Université de Montréal

Exploring the prosodic and syntactic aspects of Mandarin-English Code-switching

par Ruowei Liu

Département de linguistique et de traduction
Faculté des arts et des sciences

Mémoire présenté
en vue de l’obtention du grade de maîtrise
en Linguistique

novembre, 2020

Résumé

L'alternance codique (Code-switching, CS) est l'un des comportements naturels les plus courants chez les bilingues. Les linguistes ont exploré les contraintes derrière l'alternance codique (CS) pour expliquer ce comportement. Au cours des dernières décennies, la recherche a plutôt été axée sur les contraintes syntaxiques et ce n'est que récemment que les contraintes prosodiques ont commencé à attirer l'attention des linguistes. Puisque la paire de langues choisie est moins étudiée dans le domaine de recherche sur la CS, les études sur la CS mandarin-anglais sont limitées en ce qui concerne les deux contraintes. Ainsi, cette étude explore à la fois les contraintes prosodiques et les schémas syntaxiques de cette paire de langues grâce à une base de données naturelle sur l'alternance codique.

Prosodiquement, l'étude applique une approche fondée sur l'information (*information-based approach)* et utilise une unité fondamentale, l'unité d'intonation (Intonation Unit, IU), pour mener l'analyse. Le résultat de 10,6 % d'IU bilingue (BIU) se révèle fiable et offre des preuves solides que l'alternance codique a tendance à avoir lieu aux frontières de l'IU chez les bilingues. Les résultats soutiennent le travail précurseur de Shenk (2006) à partir d'une paire de langues inexplorée (mandarin-anglais). De plus, cette étude développe des solutions au problème de subjectivité et au problème d'adéquation de la base de données afin de renforcer la fiabilité des résultats. D'un point de vue syntaxique, l'étude examine les schémas syntaxiques aux points de CS de la paire de langues mandarin-anglais en utilisant des données recueillies auprès d'une communauté bilingue rarement étudiée. Un schéma syntaxique spécifique à cette paire de langues a été observé en fonction des résultats, mais l'étude suggère que ce schéma ait perturbé les résultats finaux. L'étude comporte une analyse avec les résultats de l'aspect prosodique et de l'aspect syntaxique. Lorsque les résultats divergents sont éliminés, on peut observer un résultat plus solide qui soutient davantage l'argument de la contrainte prosodique.

Mots-clés : Code-switching, mandarin-anglais, Intonation Unit, les contraintes prosodiques, les schémas syntaxiques

Abstract

Code-switching (CS) is one of the most common natural behaviors among bilinguals. Linguists have been exploring the constraints behind CS to explain this behaviour, and while syntactic constraints have been the focus for decades, prosodic constraints were only studied more in depth recently. As a less common language pair in CS research, studies on Mandarin-English CS are limited for both constraints. Thus, this study explores the prosodic constraints and syntactic patterns of this language pair with a natural CS database.

Prosodically, this study applies the information-based approach and its fundamental unit, Intonation Unit (IU), to conduct the analysis. The result of 10.6% bilingual IU (BIU) proves to be reliable and offers solid evidence that bilinguals tend to code-switch at IU boundaries. This supports the pioneer work of Shenk (2006) from the unexplored Mandarin-English language pair. In addition to this, the study develops solutions to deal with the subjectivity problem and the database appropriateness problem in this approach to strengthen the validity of the results. Syntactically, this study investigates the syntactic patterns at switching points on the Mandarin-English language pair using data collected from a rarely investigated bilingual community. Based on the results, a syntactic pattern specific to this language pair was observed and this study suggests it disrupted the final results. This study conducts an analysis with the results of both the prosodic aspect and the syntactic aspect. When the interfering results are eliminated, a more solid outcome can be observed which provides greater support to the prosodic constraint argument.

Keywords: Code-switching, Mandarin-English, Intonation Unit, prosodic constraints, syntactic patterns

Table of Contents

List of Tables

List of Figures

List of Initials

CS: Code-switching

IU: Intonation Unit

SIU : Spanish Intonation Unit

EIU: English Intonation Unit

MIU: Mandarin Intonation Unit

BIU: Bilingual Intonation Unit

EXCS: Code-switching external to IU

INCS: Code-switching internal to IU

LDC: Linguistic Data Consortium

SEAME: South East Asia Mandarin English code-switching corpus

HNR: Harmonic to Noise Ratio

Hz: Hertz

F0: Fundamental frequency

List of Abbreviations

ADJ. : Adjectif

V. : Verb

N. : Noun

ADV. : Adverb

PREP. : Preposition

DET. : Determiner

AUX. : Auxiliary

PRO. : Pronoun

NP: Noun phrase

I lovingly dedicate this mémoire to my husband,
who supported me each step of the way.

Acknowledgements

First of all, I would like to express my deepest and most sincere gratitude to my supervisor, Professor Michael Dow, for his constant patience, kindness and guidance throughout this whole research. His valuable and constructive suggestions during the planning and development of the research has been very much appreciated. He convincingly encouraged me and guided me to do the right thing, even when the road got tough. My genuine thanks also go to Professor Daniel Valois, for his support of the work and his generosity of purchasing the corpus for this research, without which the research wouldn't exist.

I would also like to take the chance to show my appreciation to the Linguistics department, for offering me this chance to study and conduct research in University of Montreal. The multicultural environment in the campus and the abundant resources of the school library have indeed broaden my vision, in both the academy and in life.

I wish to pay my special regards to China Scholarship Council and Faculté des études supérieures et postoctorales of University of Montreal, for generously offering me the scholarships, which got me through the three years during my Master. Without a doubt the kind gesture has encouraged me to pursue my academic goal and shall never be forgotten.

My sincere thanks also go to my colleagues and friends of the linguistics department, we get through all the difficulties and obstacles side by side. I would like to thank Victoria Stecum, for a job wonderfully done on the proofreading of the thesis. And last but not least, I would like to thank my family and friends, for supporting me spiritually throughout my studies and my life in general.

Chapter 1 General introduction

With the development of society and the growing trend of cultural integration, bilingual communities now make up a large part of the world population. Being one of the most common linguistic behaviors in bilinguals, Code-switching (hereafter CS) has attracted significant attention in bilingual studies during the past several decades. CS refers to the alternation of two languages within a single discourse, sentence or constituent (Poplack, 1980). Linguists have been searching for constraints behind CS to explain this natural bilingual behavior. For example, where in a sentence does a switch usually take place? For years, experts in the field have focused on syntax to explain this phenomenon, proposing many syntactic constraints as possible explanations. However, there are still a number of examples that cannot be explained by them. Shenk (2006) was among the first ones to explore solutions from prosodic aspects. By introducing the *information-based approach* to CS studies, she suggested that prosodic constraints also control CS behaviors. This proposition is corroborated by studies on certain language pairs, and will require further studies in various other language pairs in order to reach a consensus. These two kinds of constraints explain CS behaviors in a more comprehensive way and offer a possibility to explain some of the counterexamples provided by the critics of the syntax-only approach.

This study explores both the prosodic constraints and the syntactic patterns of CS behaviors with a Mandarin-English CS database since research on this language pair is still limited, especially with respect to the new prosodic perspective. Concerning the syntactic patterns, this study follows previous studies of the field, investigates the patterns at switching points, and expects to find evidence that coincides with previous research. Furthermore, since the database is collected in a rarely investigated bilingual community, novel patterns are expected to be found.

Prosodically, this study follows Shenk (2006), applies the information-based approach to conduct the analysis, and presents evidence that supports the existence of prosodic constraints. In the meantime, this study attempts to deal with several problems in previous methodology, concerning the potential subjectivity problem of phonetic measurements and the problem of database appropriateness. Thus, based on the above, three objectives are proposed for the current study:

1) Investigate syntactic patterns at switching points, find whether there is evidence that coincides with previous studies, then look for novel patterns.
2) Find whether there is evidence of prosodic constraints that provides support for Shenk's proposition from the Mandarin-English language pair.
3) Develop solutions for the subjectivity problem and the database appropriateness problem in the approach, which could help make the results more conclusive.

The study is organized as follows: in Chapter 2, a comprehensive literature review is conducted. Chapter 3 is dedicated to introducing the database used for this work. In Chapter 4, the systematic prosodic methodology and the syntactic methodology in this study are presented. Then in Chapter 5, I conduct an integral discussion on the results of the prosodic constraints and syntactic patterns. Finally, Chapter 6 concludes the whole work, connects this study to other fields and discusses expectations on future research.

Chapter 2 Research background

This chapter aims to provide a thorough literature review of the critical issues discussed in the current study. In Section 2.1, a general introduction is given on the development of syntactic aspects of CS studies. First, the definition of CS is presented. Over the past few decades, linguists focused on exploring the syntactic constraints behind CS in order to explain this behavior. Thus, several well recognized syntactic constraints are introduced. Then more research has been conducted on more language pairs, and many counter-examples to these constraints were observed. Following this, previous studies on Mandarin-English CS (the target language pair of the current study) are reviewed. A brief introduction of the syntactic, morphological and phonological characteristics of both languages is offered. Studies on this language pair are quite limited in both number and scope, yet there is evidence that either supports or challenges the syntactic constraints. Thus the problem remains in the field of CS studies, this is pointed out by Shenk (2006): syntactic constraints are able to significantly advance our knowledge of CS behavior, but many of them are flexible and have a number of attested counterexamples. Therefore, syntactic constraints alone might not be enough to explain CS behavior in a comprehensive way.

Shenk (2006) was one of the first in proposing that, in CS, prosodic constraints are equally important to syntactic constraints, and can together provide a more complete explanation. Section 2.2 is dedicated to the prosodic aspects in CS studies, as follows. Shenk (2006) applied the information-based approach from monolingual studies in her analysis, and found evidence of the prosodic constraint. This finding received support from certain studies of different language pairs.

Three critical issues observed in the studies mentioned above have influenced this study's theoretical basis and methodology. Section 2.3 provides a thorough discussion on these problems. The first problem revolves around the long-existing debate over the definitions of single-word CS and borrowing. The second problem stems from the fact that in previous studies who applied the information-based approach, their methodologies suffered certain impact from the subjectivity problem, both in monolingual studies and in bilingual CS studies. The third problem is that there are currently no proper criteria to measure whether a database suits CS studies that apply the information-based approach. This third issue has led to major misunderstandings in previous studies.

Finally, Section 2.4 sums up the previous three sections based on which three objectives concerning both syntactic and prosodic aspects are proposed for the current study.

2.1 Development of CS studies on syntactic aspects

2.1.1 Definition

Code-switching, among several language contact phenomena, has drawn great attention in the domain of bilingual studies over the past few decades. Throughout these years, a number of linguists have offered their own definitions of this bilingual behavior. Though the specific definitions can differ from one another, they share the same idea. For instance, Bullock and Toribio (2009) broadly defined code-switching as the ability on the part of bilinguals to alternate effortlessly between two languages. Then there is the well-known metaphor made by Grosjean (1998), "a bilingual is like two monolinguals in one (p. 3)". Linguists regard CS as a way of showing bilinguals' ability in both languages: bilinguals have to reach certain advanced levels in both languages to switch between them and still maintain their prescribed norms. For example:

(2-1) ***Spanish-English code-switching*:**

Sometimes I'll start a sentence in Spanish ***y termino en español***.

"Sometimes I'll start a sentence in Spanish and I finish in Spanish."

(Poplack, 1980)

2.1.2 Development

2.1.2.1 Syntactic constraints

With the bilingual community in the world growing rapidly, linguists have been trying to explore the reason behind CS. In the beginning, it was thought that CS was not controlled by any grammatical rules or syntactic constraints, based on observations on Spanish-English CS (Labov, 1971; Lance, 1975). Shortly after that, the rule-governed system began to flourish, and grammatical rules such as categorical constraints were considered to govern CS behavior

(Gumperz, 1977). In his study, Gumperz tested the acceptability of a number of CS utterances, and proposed several grammatical rules. For example, the *Conjunction Constraint* states that *"in conjoined phrases, both co-ordinate and subordinate conjoined sentences can be freely switched, but the conjunction always goes with the second switched phrase."* Two pairs of sample phrases are listed below, between the two phrases in each pair, the first one is accepted while the second one is not. Note that in Gumperz (1977), the author provided all examples in English, even the switched parts were provided in italicized English equivalents:

(2-2). (a) I was reading a book ***and she was working***.

(b) * I was reading a book and ***she was working***.

(2-3). (a) I wanted to stop smoking ***but I couldn't***.

(b) * I wanted to stop smoking but ***I couldn't***.

Shortly after this, more syntactic constraints were proposed based on truly natural speech data collected from large bilingual communities. Poplack (1978) was one of the pioneers in this field, she proposed two syntactic constraints: the *Free Morpheme Constraint* and the *Equivalence Constraint*. The *Free Morpheme Constraint* indicates that codes may be switched after any constituent in discourse provided that constituent is not a bound morpheme. It is shown in the two examples below that, the first one is accepted while the second one is rejected. In the second example, a Spanish bound morpheme -idendo (*'-ing'*) is affixed to an English root 'eat' which is unacceptable according to the *Free Morpheme Constraint*.

(2-4) una buena ***excuse***

*'a good **excuse**'*

(2-5) * eat – ***iendo***

'eating'

The *Equivalence Constraint* indicates that CS tends to occur at points in discourse where juxtaposition of L1 and L2[1] elements does not violate a syntactic rule of either language, i.e. at points around which the surface structures of the two languages map onto each other. In other

[1] L1 = Language 1, L2 = Language 2.

words, should a constituent be generated by a rule that belongs to L1 and yet is not shared by L2, then a switch would be inhibited here. Figure 1 provided by Poplack (1980) illustrates this constraint. In Figure 1, between line A and B, dotted lines indicate permissible switch points, arrows indicate ways in which constituents from both languages map onto each other. Line C is the transcription of the speaker's actual utterance.

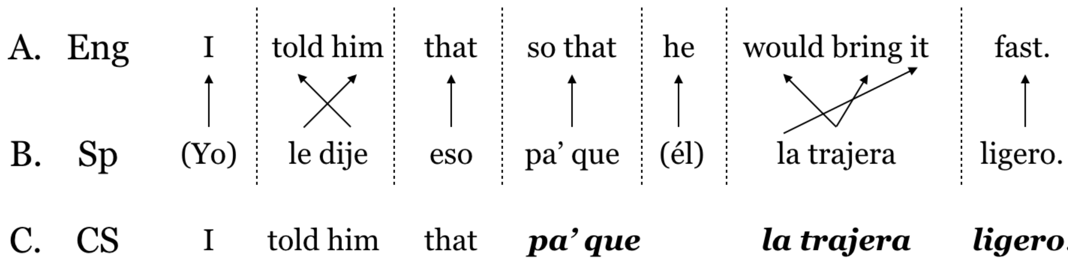


Figure 1 Permissible CS points, Poplack (1980)

Following Poplack's publications, more syntactic constraints were proposed based on work on natural speech of different language pairs (Di Sciullo, Muysken, & Singh, 1986; Kachru, 1978; Pfaff, 1979; Singh, 1983; Sridhar & Sridhar, 1980; Timm, 1978). Linguists at the time reached a general consensus: with respect to syntactic constraints, grammatical CS utterances could be produced. However, it did not take long before more language pairs were explored, and counter-examples to these constraints appeared.

