etc., here we will Content that is unrelated to a double statement pair is called noise. Statistical alignment is a method of high accuracy and efficiency in current alignment methods. According to our statistics, traditional statistical length alignment on standard flat text The accuracy and recall rate of the method are both around 90%. On the bilingual page, due to the above characteristics, the accuracy and recall rate of the method drop to About 55%. Much scholarship on code-switching (CS) has focused on the development of children’s bilingual ability (Azlan & Narasuman, 2013). Many of these studies, however, adopt the survey to conduct their study. This study explored how code-switching effects the development of language learning of two Chinese-English bilingual children in the Philippines, employing an ethnographically based approach to observe the participants, which lasted 4 months. The findings of the study revealed that the more progress made in English, the more frequencies children take English as matrix language (ML); both inter-sentential and intra-sentential types are presented as reported in bilinguals; CS is a helpful teaching and learning strategy for low proficiency learners. In view of the results, it is obvious that code-switching is a successful communicative strategy which accounts for the development of bilingualism. It is strongly recommended that we should do more research on the implication of this phenomena has on language acquisition right from childhood[1].

Keywords: Web, Corpus, Code-switching (CS), Chinese-English Bilinguals

1. Introduction
Code-switching (CS) is generally defined as the use of two or more languages, variety, or style by a speaker within one conversation, or between different interlocutors or situations (Romaine, 1989). For their speech development, young children can acquire two or more linguistic forms more easily with the help of code-switching in a multilingual context[2].

This opinion is supported by some language instructors who support bilingual instruction in the form of code-switching believe it to be extremely useful to students in many different aspects, especially in the teaching of beginner students (Ahmad & Jusoff, 2009). It is effective for students to practice code-switching during their language learning and teachers are encouraged to use code-switching when teaching students of low proficiency (Azlan & Narasuman, 2013). The use of code-switching in the language instruction has multiple functions that support an effective learning process (Fachriyah, 2017). Moreover, some researchers, such as Alam (2006), Hossain and Bar (2015), have also found code-switching can help the bilinguals to keep fluency during their engagement in conversation with other bilinguals[3].

Thus, instructors are advised to see code-switching as “a bridge between two languages that the students are learning” (Faltis, 1989, cf. Brice & Roseberry-McKibbin, 2001) as it is an essential tool for language teaching and an “interactional strategy” (Dulm, 2007) for social interactions.

Code-switching of bilingual children, as an important part of language learning process, has drawn the attention of many scholars. However, the previous research in this field has been limited in the general functions, patterns and reasons for code-switching in the bilingual children’s linguistic performance (Koban, 2013; Klapicová, 2017; Jie, 2017). A considerable amount of research has been carried out but little research focused on how the code-switching effects the development of language learning of Chinese-English bilingual children in a multilingual context.

The current study, therefore, aims to fill in this gap by exploring the process of CS undertaken by two Chinese 4-year-old children and the function of CS on their English learning when they live and study in the Philippines. The research questions were as follows:

(a) What are the frequencies of matrix language (ML) and embedded language (EL) during intra-sentential CS in different periods?

(b) What are the types of CS? Whether both inter-sentential and intra-sentential types are presented as reported in bilinguals.

(c) What are the functions of CS on children’s English learning?

To deal with the research questions more objectively and clearly, the study was guided by Myers-Scotton’s (2002) Matrix Language Framework model (MLF). Based on the linguistic and syntactic features of CS, there are two kinds of CS: inter-sentential and intra-sentential CS (Brice, 2000). The inter-sentential code-switching is used to refer to language switch at sentence boundary, while intra-sentential code-switching is used to describe language alternation produced within a sentence.

The MLF model, concerning the intra-sentential code switching, holds that one language acts as a dominant or matrix language (ML) and the other as a subordinate or embedded language (EL). It is the matrix language that plays a key role in setting the sentence frame where CS arises, while the embedded language material is considered as having been inserted in a matrix language frame. Since the distinction of the ML versus the EL is the basic principle governing code switching within the MLF model, identifying the ML in code switching utterances becomes obviously crucial. “A frequency criterion” is the one for determining the Matrix Language: the language which has the higher frequency of words in the sentence is the dominant language.
In the present study, the author mainly analyzes the amounts of ML and EL of the code-switched data in different stages, types and functions of CS, aiming at exploring the development of English learning of two Chinese children.