2.1.2.2 Counter-examples

Bentahila & Davies' (1983) results from a study on highly proficient French-Arabic bilinguals in Morocco pose a challenge for two of the most well-accepted syntactic constraints: the *Conjunction Constraint* and the *Equivalence Constraint*. First, the *Conjunction Constraint* indicates that the conjunction always goes with the second switched phrase. However, according to Bentahila & Davies (1983), the constraint did not apply to French-Arabic CS. The following examples show that the conjunction word goes with the second phrase, as well as the first phrase.

(2-6) ʕandna bzzaf ddrija ***et je ne m'entends pas avec ma mère***.

'We have a lot of children ***and I don't get on well with my mother****.'*

(2-7) ʕTeitulu w ***il l'a analysé***.

*'I gave it to him and **he analyzed it**.'*

As for the second constraint challenged by Bentahila & Davies (1983), the *Equivalence Constraint*, which allows CS to happen only when there exists surface structure equivalence between two languages, the authors found evidence that claimed otherwise. For example, the adjectives have different positions in Arabic and in French. In Arabic, adjectives usually follow the head noun whilst in French, many adjectives do follow the noun yet some others must precede the noun. According to the *Equivalence Constraint*, a switch can only take place when the adjective follows the noun. However, in Bentahila & Davies' data, phrases that contradict this were easily found.

(2-8) j'ai vu un ancien ***tilmid djali***.

*'I saw an old **student of mine**.'*

(2-9) c'est le seul ***ustad***.

*'it is the only **teacher**.'*

With more language pairs investigated, many counterexamples questioned the universality of the syntactic constraints (Bentahila & Davies, 1983; Berk-Seligson, 1986; Bokamba, 1988, 1989).

2.1.3 Research on Mandarin-English CS

In research done on Mandarin-English CS, the target language pair of the current study, linguists found evidence that supports the syntactic constraints. However, a number of counter-examples also exist. Very few studies can be found based on this pair, as Mandarin-English has not been one of the major language pairs in this field. As of now, research on Mandarin-English CS has focused on syntactic aspects. Among the studies conducted on other language pairs, some focused on proposing original syntactic constrains. Others focused on finding evidence to either support or oppose the existing syntactic constraints. Studies on Mandarin-English CS mainly focus on the latter. No syntactic constraints were proposed based on this language pair. To be more specific, with regard to this language pair, linguists found evidence mainly by focusing on syntactic patterns at switching points.

Since Mandarin is not the lingua franca as is English, it is necessary to provide a general introduction to several basic syntactic, morphological and phonological characteristics of this language. Note that the introduction of this language is not the focus of the current study, thus only a brief description is given on several important characteristics along with comparisons to their counterparts in English. One should refer to handbooks on Mandarin basics for a detailed and systematic introduction, such as Chen & Tzeng (1992), Lin (2001), Huang et al. (2014) etc.

2.1.3.1 Brief introduction of Mandarin

Syntax & Morphology

Mandarin is an SVO language, just like English. The subject (S) precedes the verb (V), which in turn precedes the object (O) (Lin, 2001). However, under several circumstances, many Mandarin sentences have no subject. In many other languages, the subject may either be presented or ellipted. In Mandarin, it is a norm for the subject to be absent under certain conditions (Chen & Tzeng, 1992). For instance, regarding climatic phenomena such as thunder, rain, cloud etc., a formal subject is needed in an English sentence. In Mandarin, no subject is allowed. A sample phrase in Mandarin is given as well as its counterpart in English.

(2-10)	**Mandarin:**	下	雨	了。
	IPA:	**/ɕja4/**	**/y3/**	**/lɤ5/**
		pour	*rain*	*PARTICLE*
	English:	'It's raining.'		

Typologically speaking, in contrast to languages such as English, Mandarin stands out as a language without a great number of affixational morphological processes. The average word is not made up of multiple components called *morphemes.* Rather, most words consist a single morpheme, or a single character, which is also a single syllable. A small portion of words consist of two morphemes, or two characters, and thus two syllables. There is very little inflectional morphological complexity in Mandarin. There are few overt syntactic expressions of tenses, subject-verb agreement, case, gender or number markings as there are in inflectional languages. Meanings and the semantic role of various constituents rely on word order and lexicon (Lin, 2001). For instance, the verb 'go' in English changes into different forms under different circumstances, such as 'went', 'going', 'goes' etc. Yet its counterpart in Mandarin, the verb '去

(/tɕʰy4/)' remains exactly the same in all circumstances. To specify that the action took place yesterday, people simply add a time noun 昨天 (/tsuɔ2 tʰjɛn1/), '*yesterday*'[2] in the sentence. To sum it up, while the general lack of inflectional morphological complexity of word formation in Mandarin is apparent from a cursory glance at the language, two specific examples on nouns and determiners are given, to further illustrate this major feature in this language. For example, the word 书 (/ʂu1/), '*book*' and its counterpart in English:

A. Nouns

English	**Mandarin**
book/books	书 (/ʂu1/)
one **book**	一本[3]书 (/i1 pən3 ʂu1/)
two **books**	两本书 (/ljɑŋ3 pən3 ʂu1/)
many **books**	很多本书(/xən3 tuɔ1 ʂu1/)

Table 1 Presentation of the word '书 (/ʂu1/), *book*' in Mandarin and its English equivalent

It is obligatory in many languages to mark nouns for a singular/plural distinction, such as *book/books* in English. In Mandarin, this is unnecessary for most nouns, thus it involves no inflectional morphological complexity within a word. If one needs to express the concept of plurality, it is typically accomplished by a separate word, such as 一些 (/i4 ɕjɛ1/), '*some*' and 很多 (/xən3 tuɔ1/), '*many*'.

[2] The word '昨天 yesterday' contains two morphemes, or two characters, thus two syllables. The morpheme '昨'means last, the morpheme '天' means day. In ancient Chinese, solely the morpheme '昨' can refer to the meaning 'yesterday'. But in modern Chinese (Mandarin), these two morphemes need to appear together to mean 'yesterday'.

[3] In this expression, solely the first morpheme '一' means 'one'. The following morpheme '本' is a quantifier. In Mandarin, most nouns are set with specific quantifiers. Here the noun is '书, book', thus the specific quantifier '本' is applied.

B. Determiners

The simplicity also applies to determiners in Mandarin. Some Mandarin determiners play multiple functional roles vis-à-vis their English counterparts (Ong & Zhang, 2010). For example, in English, the article 'the' denotes definiteness, and articles like 'a' or 'an' denote indefiniteness. Articles like this do not exist in Mandarin. Besides, a single determiner in Mandarin can possess multiple counterparts in English. For instance, the determiner '一个 (/i2 kɤ4/, *one*)' in Mandarin can be translated in English as the numeral 'one', or the indefinite articles 'a' and 'an'.

Phonology

In Mandarin, morphemes are each one syllable long, one character and have one tone. The tone system is comprised of four lexical tones plus a variable (or neutral) tone. These tones can be represented with tone letters, as developed by Chao (1930), illustrated in Table 2. To get a more direct overview of the description of tone 1-4, Chao (1968) also provides a schematic presentation of them, illustrated in Figure 2. The description on the left does not refer to any specific value, it just represents the relative tone level, and these levels are achieved slightly differently from individual to individual.

Tone	Description	Tone letter
1	High level	-
2	High rising	/
3	Falling-rising	∨
4	High falling	\
0	No tone	(blank)

Table 2 Illustration of the description and the letter of each tone

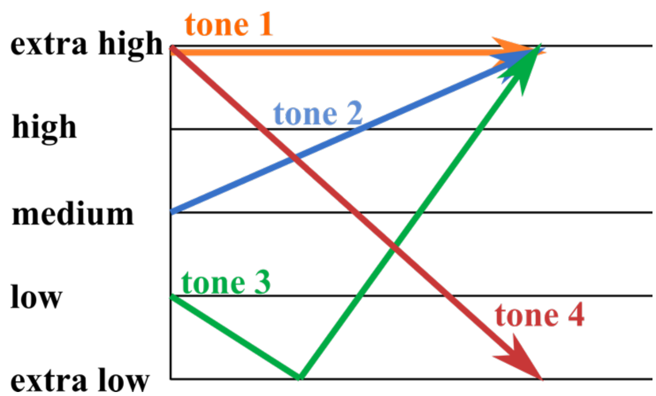


Figure 2 Schematic representation of tones 1-4 in Mandarin.

At first, in written Mandarin, morphemes could only be represented by characters, the logographic writing system. Then, the *Pinyin* system was introduced to the public. This is a newly developed transcription system that uses roman letters to represent the pronunciation of each morpheme. When transcribed with characters, no tones are indicated; when transcribed with corresponding *Pinyin* letters, tones are indicated accordingly with tone letters. The example retrieved from speaker ***03NC05FAX*** illustrates the difference between the two transcription systems[4]:

(2-11) **Characters:**	我	知	道	他	在	新	加	坡。
Pinyin:	wǒ	zhī	dào	tā	zài	xīn	jiā	pō
IPA:	**/uɔ3/**	**/ʈʂɚ1/**	**/tɑʊ4/**	**/tʰa1/**	**/tsaɪ4/**	**/ɕɪn1/**	**/tɕja1/**	**/pʰɔ1/**
	I	*know*		*he*	*in*	*Singapore*		

‘I know he is in Singapore.’

2.1.3.2 Previous studies on Mandarin-English CS

It is mentioned at the beginning of Section 2.1.3 that, few original syntactic constraints are proposed based on Mandarin-English CS. Rather, studies mainly focused on investigating

[4] IPA is offered for the sake of pronunciation illustration. For each syllable, its lexical tone is illustrated immediately after the IPA letters, realized in numbers 1-5. For all the Mandarin syllables in the current study, the same criteria applies.

syntactic patterns at switching points, in order to find evidence to either support or challenge the existing constraints. The current section offers a review on the previous studies.

Lu (1991) examined truly natural bilingual discourses in Mandarin-English CS. He analyzed a one-hour audio-taped recording of a meeting of 12 highly proficient Mandarin-English bilingual officers of the Mandarin college fellowship. His intent was to examine the interrelationship between form and function in discourse. The analysis was conducted from both structural and functional aspects, however, the latter is not discussed here since the current study focuses on the structural outcome. Lu (1991) found that bilinguals tend to let CS fall on certain types of syntactic patterns, namely ADJ+NP, DET+N/NP and V/VP+N/NP (that is to say, bilinguals would produce the former syntactic category in one language, then the latter category in the other language). Lu (1991) pointed out that those were the most frequent types based on his corpus, whether it concerns simultaneous bilinguals[5] or those who are dominant in either one of the languages. Even though no specific examples of the utterances were given in his paper, Lu (1991) suggested that his results, as mentioned above, are all equivalent syntactic structures of Mandarin and English. Therefore, these findings seem to support Poplack's (1980) *Equivalence Constraint*. Lu (1991) suggested two reasons behind the existence of the patterns. The first is ease of expression, that human beings tend to choose the expression that they are most familiar with. The second is due to the influences of the interlocuters. However, he did suggest that future studies should take a closer look into the patterns to test whether there are contradictions to the syntactic constraints.

Another early work (Tan, 1988) on a multilingual household finds that DET+N/NP appears the most frequently. Surprisingly, it is always "Mandarin DET+English N/NP", and never the other way around. In this study, a solid explanation as to why this phenomenon exists is not presented. Tan's (1988) results support a number of syntactic constraints proposed earlier. For instance, the *Equivalence Constraint* proposed by Poplack (1980). As an example, a switch takes place between the Mandarin determiner '他们的 (/tʰa1 mən5 tɤ5/, *their*)' and the English

[5] Simultaneous bilinguals are children who are exposed to more than one language prior to age three. They develop two or more languages equally, or nearly equally, through exposure and frequent opportunities to use each language.

noun ‘parents’ in the following utterance. Two surface structures map onto each other in these two languages, this is allowed by the *Equivalence Constraint*.

(2-12) **Utterance:** 他 们 和 他 们 的 parents.

Pinyin: tā men hé tā men de

IPA: **/tʰa1/ /mən5/ /xɤ2/ /tʰa1/ /mən5/ /tɤ5/**

they *and* *their* *parents*

‘They and their parents.’

Following Tan (1988) and Lu (1991), Ong & Zhang (2010) examined a database obtained from 140 simultaneous Mandarin-English bilinguals aged from 16 to 25. The authors came to the same result that the "Mandarin DET+English N/NP" pattern appears most frequently. They also offered an explanation from the perspective of language activation and inhibitory control (Green, 1998a, 1998b, 2007). They claimed that the existence of the "Mandarin DET+English N/NP" pattern is due to the influence of both the *Lemma versatility filter* and the *Grammatical feature filter*. The *Lemma versatility filter* refers to the principle of speech economy proposed by Clyne (1991), according to which, bilinguals subconsciously prefer simpler forms of speech from either lexicon. Section 2.1.3.1 presents that Mandarin determiners allow for more economy of speech compared to English ones, which explains why people tend to use Mandarin determiners. Furthermore, the *Grammatical feature filter* indicates that bilinguals prefer to use the language that possesses more grammatical features. To be more specific, Mandarin nouns do not inflect for number whereas English nouns often have a plural form. Thus, from an efficiency standpoint, people tend to use English nouns (for more details, refer to Ong & Zhang (2010)). Both filters were activated by the semantico-syntactic and morpho-syntactic dissimilarities between Mandarin and English, which is why the "Mandarin DET+English NP" pattern is the most frequently attested switch point.

The above-mentioned studies all find evidence that supports the existing syntactic constraints. Zheng (2005), on the other hand, studied natural discourses of Chinese-Australian bilingual children by conducting interviews at three primary schools in Victoria. He found certain CS patterns that contradict several syntactic constraints such as the *Free Morpheme Constraint* proposed by Poplack (1980). This constraint claims that codes may be switched after any constituent in discourse provided that constituent is not a bound morpheme, yet Zheng (2005) found several instances of CS between an English lexical form and a Mandarin bound morpheme. For example, '的 (/tɤ5/)' is a grammatical bound morpheme in Mandarin, when following another lexical bound morpheme '有趣 (/joʊ3 tɕʰy4/)', they become an adjective '有趣的 (/joʊ3 tɕʰy4 tɤ5/), *interesting*'. According to the constraint, a switch cannot take place between these two morphemes, yet counterexamples are found in Zheng's (2005) corpus. A solid explanation on why these contradictions exist was not offered by the author.

(2-13) **Utterance:**	没	有	很	interesting	的。
IPA:	**/meɪ2/**	**/joʊ3/**	**/xən3/**		**/tɤ5/**
	not	*have*	*very*		*(bound morpheme)*

'That is not very interesting.'

Syntactic studies focusing on Mandarin-English CS are limited. The current research means to follow previous studies and investigate the syntactic patterns at switching points, in order to verify the previous outcomes and in hopes of finding new patterns. To do so, this study relied on a corpus[6] and conducted research on a rarely studied bilingual community in Singapore and Malaysia. This is the first objective of the current study. Note that this is not the main objective of this study. First, it means to provide additional information for the prosodic analysis, for example, the result of syntactic patterns might provide information on the interference that might affect the prosodic outcome. Second, it means to make the most of the corpus, and examine whether the frequent syntactic patterns are aligned with findings in previous studies. In CS studies generally, Mandarin-English CS included, the problem remains in the field. Shenk (2006) pointed out that syntactic constraints are able to significantly advance our knowledge of CS behavior, but many of them are flexible with a number of attested counterexamples. Syntactic constraints alone might not be able to explain the CS behavior in a comprehensive way. Section 2.2 is dedicated to a comprehensive introduction to the new prosodic aspects of CS studies, which paves the way for a possible solution to this problem.

[6] This refers to the SEAME corpus purchased and applied in the current study. It is thoroughly presented in Chapter 3.

2.2 Prosodic approaches

Shenk (2006) was one of the first to propose that in CS studies, in addition to syntactic constraints, prosodic constraints not only must exist but are also equally important. She applied the information-based approach from monolingual prosody studies, and used its fundamental unit, Intonation Unit (IU), to analyze a natural CS speech corpus. The IU has been characterized as "a sequence of words combined under a single, coherent intonational contour that plays an important functional role in the production and comprehension of language" (Chafe, 1994, p.62). The following example is retrieved from speaker ***NI18MBP*** of the current database. The speaker is participating in an interview, and is talking about his career choice after graduation. It is transcribed with respect to the transcription convention introduced in Chapter 3, each line represents a single IU:

(2-14) a. …(0.9s) I don't know,

b. .. I don't –

c. … think my =

d. … parents can

e. .. stop me =.

Most IUs contain 2-5 words, they usually do not demand a specific syntactic structure. They can be truncated at any point, even within a word, thus sometimes, an IU could contain only one or two syllables. This will be further discussed in Section 2.2.2, as well as the features to identify an IU.

Shenk described her result as *shocking*: the vast majority of CS happens at IU boundaries, and very few take place within IU. Thus she came to the conclusion that there exists a prosodic constraint that controls CS behavior, in which IU is the fundamental unit: bilinguals prefer to conduct CS at IU boundaries rather than inside IU. Shenk (2006) also pointed out that the prosodic constraint does not deny the validity of syntactic constraints, rather, together they offer a more comprehensive explanation of CS behavior. To better understand this, several related issues need to be illustrated: the information-based approach from the monolingual prosody studies, the fundamental unit IU, and how they are applied by Shenk to bilingual CS studies. The three critical issues are explained from Sections 2.2.1 to 2.2.3.

2.2.1 Information-based approach

2.2.1.1 General background of prosodic studies

In the field of monolingual prosodic studies, three strands of research now share a state of "more of less peaceful coexistence" (Couper-Kuhlen, 1986), developed from three different methodological approaches. Firstly, the *grammar-based approach* focuses on finding the evidence to support that prosody is a part of syntax, using mostly introspection and constructed examples. Grammar is considered to be the central processing unit, whereas the definition and identification of prosodic functions seem more gradient rather than categorical (Halliday, 1967; Halliday & Matthiessen, 2013; Hirschberg & Ward, 1992; Pierrehumbert, 1980; Ward & Hirschberg, 1985). Secondly, the *contextualization-based approach*, based on close observation of real and natural discourse, stands in complete contrast to the first approach and thinks of prosody as totally independent from grammar. This strand focuses on investigating how prosodic units contribute to the prediction of the development of the speech (Auer & Di Luzio, 1992; Cook-Gumperz & Gumperz, 1976; Gumperz, 1982). Finally, the information-based approach also stands in contrast to the *grammar-based approach*, and agrees with the *contextualization-based approach* that prosody and grammar should be treated independently. Based on natural discourses, this strand focuses on establishing the appropriate relationship between the prosody and the consciousness. In other words, it seeks evidence to relate prosodic units in utterances with information flow in human minds (Chafe, 1979, 1980, 1993, 1994).

Given the importance of the information-based approach and its relative novelty in bilingual studies, the current section offers a rather comprehensive introduction on this approach. Other than that, this study also aims to explain why this approach is most suitable for the current research.

2.2.1.2 Information-based approach

Human beings engage in natural discourse activities all the time. The activities can be non-reciprocal speech such as monologues, where one single person talks by himself. There are also semi-reciprocal speeches where interaction remains low, such as interviews during which the interviewer speaks much less than the interviewee. There are also interactive speech acts, such as conversations and discussions, in which all participants fully devote themselves to the interaction. When participating in all these kinds of speech acts, three different kinds of states exist in the human mind. The first one is the focal state, that is, an active state where lies the limited amount of information that people focus on at one time. Then under the focal state is the peripheral state, this is a semi-active state which provides context for the current focus and suggests opportunities for next moves. Finally, the unconscious state is an inactive state where lies a vast amount of information, some of which will be brought up to the front to reach the focal state, whereas some will stay unattended for the moment (Chafe, 1994). Thoughts and utterances are created during the continuous and constant interplays and interchanges between these states. For a clearer view of the relationship between the states and the utterances, based on Chafe's (1994) description, the current study developed a schematic diagram, see Figure 3:

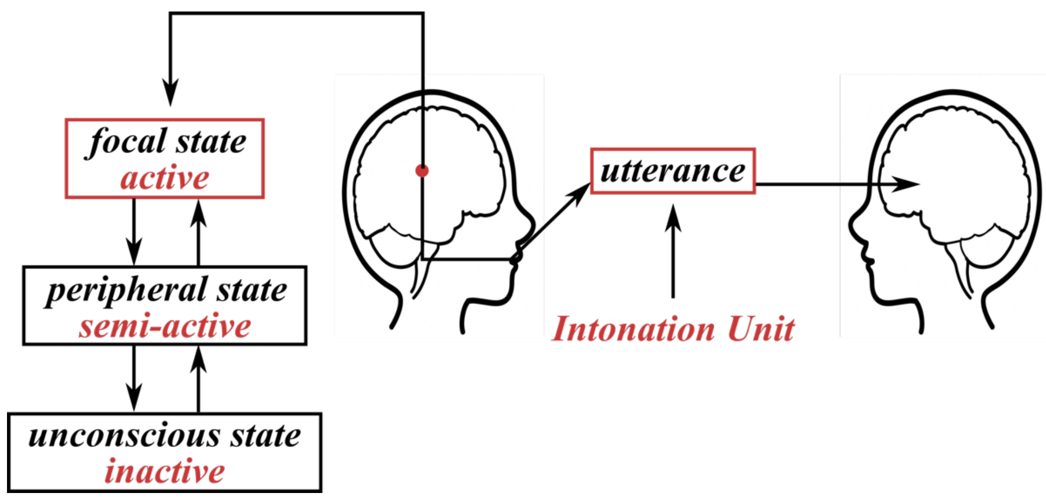


Figure 3 The three states of information flow in mind and expression in IUs

Figure 3 illustrates that the three states exist at the same time in the human mind, yet only the focal state is activated and thus, is closest to the brain surface. The other two states support the first one. Pieces of information existing in the focal state get expressed immediately in utterance form. Note that the three states coexist in a dynamic way when a human being is engaging in speech acts. That means the information between these states is constantly being exchanged. Just like the flow of waves that never stops in an ocean, when engaging in speeches, the flow of information never stops in the human mind. That is the theoretical basis of the information-based approach. As the founder of this approach, Chafe (1979, 1980, 1993, 1994) proposed to establish a corresponding relationship between the flow of information in the human mind and the prosodic units in utterances. Note that the utterances created by people can be recorded and transcribed. A number of their aspects can be studied such as fundamental frequency, pitch, intensity, etc. Thus, with this relationship established, linguists are able to investigate what happens in human minds via physical data.

2.2.1.3 Why this approach suits this study

Previous sections mentioned that the information-based approach is developed from truly natural discourse, and not from introspections or constructed examples. It has been proved to be an appropriate approach to study prosodic units of natural human speeches through many previous research (Chafe, 1994; Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1993; Shenk, 2006; Urrea, 2012). Therefore, it is adopted in the current work. Approaches and theories in the field were mostly developed based on the English language, some of them point out their applicability to other European languages (Brazil, 1985; Cook-Gumperz & Gumperz, 1976; Crystal, 1969, 1975; Halliday, 1967; Halliday & Matthiessen, 2013; Nespor & Vogel, 1983; Pierrehumbert, 1980; Selkirk, 1984). Among the number of scholars who studied approaches in the field of prosodic studies in natural human speech, Chafe was the only one who specifically pointed out that his information-based approach is universally applicable to languages and even tonal languages. Since Mandarin is a tonal language and the target language of this study, this approach was best suited the current study. Note that the applicability of this approach to languages other than English, including Mandarin, has been supported by various studies throughout time (Du Bois et al., 1993; Shenk, 2006; Tao, 1996). Finally, from a more practical point of view, the information-based approach remains one of the most popular approaches applied in prosodic studies of natural discourse. More importantly, in the recently developed field of prosodic constraints in CS, this is the approach that has been applied the most, and if anything, the only.

2.2.2 Intonation Unit

A crucial fact about human speech is that it is not constituted of continuous, uninterrupted flow, but rather, it is expressed in spurts. Of course, one major reason for these spurts is physiological in nature: human beings need to breathe. Apparently breathing is not the only thing that segments human speech, because if that were the case, interruptions of vocalization would be done in regular intervals (Chafe, 1994). The truth is, we speak at a natural pace. Our speech is segmented into spurts with various lengths at irregular time points, which sounds random and arbitrary. In fact, it is not at all arbitrary. The segments correspond with the active information in the focal state of the human mind, as introduced in the previous section. During the production of natural speech, information in the focal state is expressed as the utterance. While engaging in natural speech, the three states in the human mind interact with each other constantly, and thus information in each state changes very quickly and continuously. This is why the amount of active information in the focal state is usually very limited. Consequently, the corresponding utterance is also quite limited, such that only a few words are included in it, sometimes even just a few syllables. Therefore, this kind of segment cannot represent a sentence or a phrase, at most it is a prosodic unit that represents whatever amount of information that is the most active in the mind at that moment. Chafe (1994) named this prosodic unit the Intonation Unit (IU). This is the fundamental unit used in the information-based approach.

Chafe & Tannen (1987) defined IU as the sequence of words combined under a single coherent intonational contour that plays an important functional role in the production and comprehension of language. It is a unit that exists in mental and linguistic processing, which expresses the information that is processed by the human brain. Previous sections explained that IU represents whatever exists in the active focal state. During an interactive speech, the IU first exists in the speaker’s consciousness, then gets expressed as the utterance, and is finally perceived in the listener’s consciousness. Since IU represents the active information in human minds, it exists in a natural and interactive way in human behaviour, thus it is hard to define a specific length for it. Sometimes an IU contains several words, sometimes it contains merely one or two syllables that do not count as a word. The average duration of a syllable is undefinable. Even Chafe (1994) himself simply proposed the general referential length for a syllable in

English, that is 0.35s. Based on our preliminary work in the current study, our finding agrees with Chafe's since, generally speaking, there are 2-3 syllables in a word, and 2-3 words in an IU. The length of an English IU in the current corpus is roughly 2-4s based on our estimation. As for Mandarin, based on our preliminary work, the length for a syllable is generally 0.3s. Section 2.1.3.1 introduces that a word in Mandarin contains solely one syllable. Thus an IU in Mandarin is roughly estimated to last 1-3s. Note that the above discussion simply means to offer a rough idea of how long an IU should generally be, it does not apply to every IU and it certainly does not set any length constraint. One might say that the fundamental unit of an IU is a syllable. The boundary between two IUs can be between two words, or between two syllables of a word. An example from Shenk (2006) is presented to further explain this.

(2-15) Etta: a. El Bo,

'Bo'

b. resp--

'resp--'

c. se mueve mucho,

'he moves a lot'

d. pero no hace nada.

'but he doesn't do anything'

This example comes from a conversation, in which *Etta* is talking in Spanish about the dancing ability of a nonpresent third party named *Bo*. Each line represents a single IU[7]. This example illustrates that three IUs in *line a, c, d* vary in number of words. The IU in *line b* is clearly a truncated word, *Etta* might be interrupted by someone else or she simply changed her mind while talking. Either way, an IU containing a single syllable was created. Similar situations happen constantly in natural speeches. This is why one cannot simply delimit IUs by punctuation or even by word. Obtaining an accurate and categorical delimitation and identification of an IU is complicated and technical work.

[7] The transcription convention is introduced in Chapter 3 and presented in Appendix B.

With advances in phonetic software, the identification of prosodic units is becoming considerably more accurate. For the fundamental unit in information-based approach, Chafe (1994) proposed a categorical way to identify the IU. The first step is to identify the critical features. Then the second step is to apply the features to identify the IU. According to Chafe (1994), the IU possesses four critical features, namely the *change in fundamental frequency* (perceived as pitch), the *change in syllabic duration* (perceived as the shortening or lengthening of syllables or words), the *alternation of vocalization with silence* (perceived as pausing) and the *change in voice quality of various kinds* (mostly a creaky voice). An introduction to each of the features is given in the following section. Note that the examples for the illustration are extracted from the database of the current research.

2.2.2.1 Change in fundamental frequency

The *change in fundamental frequency* (F0), realized in Hertz, is a key feature in identifying the IU boundaries (Chafe, 1994). Two critical characteristics are necessary to define a change in F0, namely the *pitch reset* and the *pitch declination*. Figure 4 illustrates both characteristics of the change in F0:

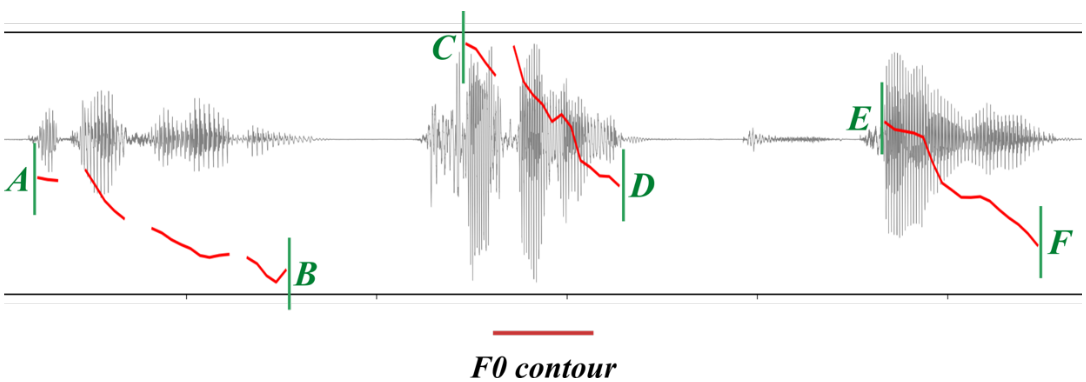


Figure 4 Example of the change in F0 retrieved from speaker ***NI18MBP***

1) Pitch reset

Figure 4 indicates that three separate F0 contours are observed, six pitch values at the beginning and the end of them are marked with letters *A, B, C, D, E, F*. A pitch reset refers to the resetting of the baseline pitch level, which is manifested in the form of rising at the beginning of an IU in contrast with the end of the preceding IU. In this example extracted from speaker

NI18MBP, two pitch resets are observed, the first one is between point *B* and point *C*, the second one is between point *D* and point *E*.

2) Pitch declination

Pitch declination refers to the global tendency for the F0 curve to decline with time, despite local rises and falls (Schuetze-Coburn et al., 1991). In this example illustrated in Figure 4, all three pitch contours possess the declination tendency, the first contour between point *A* and point *B*, the second one between point *C* and point *D*, and the third one between point *E* and point *F*.