2. Methodology

2.1. Participants

The observed children on whom the present study is based include my twins: Kaikai (son) and Xinxin (daughter). Both of them have the same background: 4 years old; having already acquired the systematic rules of Chinese, their native language; speaking Chinese fluently; the contact with English was minimal in China; first time to go abroad; learning English in the kindergarten in the Philippines[4].

2.2. Data Gathering

This research employed an ethnographically based approach which allowed the researchers both to observe the participants and to elicit their opinions through informal interviews.

The collection of the data took four months, from August to November in 2018. The data was collected by means of direct observation and note taking. Then the researcher picked up 300(including 150 utterances spoken by KK and 150 utterances spoken by XX) valid data which would be analyzed in the present study.

The participants were observed from the following:

(a) different social spaces: home, kindergarten, amusement park, supermarket, and outside of home during playing;

(b) different interlocutors: mother, grandmother, teachers, tutor, friends, classmates, neighbors, between themselves;

(c) different activities: drawing pictures, playing games, story-telling, everyday conversation, doing assignments, watching TV, listening to tapes.

2.3. Data Analysis

The researcher would analyze the data with some examples, analysis and interpretations. And the data were divided into three periods according to the developing stages of English learning of the twins.

Period 1: August to September: initially get to know English

Period 2: October: gradually imitate English speaker

Period 3: November: actively use English

In order to facilitate clear, accurate, and comprehensive tabulation of records, three periods were numbered sequentially and coded as P1 for period one, P2 for period two, and P3 for period three; K1 for the first CS created by KaiKai and X1 for the first CS produced by XinXin, and so forth. All the utterances of CS were numbered in sequence from P1-K1 to P3-K150, P1-X1 to P3-X150.

3. Results and Discussion

A thorough and rigorous reading and analysis of 300 verbal records of the twins from August to November in 2018 and informal interview resulted in several themes which are presented in this section.
under three cardinal categories, namely Frequency, Types and Functions of CS.

3.1. The Occurrence Frequency of ML and EL

To explore CS of the children, the researcher was guided by Myers-Scotton’s (2002) Matrix Language Framework model (MLF) that has a key concept: matrix language (ML) and embedded language (EL), which, to some extent, reflects the development of the children’s English proficiency. The following table shows the frequency of ML and EL during intra-sentential CS of the twins in three periods.

Table 1. Summary of ML and EL in CS of the twins in three periods.

|            | Frequency | Total |
|------------|-----------|-------|
|            | P1  | P2  | P3  |       |
| Chinese as ML | 28 | 77  | 90  | 195   |
| and English as EL |     |     |     |       |
| English as ML  | 0  | 20  | 85  | 105   |
| and Chinese as EL |     |     |     |       |
| Total         | 28 | 97  | 175 | 300   |

As shown in table 1, in P1, Chinese is the Matrix Language (ML), however, it only garners a total frequency of 28 occurrences in the corpora of CS created by the twins. This is because the participants initially get to know English during this period. In addition, they just insert some single words into their utterances.
3.2. Types of CS Observed

A total of 300 instances of CS were found in the transcribed episodes. Table 2 shows the percentages of the inter-sentential and intra-sentential CS.

| Table 2. Number of intra-sentential and inter-sentential CS. |
|---------------------------------------------------------------|
|                | Number | Percentage |
|-----------------|--------|------------|
| Intra-sentential CS | 285    | 95%        |
| Inter-sentential CS | 15     | 5%         |
| Total           | 300    | 100%       |

As shown in table 2, both inter-sentential and intra-sentential types are presented in the twins’ utterances as reported in bilinguals. The majority of CS examples take place within a sentence, which occupying 95% while inter-sentential CS has got 5%. That is because they are just beginners of English learning with a low proficiency. There are some examples involving inter-sentential CS, most of which are produced in P3.