Using the *pitch declination* within a single contour and the *pitch reset* between two contours, an IU boundary can be identified. In Figure 4, two IU boundaries are identified: the boundary between point *B* and point *C*, and the boundary between point *D* and point *E*. Note that in the case of Figure 4, we identify two IU boundaries, not three IUs. While the *change in F0* feature can define IU boundaries, it could be the boundary of one single IU or several IUs. For instance, it is possible that only one IU exists between A and B, it is also possible that multiple IUs exist between these two points. Other features need to be applied to make further identification.

2.2.2.2 Change in syllabic duration

Change in syllabic duration, occurring in a matter of seconds, is another major cue to indicate the IU boundaries (Chafe, 1994). For the sake of simplicity, the term syllabic duration is represented by D. According to Chafe's (1994) definition, suppose there is a syllable with the duration of D_x[8], if D_x < 0.15s, it is identified as a short syllable; if 0.15s < D_x < 0.35s, it is a normal syllable; if D_x > 0.35s, it is defined as a long syllable. Figure 5 illustrates the waveform of a Mandarin IU with a complete syllabic duration change retrieved from speaker ***NI20MBP***, English translation given above.

[8] The finding out of this value will be described later.

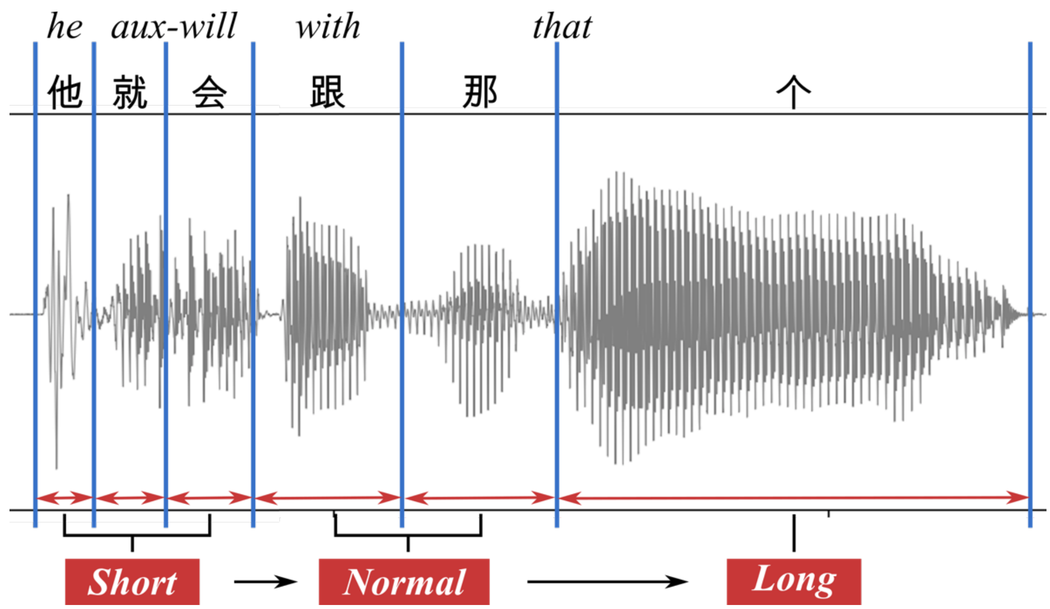


Figure 5 Example of syllabic duration change in an IU retrieved from speaker ***NI20MBP***

Using blue vertical lines, the IU is divided up into syllables. The IU begins with three short syllables (*他 tā* ***/tʰa1/*** *就 jiù* ***/tɕjoʊ4/*** *会 huì* ***/xweɪ4/****, 'he will'*), followed by two normal syllables (*跟 gēn* ***/kən1/*** *那 nà* ***/na4/****, 'with that'*), then it finishes with a very long syllable (个 gè **/kɤ4/**, 'that'). This pattern of acceleration-deceleration is characteristic of many IUs. In other words, an IU proceeds from reduced-length syllables, through normal-length syllables, to extended-length syllables. This may in some instances be the primary evidence for their delimitation (Chafe, 1994). However, a complete duration change is not always necessary in identification, the second part of the change can also define the boundary of an IU: several normal syllables followed by a long syllable always define the end of an IU. Chafe (1994) pointed out that the syllable length does not remain the same for different speakers, it is necessary to take the individual speaking rates into consideration when applying this feature. However, he did not elaborate on the specific procedure. In Chapter 4, this study investigates the speaking rates of the current database, and clarifies how the syllable length is decided for different speakers.

2.2.2.3 Alternation of vocalization with silence (Pause)

Alternations of vocalization with silence, perceived as pausing for a matter of seconds, is another cue that indicates the boundary of an IU (Chafe, 1980). In natural spoken discourse, the placement and timing of pauses convey significant information, which is critical in establishing the discourse production process and the orientation of the ongoing conversational interaction (Du Bois et al., 1993). Pauses have a rather clear presentation, and thus are easy to recognize. According to the convention set by Chafe (1994), a total of three types of pauses exist:

1) Short pause ($t < 0.2s$): it indicates a brief break in speech rhythm. It is a very short, barely perceptible pause which lasts about 0.2s or less, indicated in Figure 6. Four short pauses with the length of respectively 0.2s, 0.1s, 0.1s, and 0.2s are identified in this period.

Figure 6 Representation of four short pauses retrieved from speaker ***05NC10MAY***

2) Medium pause ($0.3s < t < 0.6s$): it indicates a pause of medium length, which is noticeable but not very long. Generally, it lasts between 0.3 and 0.6 second inclusively, indicated in Figure 7. Three medium pauses with the length of respectively 0.5s, 0.5s and 0.4s are identified in this period.

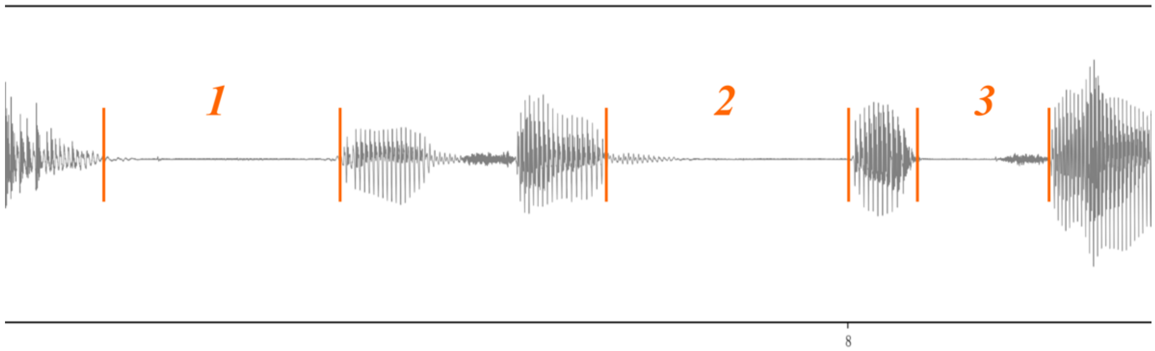


Figure 7 Representation of three medium pauses retrieved from speaker ***NI20MBP***

3) Long pause (t > 0.7s): it represents relatively long pauses which last longer than 0.7 second, indicated in Figure 8. A long pause with the length of 0.9s is identified in this period.

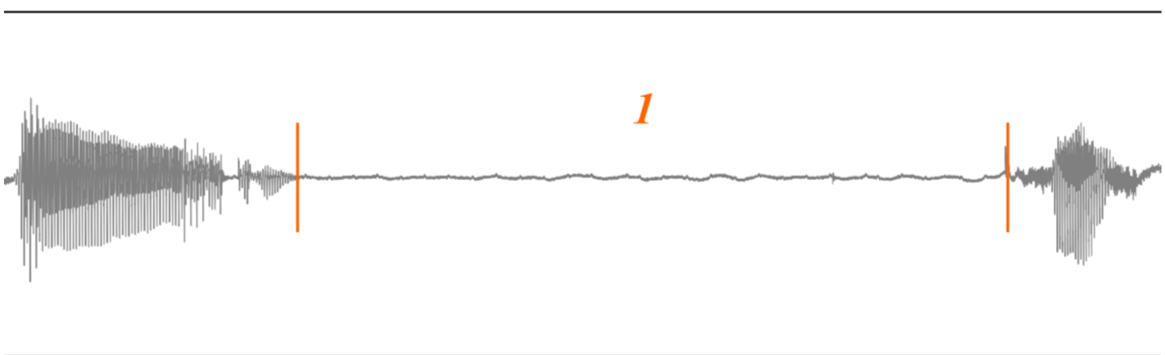


Figure 8 Representation of one long pause retrieved from speaker ***04NC07FBX***

Du Bois et al. (1993) specifically points out that pauses function as a cue to aspects of discourse production and conversational interaction, and not simply as a raw acoustic fact. Therefore, one should always make the distinction between pauses and the other kind of brief silence (lexically or phonologically required silence), such as a voiceless stop inside a word. For instance, the /p/ in the word *carpenter* is a voiceless stop. When pronounced emphatically, a clear brief silence could easily be captured, yet it should not be classified as a pause. Just as the *change in syllabic duration* feature, the possible influence of different speaking rates is discussed in Chapter 4.

2.2.2.4 Change in voice quality (Creaky voice)

According to Chafe (1994), the *change in voice quality* of various kinds works as another cue for IU boundaries, the most common kind is a *creaky voice*[9]. Creaky voice refers to a number of different kinds of voice production, such as low subglottal pressure, compressed vocal folds, skewed glottal pulses and sharp harmonics etc. (Keating, Garellek, & Kreiman, 2015). Generally speaking, it is a voice quality caused by a distinctive phonation type which involves low-frequency vocal fold vibration (Raitio, Kane, Drugman, & Gobl, 2013). Ishi et al. (2008) points out that it can be perceived as a rough quality with the sensation of additional impulses. Chafe describes a creaky voice to be a laryngealization or a "fry". IUs often end and sometimes begin with creaky voice, which provides the clue to their delimitation[10].

Chafe (1994) points out that one can identify an IU with any or all of the features, indicating that under some circumstances some of the features might not be present in an IU. The following section is dedicated to the introduction of prosodic studies in CS research, especially the work of Shenk (2006).

2.2.3 Prosodic studies in CS

Shenk (2006) was one of the first to explore prosodic constraints in CS studies. She studied a Spanish-English CS database containing one hour of natural discourse data from four competent bilinguals of Mexican heritage living in Southern California, yielding a total 782 analyzable IUs.

It is illustrated in Figure 9 that, in Step 1, Shenk applied the features proposed by Chafe (1994) and identified the IU. Then, in Step 2, the original data is divided up into different IUs. Next, in Step 3, Shenk categorized the IUs according to whether they are monolingual Spanish

---

[9] Since the creaky voice is the only kind of voice quality change mentioned in Chafe (1994), it is the one that is discussed in the current study.

[10] Note that "a fry" is the entire description given by Chafe. Other previous studies in CS studies did not expand on this issue either. Thus Section 2.3.2 points out that this might bring subjectivity problem in IU identification. Then in Chapter 4, this study provides physical indices of a creaky voice, as well as how they are implemented in identifying IU boundaries.

IUs (SIU), monolingual English IUs (EIU) or bilingual IUs (BIU). It is marked in Step 3 that, two types of CS exist: 1) between the yellow rectangle and the green rectangle, CS occurs. Since it is at the boundary of two different IUs, it is categorized as a CS external to IU, i.e., CS at IU boundaries. 2) within the gray rectangle that follows, CS also occurs. Since it is inside one single IU, it is categorized as CS internal to IU.

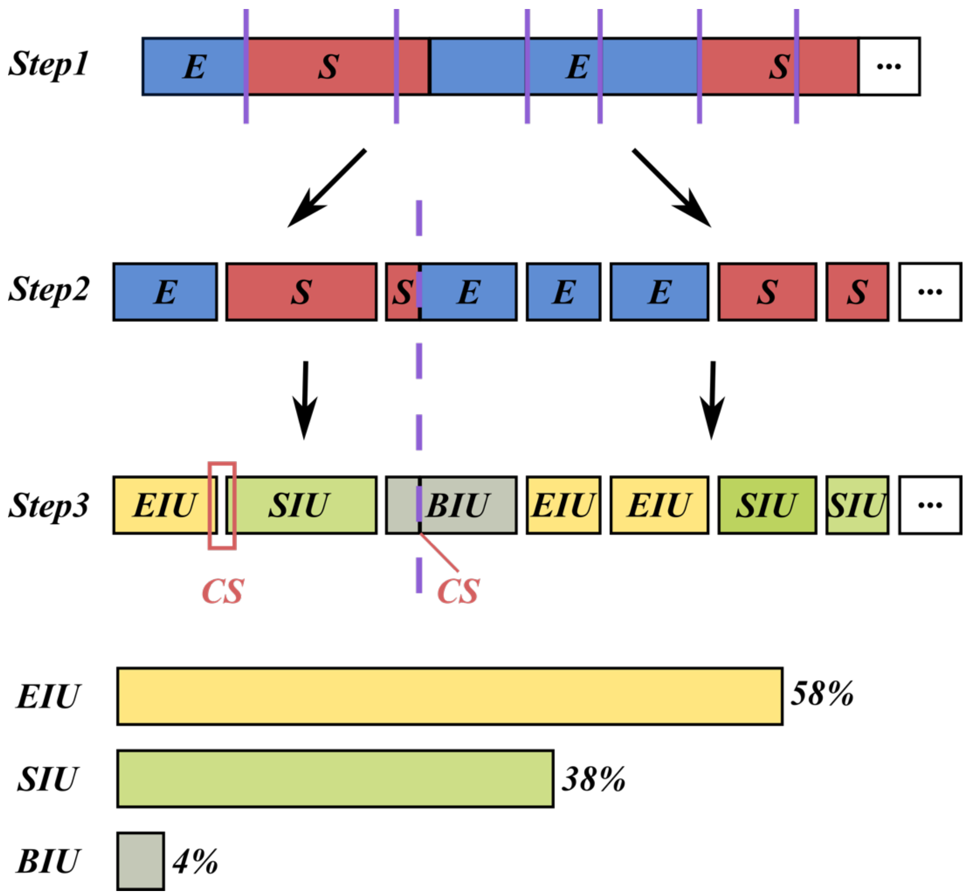


Figure 9 Illustration of the identification and classification of different IUs in Shenk (2006)

Shenk then calculated the percentages of all three types of IUs. Figure 9 illustrates that the monolingual English IU (EIU) takes up 58%, the monolingual Spanish IU (SIU) takes up 38%, and the bilingual IU (BIU) takes up a mere 4% of total IUs. This suggests that speakers are producing monolingual IUs 96% of the time. In other words, bilinguals code-switch massively at IU boundaries. The results led Shenk (2006) to conclude that the most robust boundary correlating with CS is prosodic in nature, and thus the IU should be taken as an important factor when looking into when and where a CS might occur.

The results in Shenk (2006) suggest that prosodic constraints control CS behavior, the fundamental unit is IU in information-based-approach, and bilinguals tend to switch at IU boundaries rather than within IU. Of course, more research on this matter needs to be conducted to support the proposition. The key issue is that syntactic constraints and prosodic constraints should work together to offer an explanation behind CS behaviour. Then again, this new theory still requires support from many more future studies to be confirmed. After that, should there exist a phenomenon that contradicts syntactic constraints, perhaps prosodic constraints could offer an explanation.

Following Shenk (2006), more studies began to apply the information-based approach and the fundamental unit IU in CS studies of different language pairs, with the hope to find evidence to either support or challenge this novel perspective in the field. Cacoullos & Travis (2010) mentioned an unpublished manuscript, in which the author applied the IU in the study of Spanish-English CS and found that people code-switch at IU boundaries 90% of the time. Travis & Cacoullos (2013) also confirmed the significance of applying IU into CS studies. Both Mettouchi (2008) and Vargas (2008), who conducted studies respectively on Berber-French CS and French-English CS, found that CS massively occurs at IU boundaries, and supported Shenk's proposition (2006). Myslin & Levy (2015) studied Czech-English CS and found that people code-switch at IU boundaries 77% of the time. Urrea (2012) looked into Spanish-English CS, her data suggested that 79% of the CS occurs outside of IUs, in other words, at IU boundaries. Manfredi et al. (2015) followed Shenk's (2006) work and found that people do massively code-switch at IU boundaries. However, these last three studies did not turn out to obtain a BIU rate as low as Shenk's original work, all the authors have one point in common: based on their corpus, those who switch inside an IU are those with single-word CS. The authors tend to consider these situations as exceptions or special cases.

To sum up, all these studies on different language pairs suggest that the IU does play a significant role in CS behaviour, and prosodic factors could set certain constraints in CS utterances. Although it is still a novel theory that needs more verifications from many more language pairs. Since it has not been verified in Mandarin-English yet, the current study means to provide evidence in this aspect. Thus, the second objective of the current study is to investigate the Mandarin-English bilingual corpus and provide evidence on whether the prosodic constraint exists in this language pair. This is the main objective of the current study.

2.3 Problems

Section 2.1 and Section 2.2 illustrate the development of CS studies in the syntactic and the prosodic aspect respectively. The objectives of the current study concerning these two aspects are also presented accordingly. Yet three problems exist in the field of prosodic studies: the debate over the difference between single-word CS and borrowing, the subjectivity problem in prosodic measurements that exists in previous methodology, and the lack of a parameter to help estimate the appropriateness of a database to this specific analysis. The current section discusses all three problems, by defining the position on the first problem, and proposing the solutions for the second and third problems. Providing solutions to help solve these two problems is the third objective of the current study.

2.3.1 Single-word CS and borrowing

CS is defined as a behaviour that consists of fragments alternating from one language to another within a single utterance. Concerning single-word switches though, there has been a long-standing debate over whether they should be classified as instances of CS or borrowing. Since the current research focuses on studying CS behaviour, should single-word switches be categorized as instances of borrowing, they would no longer count as the target of the current research. So, it is crucial to examine this issue, and define the study's position on it prior to data processing, in line with previous studies in the field.

Two approaches take two completely different stances on this issue. One suggests that CS and borrowing are two fundamentally distinct processes, and single-word insertion should be classified as borrowing rather than CS (MacSwan & Colina, 2014; MacSwan, 1999; Poplack & Dion, 2012; Sankoff & Poplack, 1981). The other approach sees CS and borrowing as part of the same diachronic continuum, and makes no distinction between single-word CS and phrasal CS. (Myers-Scotton, 1993, 2002, 2006; Thomason, 2003; Van Coetsem, 2000).

For the former approach, researchers find two characteristics that distinguish CS from borrowing. The first one is the number of words contained in the alternated fragments. Sankoff et al. (1990) pointed out that to define a CS, the fragments from both languages should consist of more than a single word. When a single word of one language appears in a sentence that is otherwise entirely in the other language, this is considered to be the result of borrowing. The second characteristic is the difference between the processes involved in each phenomenon. CS requires both languages to retain their own grammatical rules, while borrowing involves the grammatical structure of one language only (the recipient language). The other language plays solely an etymological role (Poplack & Meechan, 1995). Poplack and Meechan (1995) identified two types of borrowing, established borrowing and nonce borrowing. Established borrowing is "the adaptation of the lexical material to the morphological and syntactic (and usually, phonological) patterns of the recipient language" (Poplack & Meechan, 1995, p.208). Nonce borrowing is defined as an insertion from another language that happens a single time, done by a speaker in a reasonably representative corpus. Thus, in this approach, a true single-word switch is usually considered to be an established borrowing. If this switch is not accepted by a bigger community, and is just a spontaneous insertion by one speaker, it would be categorized as nonce borrowing.

However, the other approach claims that borrowing and CS should not be considered as two distinct elements (Heller, 1988; Myers-Scotton, 1993). Researchers who agree with this approach suggest that borrowing arises originally as CS and these two phenomena are part of the same developmental continuum. A description in Myers-Scotton (2006) illustrates clearly the relationship between CS and borrowing:

> *"There is a continuum of embedded language*[11] *elements in bilingual clauses, with single words as one end point and full phrases as the other. Further, many singly occurring words that are codeswitches could (and do) become established borrowings if they are adopted by trend-setters." (Myers-Scotton, 2006, p. 254)*

Myers-Scotton (2002) underlined that in both CS and borrowing, there is interaction between two languages and both languages are active. She also pointed out that these two phenomena undergo the same morphosyntactic procedure, and there is no difference between single-word CS and phrasal CS.

---

[11] Myers-Scotton (1993) labels the language that is less active and does not play a dominant role in CS as embedded language.

There have always been debates on how to treat CS and borrowing. Both approaches have received great support from a number of researchers (Backus, 1996; Park, 2006; Treffers-Daller, 2005). Regarding research on CS studies that apply the information-based approach (though it began quite recently and the number of studies is relatively small), researchers would still choose to follow either of these two approaches to deal with the issue of single-word switches. For instance, Urrea (2012) reviewed the approach that suggests a clear distinction between CS and borrowing, and pointed out that there exists controversy over this issue. She made a “conservative move” and chose to exclude all single-word English origin nouns from the count of CS. However, she did not disclose the reason why she only excluded the *English origin nouns*, but included other single-word switches such as verbs, adjectives, etc. Cacoullos (2010) also followed this approach and chose to exclude all single-word English-origin nouns without specifying why. On the other hand, some researchers, such as Myslin & Levy (2015) and Manfredi et al. (2015) conducted their studies using the other approach. They made no difference between single-word switches and phrasal switches, thus no category was excluded from their studies. All kinds of switches were investigated.

Even though these studies chose to follow different approaches and made different decisions concerning their data processing, it did not affect their conclusions: they all came to the conclusion that IU plays a significant role as the prosodic constraint in CS behaviours. Therefore, based on the literature, the controversy over CS and borrowing does not have a great impact on CS studies that apply the information-based approach. The current study follows Myers-Scotton’s position on this issue, and thus all kinds of switches are included in the count of CS.

2.3.2 Subjectivity problem of the methodology

Section 2.2.2 introduces that the IU possesses four critical features. In order to identify the IU, one first needs to process the audio data and identify the features, then apply the present features to help determine IU boundaries. Several problems are observed in the methodologies of previous studies, including Chafe (1994) who proposed the approach, and the subsequent studies who applied the same approach. The problems concern 1) the principled way to identify the features, and 2) the application of the features, such as choosing the number of necessary features for IU identification, deciding which features to use to identify an IU, and evaluating what the alternative options are when one or more of the features do not possess a clear presentation. If these problems are not resolved, the analysis might rely on the researcher's experimental judgment which brings along a subjectivity problem. The problem brought on by this downside is that different researchers might come to different results despite using the same database. One then loses confidence in the conclusion. The current section conducts a discussion on this subjectivity problem.

2.3.2.1 Subjectivity problem on feature identification

Among the four features, the identification of the *change in fundamental frequency* and the *pause* does not suffer much from the subjectivity problem because the characteristics of both features are straightforward. Section 2.2.2 presents that for the *change in fundamental frequency* feature, one simply needs to make a clear reading of the pitch reset and the global declination tendency. For the *pause* feature, one needs to listen to the audio while reading the waveform to avoid mixing up the voiceless stop and the actual pause. By respecting the above mentioned procedures, the judgments should not differ from one researcher to another.

The subjectivity problems exist in the identification of the *change in syllabic duration* and the *creaky voice*. When identifying the *change in syllabic duration*, the problem arises from two aspects. The first one concerns the delimitation of the syllables and the second one has to do with the duration categorization. With regard to the first aspect, syllables are very small units that take a very short time to produce, thus when reading the waveform, it is difficult to clearly see the boundary between two syllables. Listening to the corresponding audio can solve most of the problem. Syllables can be assigned to the waveform accordingly. Yet under certain circumstances, the problem remains. See Figure 10 for an example (sample period in Section 2.2.2[12]).

Figure 10 illustrates that there exists a relatively long voicing period between the syllable *那 (nà, **/na4/**)* and the syllable *个 (gè, **/kɤ4/**)*, marked by blue line A and dotted blue line B. With

[12] It is exactly the same example as in Section 2.2.2, where the IPA of all the characteristics are indicated. Here we focus on the discussion of the subjectivity problem.

the help of listening to the audio, the starting point of the latter syllable 个 *(gè, /kɤ4/)* is initially decided at the blue line A. Yet, solely based on listening to the audio, different researchers could make slightly different decisions on the specific position of this blue line (any position between A and B). Delimitation of syllables is the first issue that is affected by subjective judgment.

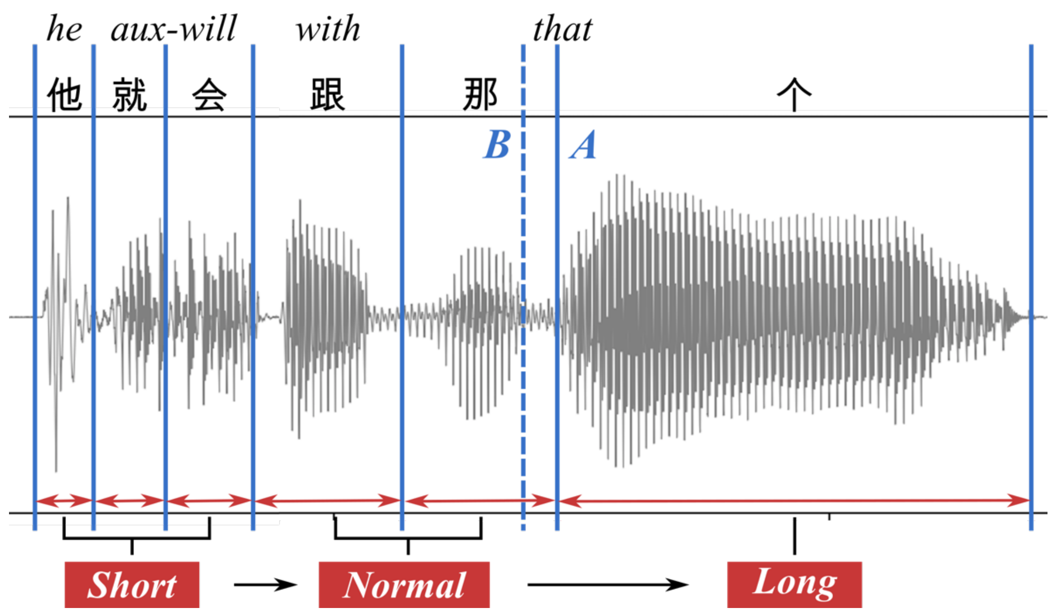


Figure 10 Example of syllabic duration change in an IU retrieved from speaker ***NI20MBP***

Concerning the second problem, the duration categorization, one obvious fact in natural discourse studies is that people speak at different rates. The syllabic duration is therefore different from one individual to another. It is possible that the referential syllabic duration for short, normal and long syllables (presented in Section 2.2.2) proposed by Chafe (1994) can represent the average speaking rate of people. Nevertheless, it is not applicable to every single person on earth. Even though he mentioned that the duration needs to be adjusted according to the speaking rate of different speakers, a practical procedure for this was not proposed. Other previous studies (Shenk, 2006; Tao, 1996; Urrea, 2012) did not address this issue. Judging the categorization thus depends on the experience of the researcher.

The other feature that suffers from the subjectivity problem is the *creaky voice*. Previous studies stated that a creaky voice possesses "obvious difference" from a regular voice. Without further specific criteria, the researcher must decide based on experience whether this *obvious difference* exists each time a possible creaky voice is observed.

2.3.2.2 Subjectivity problem on feature application

Previous studies have also potentially suffered from the subjectivity problem with respect to the application of features. For instance, when it concerns how to apply features to identify the IU, Chafe (1994) mentions that the identification of an IU may involve any or all of the features, without further explanation (Section 2.2.2). Tao (1996) directly pointed out that one single salient feature may be sufficient to clearly mark the boundary, with no statement on which salient feature it might be. Shenk (2006) categorized the *change in fundamental frequency* and the *change in syllabic duration* to be the two primary cues, and the *pause* and the *creaky voice* to be the two secondary cues. She then claimed that one can accurately identify an IU with one primary cue and one secondary cue, although she does not provide further illustration or explanation for the reason behind this categorization. She does not provide guidance on the procedure to take under specific circumstances. For example, what if there are only secondary cues and no primary cues? Other previous studies did not address this issue.

2.3.2.3 Summary

The current section illustrates that the subjectivity problem has a certain impact on both the identification and the application of features. This indicates that the methodologies of previous studies might be problematic. All these problems need to be either explained or solved before conducting the analysis. The current study makes an attempt to develop a systematic methodology to reduce the subjectivity problem as much as possible, and to increase the accuracy of IU identification. This allows the results of the current study to be more reliable.

2.3.3 Parameter for the appropriateness of the database

Shenk (2006) first processed 1h of Spanish-English bilingual audio data and segmented the whole audio into different IUs: SIU, EIU or BIU (Section 2.2.3). Then she calculated the percentage of BIU among all three types of IU. Since the BIU percentage is only 4%, Shenk suggested the existence of the prosodic constraint in CS, and concluded that bilinguals massively code switch at IU boundaries, rather than within the IUs.

Following Shenk (2006), Urrea (2012) studied 4h of Spanish-English bilingual audio data of 28 participants for her doctorate dissertation. Just like Shenk, Urrea also processed her audio data first and segmented the whole audio into different IUs: SIU, EIU or BIU. The next step she took differs from the method chosen by Shenk (2006). Urrea did not simply calculate the BIU percentage. Rather, she located all the CS in the whole audio and put them into two categories: one is the CS within the IU (in other words, CS in BIU), the other is the CS occurring at IU boundaries. She then calculated the percentage of the CS in BIU among all CS and obtained a result of 21%. According to Urrea, this result indicates that most of the CS takes place at IU boundaries, and only a small part of them occur within IUs. Thus, she supports Shenk's (2006) suggestion that the prosodic constraint plays an important role in controlling CS behaviours.

However, Urrea then made a comparison between her result and Shenk's result. She directly compared her 21% (that is CS in BIU among all CS) to Shenk's 4% (that is BIU among all IU). She pointed out that Shenk's result is far too low, then indicated this might be because Shenk's database is community specific. This conclusion might be problematic and is discussed in the following paragraph.

Both Shenk and Urrea applied the information-based approach and the fundamental unit (IU) in their analyses. In fact, an analysis like this sets certain demands on the database: 1) it needs to be a truly natural bilingual speech; 2) it needs to contain a sufficient amount of CS. Note that bilinguals do not code switch constantly in a natural speech. If there are only a few CS in several hours of audio, analyzing this audio does not promise any persuasive outcome[13]. Considering the two demands mentioned above, Urrea's comparison might be problematic, for two reasons. First of all, based on the introduction above, Urrea and Shenk applied different calculating methods on different objects[14], thus their results are bound to be essentially different. Secondly, Shenk and Urrea based their studies on two different databases. Both of them studied Spanish-English bilinguals, but the databases come from separate communities, using different collecting methods, at different times. The critical issue is: neither Urrea nor Shenk provided information on whether their databases met the two demands. Thus we do not know whether

[13] The specific reason is discussed in Section 5.1.