3.3. Functions of CS on English learning

According to the researcher’s direct observation and data analysis, it is obvious that CS plays a crucial role in Children’s English learning.

Firstly, CS helps them not only to reduce anxiety of English learning but also to enhance their confidence to speak English. Before we went abroad, they often refused to communicate with me in English and they were unwilling to listen to me when I speak English because they don’t understand. When we stayed in the Philippines, I told them that they can speak Chinese when they don’t know how to say in English within a sentence. Gradually, they are not afraid of speaking English. They like to go to school although all the teachers speak English. They can understand the teachers in class although they can’t speak out. They describe some interesting things in school to me with the help of CS[5].

The above description and analysis of data has shown that the higher frequency of ML (when English is as ML), the more proficiency of English. This findings seems to be consistent with the study conducted by Reyes (2004) who had also examined the relationship between language proficiency level and CS, stating that given the learner’s CS, we should correlate the kind of code switching used by learners with their level of proficiency in L2.

In the present research, there are two types of CS observed and most of CS are intra-sentential types. This findings are almost consistent with the study conducted by Yletyinen (2004) who found that intra-sentential code-switching is the most popular type to be used in teaching and learning. However, the result is incongruent with Mary’s(2012) study which found that all the children attending the survey used only intra sentential pattern[6].

It is also find that CS plays an important role in children’s English learning. Just as Suganda, Loeneto and Zuraïda(2018) held that code switching contributes to effective language learning and communication. Azad and Ali (2017) came to a conclusion that code-switching is a natural practice which can improve the fluency of the speakers enabling them to communicate in a better way after they investigated the case on four British-Bangladeshi bilinguals with an ethnographic approach. However, this findings seems to be in contrast to that of Azlan and Narasuman (2013) who stated in their study that code-switch themselves may be setting the wrong example to students and code-switching is not really an effective communicative tool to be used in the language classroom.
4. Conclusion

Taking an ethnographic approach, this study investigated the case of two Chinese-English bilingual children living in the Philippines, representing an attempt to make contributions to a better understanding of CS between Chinese and English. Based on the data, the study has the following conclusion[7]:

Firstly, code switching is an inevitable linguistic phenomenon during the process of children’s learning English. The more progress made in English, the more frequencies they take English as ML.

Secondly, both inter-sentential and intra-sentential types are presented as reported in bilinguals. However, intra-sentential types is the most employed one.

Thirdly, CS is an effective technique used by bilinguals. It is also a helpful teaching and learning strategy especially for low proficiency learners[8].

References

[1] Ahmad, B. H., & Jusoff, K. (2009). Teachers’ code-switching in classroom instructions for low English proficient learners. English Language Teaching, 2(2), 49-55.

[2] Alam, S. (2006). Code-mixing in Bangladesh: A case study of non-government white-collar service holders and professionals. Asian Affairs, 28(4), 52-70.

[3] Azad, M.N.H., & Ali, Md. M.(2017). Code-switching: An ethnographic case study of four British-Bangladeshi bilinguals. Language in India,17(5),24-35.

[4] Azlan, N.M.N.I., & Narasuman, S. (2013). The role of code-switching as a communicative tool in an ESL teacher education classroom. Procedia-Social and Behavioral Sciences, 90 (10), 458-467.

[5] Brice, A. (2000). Code switching and code mixing in the ESL classroom: A study of pragmatic and syntactic features. Journal of the Speech Pathology Association of Australia, 20(1), 19-28.

[6] Brice, A., & Roseberry-McKibbin, C. (2001). Choice of languages in instruction: One language or two? Teaching Exceptional Children, 33 (4), 10-16.

[7] Dulm, O.V. (2007). The grammar of English-Afrikaans code switching: A feature checking account (Doctoral dissertation, The Netherlands, LOT). Retrieved from www.lotpublications.nl/Documents/159_fulltext.pdf.

[8] Fachriyah, E. (2017). The functions of code switching in an English language classroom.