[14] Their calculating methods and objects are detailed in the first two paragraphs respectively in the current section, then Section 5.4 conducts a detailed discussion on this issue.

one database suits the analysis more, or whether one database is more *community-special*. The reason why neither research, nor any other previous research in this field provided such information might be that thus far, there is not an adequate parameter in the field that helps measure the appropriateness of the database for this specific analysis. This parameter needs to take the situation of the IU and the CS in the audio into full consideration, and present a universal parameter mode.

Based on a deeper investigation of the current database and the result, this study means to propose a parameter to measure the appropriateness of a database for conducting CS analysis that applies the information-based approach. Note that a database could be suitable for all sorts of research, the parameter proposed in this study only focuses on this particular analysis. It cannot be applied as a measure of appropriateness for other research. This parameter is proposed in the current study and will require support from future studies. In future work, if authors are able to ensure the appropriateness of their database under this parameter, it could help increase the persuasiveness of their results. In addition, the platforms that sell corpora could also provide their databases with this parameter. This could help researchers make more adequate choices when making a purchase.

To sum up, three problems are discussed in the current section. Section 2.3.1 points out the debate over the definition of single-word CS and borrowing. This study has defined its stance on this issue. Sections 2.3.2 and 2.3.3 point out respectively the subjectivity problem and the lack of parameters on database appropriateness, which exist in previous methodology. The current study aims to provide solutions on these problems as that is the third objective.

2.4 Summary

This chapter discusses the CS research in syntactic and prosodic aspects, concerning both the development and the problem, and proposes three objectives.

Section 2.1 discusses the development of CS studies in syntactic aspects, and the development in research of the Mandarin-English language pair. Since the research on this language pair is rather limited, this study means to follow previous studies and investigate the syntactic patterns at switching points. Since the current database is collected from a bilingual community that was rarely studied before, new patterns are expected to be found. This section also points out the general problem that exists in syntactic aspects of CS studies, that is although syntactic constraints are able to significantly advance our knowledge of CS behaviour, they cannot explain CS behaviour in a comprehensive way.

Section 2.2 discusses the new prosodic aspect in CS studies. Shenk (2006) was the first one to introduce the information-based approach from monolingual studies to bilingual studies. She applied the fundamental unit (IU) in this approach to analyze her Spanish-English database and found that bilinguals massively code switch at IU boundaries. Thus, she suggested that in CS, the prosodic constraint is as important as the syntactic constraint and controls this behaviour. Shenk's finding received great support from certain subsequent studies on other language pairs, although as a newly developed theory, it still expects further support from many more language pairs. This study means to offer support from the Mandarin-English language pair, and provide evidence to support this theory.

However, this newly developed research field faces certain problems, and Section 2.3 discusses three major issues. The first problem concerns the long-running debate over the definitions of single-word CS and borrowing. Section 2.3.1 presents the theory of different approaches over this problem, and aligns the current study's stance with Myers-Scotton's (2002) approach, making no difference between the two concepts. The second problem refers to the subjectivity problem in the methodology of previous information-based approach studies. This problem is thoroughly discussed in Section 2.3.2, the current study means to develop a systematic methodology to reduce the subjectivity problem as much as possible. The third problem is the lack of a parameter to help measure the appropriateness of a database for studies that apply the information-based approach. Section 2.3.3 presents that this has caused certain misunderstandings in previous studies, like the comparison that Urrea (2012) made between her work and Shenk's (2006) work. The current study proposes a parameter, which requires further research from future studies.

Based on the review above, three objectives are proposed:

4) Investigate syntactic patterns at switching points, find whether there is evidence that coincides with previous studies, then look for novel patterns (refer to Section 2.1).

5) Find whether there is evidence of prosodic constraints that provides support for Shenk's proposition from the Mandarin-English language pair (refer to Section 2.2).
6) Develop solutions for the subjectivity problem and the database appropriateness problem in the approach, which could help make the results more conclusive (refer to Section 2.3).

Chapter 3 Database Introduction

This chapter is dedicated to detailing the database used for the analysis. First, the appropriate features of a database that are suitable for the current research are described in Section 3.1. Then in Section 3.2, a general introduction is given on the original corpus and on the specific database used for the current study. Finally, Section 3.3 is dedicated to the introduction of the transcription convention adopted in this study.

3.1 Database features

The database requirements for this study were that it must possess two critical features. One concerns the data itself, the other pertains to the participants. With regard to the former feature, since the current research aims to study natural bilingual human speech, the database needs to contain natural discourses, and more importantly, spontaneous CS. Regarding the latter feature, the participants need to be highly-proficient Mandarin-English bilinguals. To be more specific, they must be native in both languages, neither one of them can be a second language. It is crucial for the current study that these two requirements are respected.

Until recently, very few spontaneous code-switching speech data corpora existed. Most corpora remain read-speech or induced speech. One obvious reason for this is that it is much more time-consuming to record natural data. Then again, a read-speech corpus does not usually reflect the true nature of CS because participants simply read sentences from newspapers, magazines, internet blogs, etc., it is then neither spontaneous nor natural.

The South East Asia Mandarin English code-switching (SEAME) corpus, on the other hand, is a completely spontaneous code-switching corpus, collected at Nanyang Technological University (in Singapore) and Universiti Sains Malaysia from 2009 to 2010, participants are mainly students and staffs of the universities. The SEAME corpus fulfills perfectly all the needed requirements for the current study and for that reason was purchased from Linguistic Data Consortium.

3.2 SEAME corpus and current database

The SEAME corpus is a large-scale code-switching corpus, collected by Lyu et al. (2015) and released through Linguistic Data Consortium (LDC) for research on code-switching speech recognition and other related topics. Lyu et al. (2015) claimed that the SEAME corpus is the first large-scale spontaneous Mandarin-English code-switching speech corpus to exist. According to their description, it would work well for CS studies such as robust acoustic modeling and code-switching in conversational speech. Thus, it was adopted for analysis in the current work.

The original corpus provided by the SEAME team contains both conversations and interviews. Without any editing, the whole session of each speaker was recorded and released, thus the audio contains both the speech of the speaker and the interlocutor. However, since the SEAME team used separate channels for the recording, solely the speaker's speech is clear and transcribed, the interlocutor's is not. In interviews, the interviewer only asked several short questions, thus the majority of the audio is the speech of the interviewee, which is the target speaker. In conversations, the amount of speech between two people basically break even. For each audio, the speech of the target speaker takes up approximately a half of the audio length. With conversations and interviews combined together, there are a total 156 distinct speakers in 178 hours of recordings (as well as the corresponding transcriptions) that were released. For the majority of the speakers, the audio lasts approximately one hour. To form the specific database from the original corpus for the this study, two major aspects need to be considered: data selection and participants.

3.2.1 Selection of analyzing data

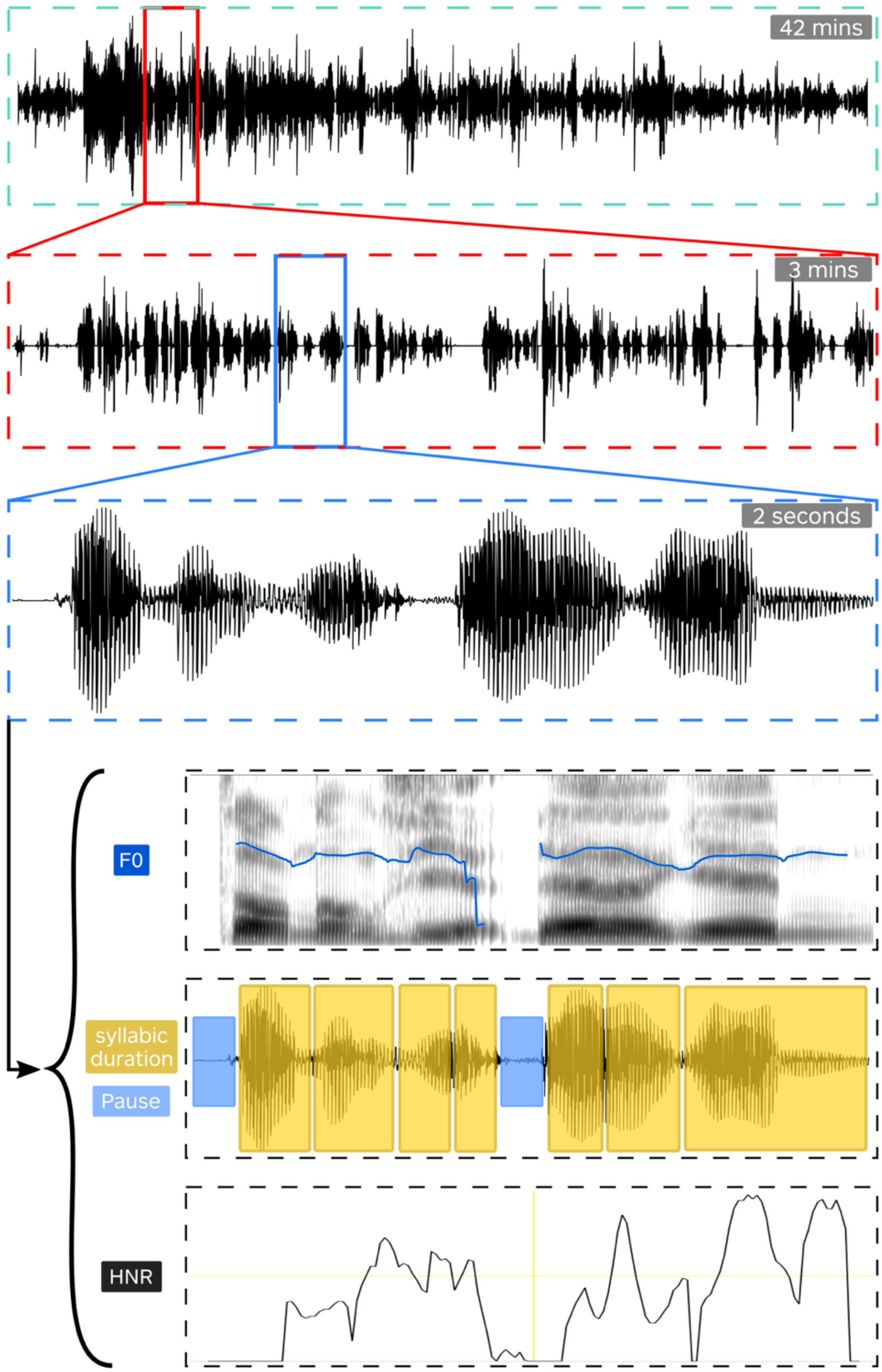


Figure 11 Procedures of data processing by Praat

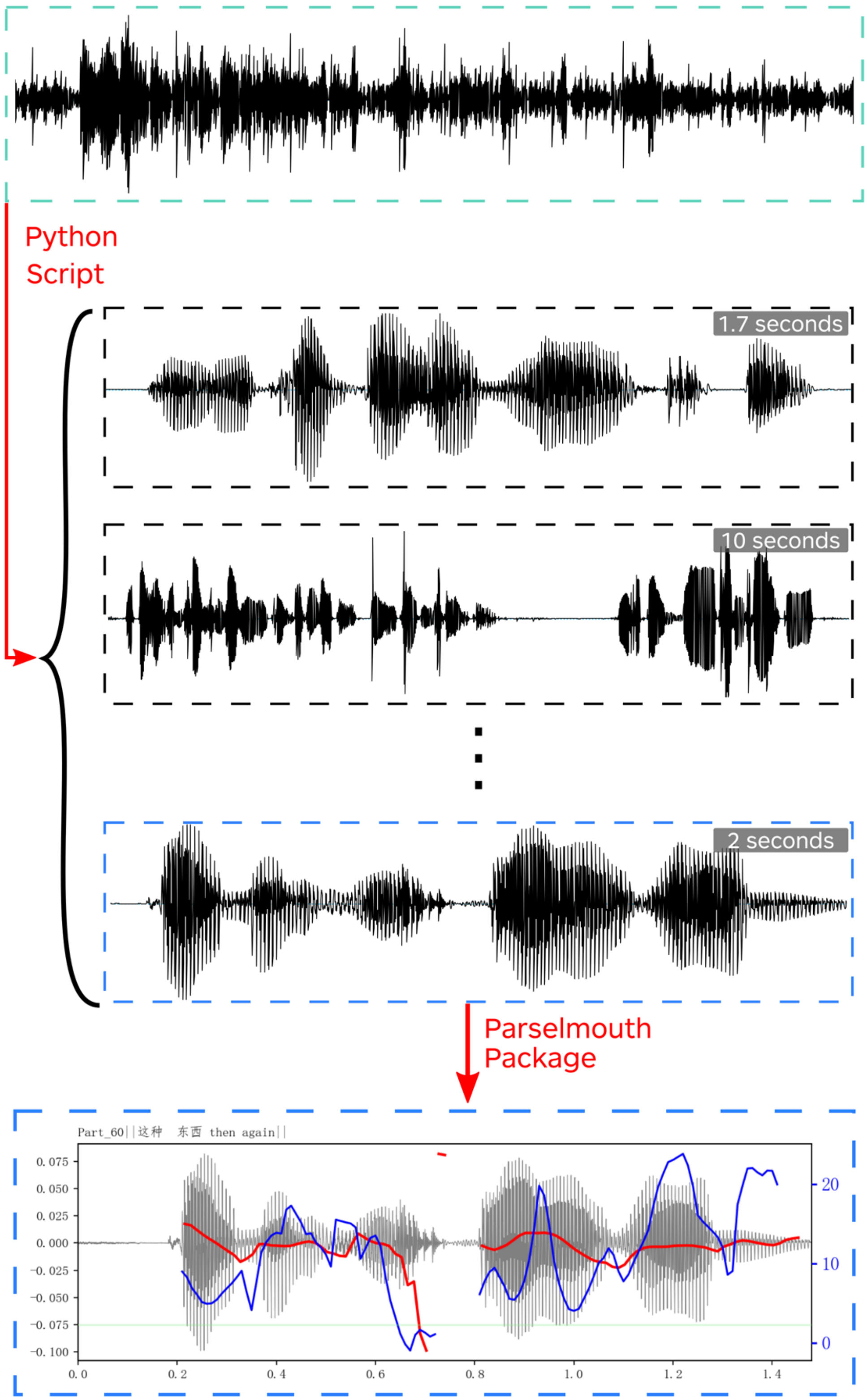


Figure 12 Procedures of data processing by Python

The phonetics software Praat and the Python script are applied for data processing in the current study. Figure 11 and Figure 12 illustrate how data is processed in Praat and Python. For Praat, first the data is manually zoomed into periods shorter than 10s, features such as F0, syllabic duration and pause can be read. For Python, the data is processed by a Python script to be separated according to a preassigned period (1-20s). After this, the Parselmouth package, a Python library developed by Jadoul et al. (2018), is used to plot the result figures, which present critical information such as the F0 and HNR curve. In the current study, Praat is applied to obtain information of syllabic duration and pause duration, and Python figures are applied to obtain information of F0 and HNR. It takes approximately 1h to process 1min of audio. With the consultation to several previous studies (Cacoullos & Travis, 2010; Manfredi et al., 2015; Mettouchi, 2008; Myslin & Levy, 2015; Shenk, 2006; Travis & Cacoullos, 2013; Urrea, 2012; Vargas, 2008), this study plans to examine 4 hours of audio, which is comparable to other theses (e.g., Urrea at 4 hours) and articles (e.g., Myslin & Levy at 3h; Cacoullos & Travis at 5.5 hours) doing the same research.

Now the critical issue is how to extract this 4h of data from the original 178h corpus. In most previous studies, the number of speakers was between 13 and 28. Then in order to avoid the possibility of encountering particular individual cases, within a limited total analyzing time, it would be more reasonable to study as many speakers as possible. Both the number of speakers and the appropriate length of period for each speaker need to be decided. The current study applies basic theory in statistics to help determine the representative length of each speaker. As long as the period is extracted in a reasonable way, it is able to accurately represent the language usage of the whole audio of a speaker. To achieve that, a convergence study is performed.

Convergence studies are usually applied to find a specific critical number of the object and apply it in the analysis to get more accurate results without any excessive calculations and time. In the current study, pointed out in previous sections, the representative length at which a period is extracted to represent the whole audio of a speaker must be decided. Six speakers, three from the interviews and three from the conversations, are randomly chosen to perform the convergence study. The whole audios of these 6 speakers (approximately 6 hours in total) are processed, it was found that extending the analysis of audio beyond 7min for the conversations and 5min for the interviews did not yield a greater BIU percentage. Selections of these lengths are thus taken as statistically representative of their larger tasks. While this finding is potentially dependent on the selection of speakers upon whose data the convergence study was performed, the speakers' results did not differ drastically from those of the group as a whole. Note that the representative length in interviews is shorter than that in conversations. This is consistent with the difference of the two speech types mentioned in the beginning of this section. In interviews, the majority of the speech in the audio belongs to the target speaker; yet in conversations, only approximately half of the speech belongs to the target speaker. Since in any sampling length of audio, the valid speech time is longer in interviews than in conversations, it is natural that the representative length in interviews is shorter than conversations.

Even though the representative length of interviews and conversations is already determined by the convergence study, one crucial factor should not be neglected since the corpus is made up of natural discourse: the impact of the session progress on language usage. The impact may be caused by the interlocutor, the topic change, or other factors. In order to respect this natural phenomenon, and to also further balance the result for the sake of cautiousness, the initial representative length (respectively 7min for conversations and 5min for interviews) was tripled. To be more specific, for speakers in conversations, the whole audio of each speaker is equally divided into three parts, named Stage 1, Stage 2 and Stage 3. At the beginning of each stage, 7min of audio is extracted. Finally, a length of 21min is obtained. Applying the same procedure, the representative length for speakers in interviews is 15min. In this way, we are able to observe more clearly the influence brought by the session progress in the results (a discussion is conducted in Section 5.4.1.2).

To sum up, the total analyzing time is around 4h. This ensures that the data processing is efficient, and more importantly, that it is aligned with the analyzing time of the majority of studies in the field. Then, the representative length for conversations and interviews is respectively 21min and 15min, in order to assure that the extracted periods are able to represent the whole audio. The next section is dedicated to the participant selection.

3.2.2 Selection of participants

An overview of the original speaker information is necessary to help make an appropriate selection among the 156 speakers. In addition to audio recordings and transcriptions, the

SEAME corpus includes sociolinguistic information on its speakers, such as age, sex and nationality. The profile of the 156 speakers of the corpus is summarized in Figure 13:

INTERVIEW				**CONVERSATION**		
category		number		category		number
gender	female	51		gender	female	31
	male	44			male	30
age	18-21	39		age	18-21	42
	22-25	39			22-25	19
	26-29	9			26-29	0
	30-33	5			30-33	0
	Not Given	3			Not Given	0
nationality	Malaysian	34		nationality	Malaysian	8
	Singaporean	61			Singaporean	53
total		95		total		61

Figure 13 Distribution of the participants of different categories

Given the result of the convergence study and the total analyzing time considered, the current study analyzes data from 16 participants. Half of them are from the interview group, the other half are from the conversation group. Then the participants are counterbalanced for sex and nationality within each group. Concerning the age, it is observed from Figure 13 that, the number of participants of each age group differs significantly. With this specific type of information kept in mind, the study analyzes data from 12 participants under the age of 25, and 4 above the age of 26. Further information on these 16 participants can be found in Appendix A. In total, this adds up to 4.8 hours of analyzed data. This is in alignment with the analyzing length of most studies in the field.

3.3 Transcription convention

Previous sections introduced that the SEAME team already provides the transcription data corresponding to the audio data. However, it is noted following common transcription convention. In other words, they simply lay out the transcriptions in text files, then indicate the starting and ending time points of each line in relation to the audio data.

When doing the IU analysis though, one needs to acknowledge and respect a great number of delicate phenomena that occur during the speech, such as a short pause, a sudden interruption, an inhalation at the end of a sentence, etc. Thus, the transcription method needs to reflect all these phenomena, and most importantly, it needs to reflect the existence of the IU. None of these could be reflected by the common transcription convention applied by the SEAME team. Consequently, a new, more suitable convention for the current study needs to be followed. Just like practically all previous studies in this field, both in monolingual IU studies and in bilingual IU studies, Du Bois et al.'s (1993) transcription convention is adopted for this research.

Du Bois et al. (1993) defined discourse transcription as the process of creating a written representation of a speech event to make it accessible to discourse research. It focuses on IUs, the fundamental unit applied in the current study. This transcription includes practically the most basic transcription information, for instance: a carriage return is used to indicate the end of an IU (the boundary between two IUs). Each IU appears on a separate line; the successive dots are used to indicate the short, medium and long pauses; laughter, inhalation, truncated words, etc. are all illustrated with different symbols. A rather detailed introduction of the symbols, an explanation of what the symbols stand for and how they are applied in the transcription can be found in Appendix B. It is worth mentioning that in the original version of transcription convention, there are many specific conventions and only part of them are identified in the database of the current study. Therefore, Appendix B includes only the conventions identified in the current study. Throughout this study, whenever a specific example is presented, the transcription method introduced in this section is applied.

Chapter 4 Methodology

4.1 Prosodic workflow

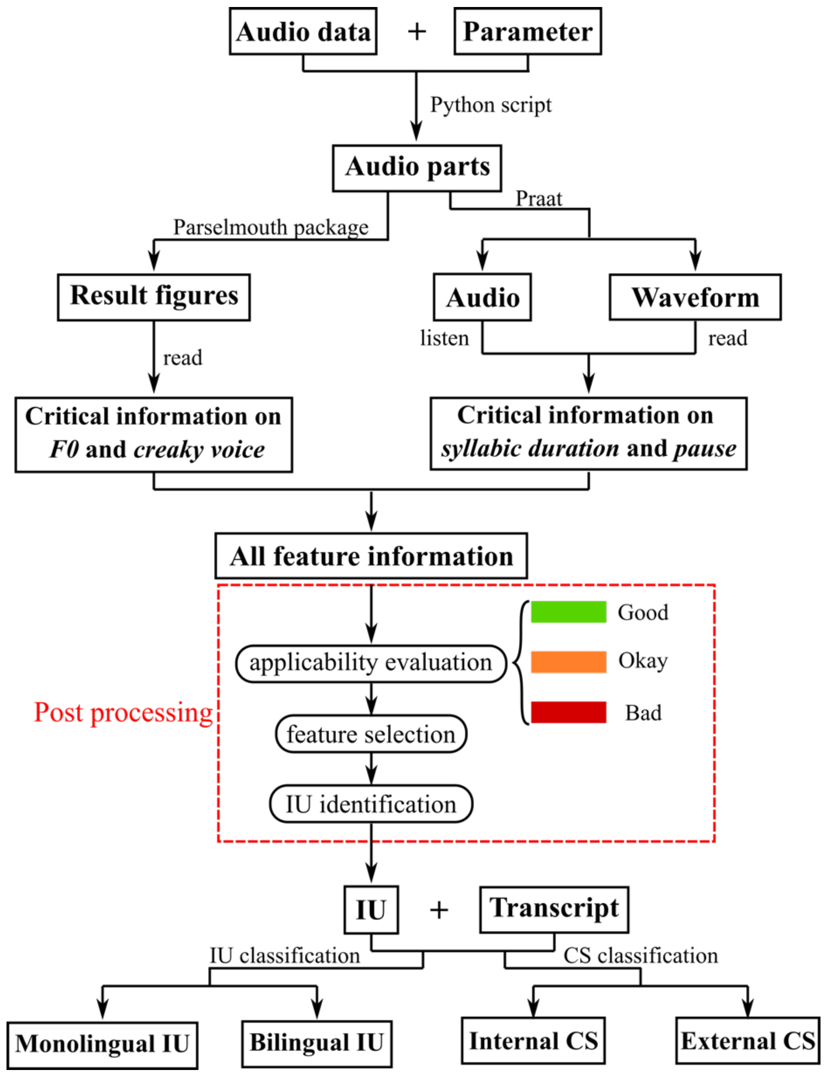


Figure 14 Workflow of the prosodic aspect

Figure 14 is the workflow of the prosodic aspect's systematic methodology developed in the current study. Chapter 3 introduced that both the audio data and the transcription data are applied in the processing, the former is applied in the prosodic analysis. First, the audio data and the preliminary parameter are processed by Python script in order to obtain different audio parts. Then, the audio parts are processed by the Parselmouth package and Praat. With the Parselmouth package, result figures are obtained[15], which provide critical information on two features, *change in F0* and *creaky voice*. With Praat, both the audio and the waveform are obtained. This allows the waveform to be read while listening to the corresponding audio, which provides critical information on the other two features, *syllabic duration* and *pause*. After this, all the feature information is post-processed using the following procedures: 1) applicability evaluation, with which the features are graded as *Good*, *Okay* or *Bad*; 2) feature selection, with which the two most applicable features[16] with the highest grading are selected; 3) IU identification, with which IUs are identified by applying the selected two features. All the above procedures are elaborated in Section 4.3. After the post processing, an analysis is conducted on IUs. With the help of the corresponding transcription, all IUs are classified to be either monolingual IUs or bilingual IUs. Then all the CSs are classified to be either internal CS (INCS, CS inside an IU) or external CS (EXCS, CS outside the IU)[17].

Note that in order to obtain readable and clear presentation of all the features for each speaker while applying Praat and the Parselmouth package, certain parameter settings need to be adjusted. Specific parameter adjusting according to each individual needs to be conducted preliminarily. First, Section 4.2 presents details on the preliminary parameter study. Second, Section 4.3 is dedicated to the introduction of the systematic methodology developed in this study, including the identification and the application of the four critical features, and also the discussion on how this methodology helps reduce the subjectivity problem (introduced in Section 2.3.2). After that, two practical examples are given in Section 4.4 to illustrate the actual analysis. Then Sections 4.5 – 4.7 present the syntactic methodology.

4.2 Parameter study of input information

Praat is applied in the current study for the analysis of the four features of an IU, namely *change in fundamental frequency*, *change in syllabic duration*, *pause* and *change in voice*

---

[15] Refer to Section 3.2.1 and Figure 12 for a clear introduction of the whole procedure, including the Python script processing and the Parselmouth package processing.

[16] This is elaborated in Section 4.3.

[17] The introduction of the concept of respectively EXCS and INCS can be found in Chapter 5, Discussion.

quality (or *creaky voice*), details are given in both Section 2.2.2 and Section 4.3. The post-processing and the analysis of the features can only be conducted with a clear and accurate presentation. Among these four features, the presentation of *change in syllabic duration* and *pause* are rather straightforward and clear, no preliminary parameter adjustment is needed. On the other hand, a parameter study is necessary for the other two features, the *change in fundamental frequency* and the *creaky voice*. Thus, in the current section, the focus is set on discussing the preliminary input parameter settings in Praat for these two features.

4.2.1 Preliminary parameter study on F0

The fundamental frequency (F0), realized in Hertz, is the frequency at which vocal cords vibrate in voiced sounds, our perception of the pitch of a speech sound usually depends on this. The initial setting instructions are obtained from the Praat official tutorial (Boersma & Weenink, 2018). It is indicated in *Intro 4.2. Configuring the pitch contour* from the official tutorial website that the standard pitch range for a human voice ranges from 75 to 500 hertz. For many low-pitched (e.g. average male) voices, one should set the floor to 75 Hz, and the ceiling to 300 Hz. Therefore, 75-300 Hz is set as the initial robust parameter for the male speakers in the current study. For many higher-pitched (e.g. average female) voices, one should set the floor to 100 Hz, and the ceiling to 500 Hz. This means that 100-500 Hz is the initial robust setting for female speakers. However, since the voice quality differs from individual to individual, slight changes are made accordingly to each of the 16 speakers of the current database. For each speaker in the current database, the pitch range is changed to adjust the final F0 contour presentation in order to obtain the most informative F0 contour. The specific F0 parameter setting for each individual is presented in Appendix A.

4.2.2 Preliminary parameter study on creaky voice

A creaky voice usually refers to a low, scratchy sound that occupies the vocal range below the modal voice (Laver, 1980). Usually, towards the end of an utterance, the vocal folds start to slow down and beat irregularly before closing, which then causes a rough voice quality, perceived as a creaky or raspy sound. Sometimes, a creaky voice can also be found at the beginning of an utterance (Keating et al., 2015).

In order to identify a creaky voice, two features can be applied, namely the *low F0 value* and the *irregular F0 value* (Keating et al., 2015). To be more specific, Boersma & Weenink (2018) defined the general F0 range for average males as 75Hz – 300 Hz. Thus, a value lower than 75Hz would be considered to be a *low F0 value*. The *low F0 value* feature can be observed directly from the F0 contour, examples are given in Section 2.2.2. In this section, I discuss the preliminary parameter that concerns the latter feature, the *irregular F0 value*, since it is difficult to define in simple words. The F0 contour can offer direct information on the *low F0 value*, yet it is slightly unpersuasive to define an *irregular F0 value* solely based on the F0 contour. Granted the F0 contour does change with some level of regulation. Human voices change constantly, and they are indeed different from individual to individual. There is no strict principled regulation of how the F0 contour changes over time. Thus, the line between a regular and an irregular F0 is rather difficult to draw. To deal with this problem, according to Keating et al. (2015) and Teixeira et al. (2013), there is a feature that can identify rather accurately the *irregular F0 value*, which is the Harmonic to Noise Ratio value (hereafter HNR value).

The HNR is an assessment of the ratio between periodic components and non-periodic components comprising a segment of voiced speech (De Krom, 1993; Murphy & Akande, 2005), expressed in dB. A high HNR value indicates a sonorant and harmonic voice sound, whereas a low HNR value denotes an asthenic voice and dysphonia, in other words, a creaky voice (Teixeira et al., 2013). Details on the application of HNR value can be found in Section 4.3.1. In the current section, the focus is set on the preliminary parameter set up to obtain an accurate HNR value in Praat. Boersma (1993), one of the authors and developers of Praat, developed an algorithm to obtain the HNR value. Applied in Praat, the algorithm then turns into four parameter settings, namely the *Time step*, which indicates the measurement interval in seconds; the *Minimum pitch*, which determines the length of the analysis window; the *Silence threshold*, which means that the frames that do not contain amplitudes above this threshold are considered silent; and the *Number of periods per window*. With these four parameters accurately set, one can find the HNR value at any specific time point required.

4.2.2.1 Time step

The *Time step* is the measurement of intervals or frame durations, realized in seconds. For the setting of *Time step*, the suggested standard value by the authors of Praat is 0.01s. Several parameters have been tested respectively: 0.5s, 0.1s, 0.05s, 0.01s. 0.005s and 0.001s, and the results of the HNR value contour are as followed:

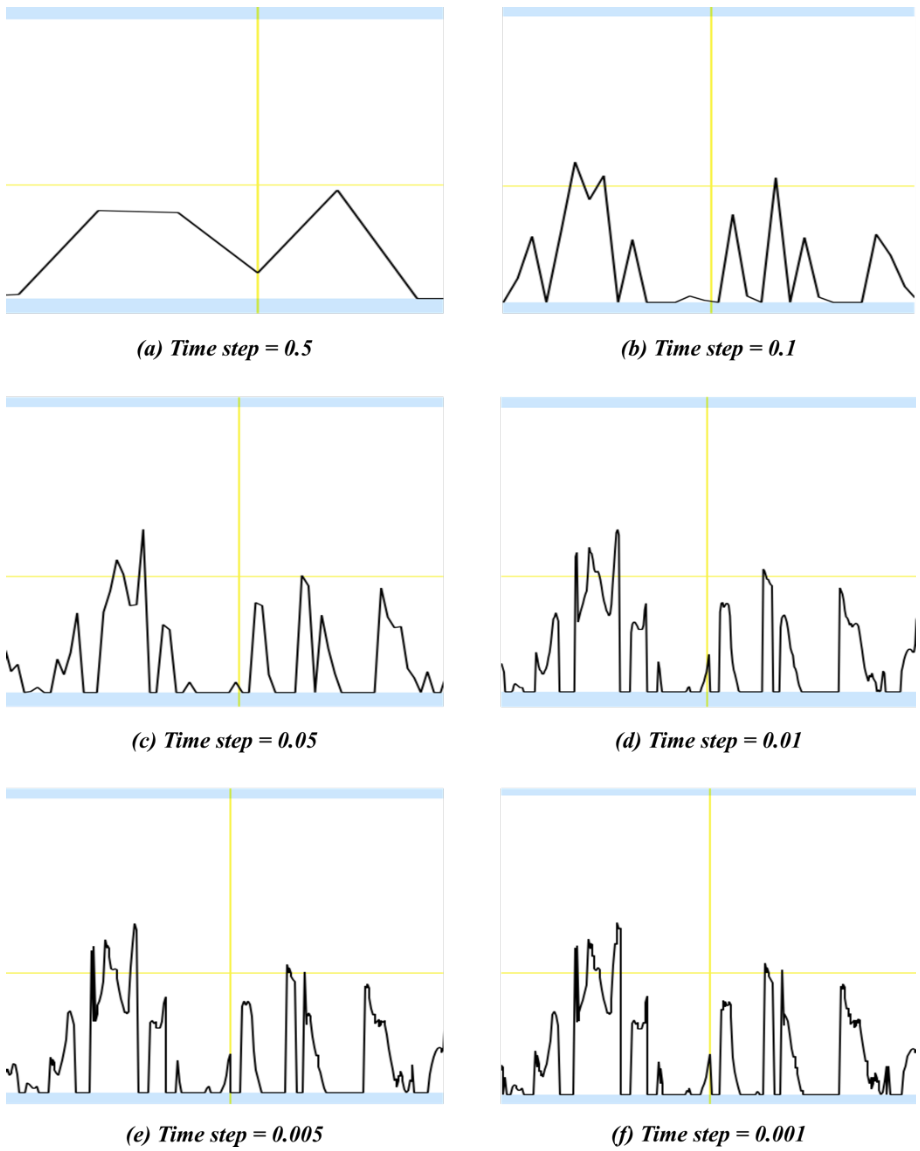


Figure 15 HNR value contour with the Time Step

As indicated in Figure 15, the contours of (e) and (f) are too specific, which present more details than needed, and thus they hinder the reading of the HNR value. On the other hand, the contours of (a), (b) and (c) are too rough, which means that they offer limited information, sometimes even false information due to the overly large time step. This can also affect the final reading of the HNR value.

Based on the comparison above, and the fact that this parameter is not individually specific, the current study respects the default setting of this parameter. The preliminary input parameter of *Time step* for the audio data for all the speakers of the current study is set on 0.01s.

4.2.2.2 Minimum pitch

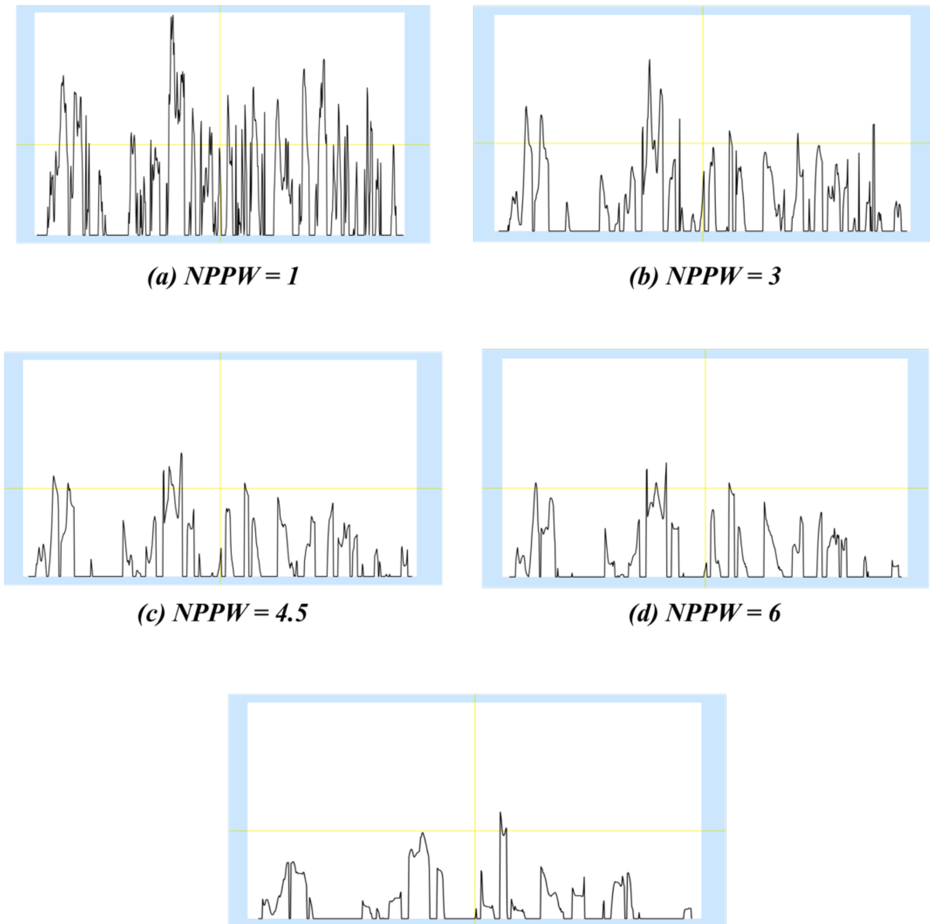
The *Minimum pitch* determines the length of the analysis window. For the setting of the *minimum pitch*, the suggested standard value by the developer of Praat is 75 Hz. However, this parameter is individually specific and the minimum pitch for the speakers are different (Section 4.2.1). Since the most suitable minimum and maximum pitch (the floor and ceiling value) have already been identified for all 16 speakers for the current study, the value of the *Minimum pitch* is set accordingly for every speaker. Information on this can be found in Appendix A.

4.2.2.3 Silence threshold

The *Silence threshold* is set due to the reason that frames that do not contain amplitudes above this threshold (relative to the global maximum amplitude) are considered silent. In a study focusing on the effect of silence in dimensional human emotion perception, Atmaja & Akagi (2020) investigated the impact brought by different silence threshold value. They set the value at respectively 0.1, 0.2, 0.3 and 0.4. The smaller the value, the tighter the filter. They found that for a regular human voice, should the value be set above 0.1, it may result in an incorrect decision to include speech as silence. Since 0.1 is also the standard value suggested by the creator of Praat, the current study decides to use 0.1 as the value of silence threshold.

4.2.2.4 Number of periods per window

For the *Number of periods per window*, this study also respect the standard value suggested by the author of Praat. This value stands at 4.5 since it is clearly pointed out in the Praat official tutorial that a value of 4.5 is best for human speech, for HNR values up to 37dB are guaranteed to be detected reliably. To be thorough, a comparison between the setting of 1, 3, 4.5, 6, 9 is made, as presented in Figure 16:



(a) NPPW = 1

(b) NPPW = 3

(c) NPPW = 4.5

(d) NPPW = 6

(e) NPPW = 9

Figure 16 HNR value contour with the Number of periods per window (NPPW)

It is observed in Figure 16 that (c) presents the most suitable contour, neither too specific with too many little bumps in the contour which disturbs the value reading, nor too flat with too little information to offer.

4.2.3 Summary

In this section, the preliminary parameter set up for the reading of the F0 contour and the HNR value is discussed. For the F0 settings, specific information on different individuals could be referred to in Appendix A. As to the HNR value settings, the four parameters will be set respectively as followed:

(1) Time step: 0.01s

(2) Minimum pitch: as indicated in Appendix A

(3) Silence threshold: 0.1
(4) Number of periods per window: 4.5

With several other robust preliminary parameters set up in Praat, the presentation of the features obtained from the audio data is readable and informative.

4.3 Prosodic methodology

Section 2.3.2 presents that there is a certain subjectivity problem in previous studies' methodology, which affects to some level the identification and the application of IU features. Based on Chafe's (1994) approach (presented in Section 2.2.2), the current study further investigates the identification and the application of the features, the Mandarin-English language pair and the current database. A systematic methodology is developed to reduce the subjectivity problem as much as possible, illustrated in the following subsections.

4.3.1 Identification & application of IU features

4.3.1.1 Identification of features

Section 2.2.2 introduces that Chafe (1994) proposed four critical features to segment an IU, the fundamental unit of the information-based approach. These features were primarily proposed for English, although Chafe pointed out that they are universally applicable, tonal languages included[18] (such as Mandarin). For the current study, several issues need to be clarified to make the identification of certain features more categorical.

(1) *Change in F0*

The first problem concerns the feature *change in F0*. The language pair of the current study is peculiar because one of the languages, Mandarin, is a tonal language. Section 2.1.3.1 introduces that Mandarin possesses lexical tones. Both the lexical tone and the phrasal intonation are primarily realized in pitch (Chang, 1958; Shen, 1990). Thus, in order to make

[18] For the two target languages of the current study, the applicability of the four features has been verified by research on the English language (Cameron-Faulkner et al., 2003; Du Bois et al., 1993; Gee, 2005; Shenk, 2006) as well as on Mandarin (Tao, 1996).

more accurate identification of the feature *change in F0*, it is necessary to clarify the relationship between the lexical tone and the phrasal intonation in Mandarin, especially whether the former affects the latter or not.

Firstly, a certain consensus should be discerned. There is great interplay between lexical tones and phrasal intonation, with the global intonation considerably influencing the lexical tone (Tao, 1996). Like in other languages, the phrasal intonation in Mandarin expresses various aspects of modality, emotion and attitude. With respect to the relationship between lexical tone and phrasal intonation, among many characterizations, the one given by Chao (1968) remains well-known and valid to present day:

"The question has often been raised as to how Chinese can have sentence intonation if words have definite tones. The best answer is to compare syllabic tone and sentence intonation with small ripples riding on large waves." (Chao, 1968, Page 39)

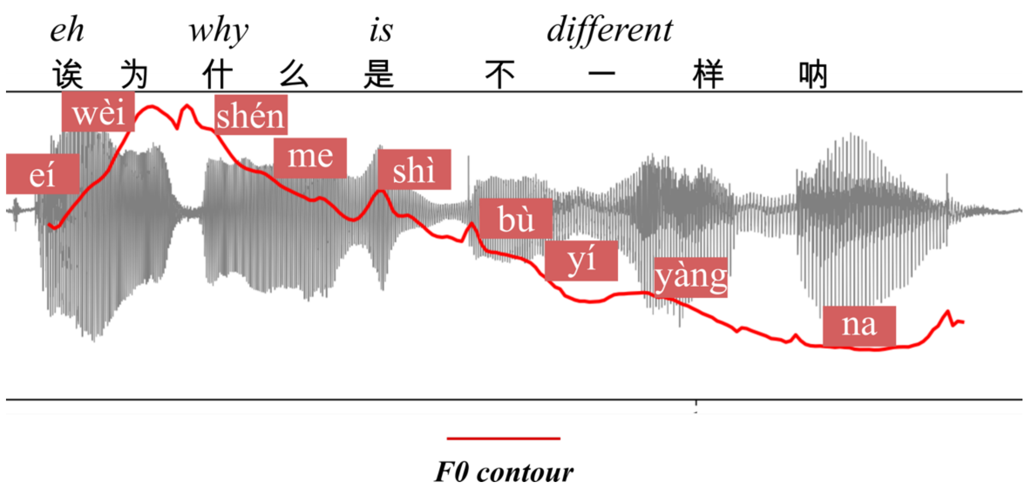


Figure 17 Example of a Mandarin IU retrieved from speaker ***04NC07FBX***

(4-1)	**Characters:**	诶	为	什	么	是	不	一	样	呐？
	Pinyin:	eí	wèi	shén	me	shì	bù	yí	yàng	na
	IPA :	**/eɪ2/**	**/weɪ4/**	**/ʂən2/**	**/mɤ5/**	**/ʂɚ4/**	**/pu4/**	**/i2/**	**/jɑŋ4/**	**/na5/**
		eh		*why*		*is*		*different*		*Particle*

'Eh, why is it different?'

Figure 17 illustrates that the first syllable "eí" is a high rising *tone 2*, followed by "wèi" a syllable with a high falling *tone 4*. Since the *tone 4* has relatively the highest pitch among the five lexical tones. Therefore, the second syllable occurs in a slightly higher pitch than the first one, and possesses the highest F0 value in the contour. It is easy to observe that the three syllables "shì" "bù" "yàng" following the second syllable possess the high falling *tone 4*, though not in a continuous series. They all create small ripples in the contour. It is observed from the contour that, even with the local small ripples, the global pitch contour is still easy to identify to be in a declination tendency. This illustrates how the F0 curve has the global tendency to decline with time, despite local rises and falls, and further indicates that the lexical tone in Mandarin does not affect the identification of the *change in F0* feature.

(2) *Change in syllabic duration*

The second problem concerns the *change in syllabic duration*. Section 2.2.2 presents that Chafe (1994) generally defined the duration for short, normal and long syllables. He also pointed out that the figures need to be adjusted for slower and faster speaking rates, indicating that the syllabic duration is individually specific and correlates to one's speaking rate. People speak at different rates, the referential syllabic duration does not apply to everyone. Thus, the current work conducts a duration study on the database.

For the sake of simplicity, the term syllabic duration is represented by D. Then, the critical value of a *short* and *long* syllabic duration is represented by D_s and D_l respectively. Suppose there is a syllable with the duration of D_x, when $D_x < D_s$, this syllable is defined as a short syllable; when $D_s < D_x < D_l$, it is defined as a normal syllable; when $D_x > D_l$, it is defined as a long syllable. In order to determine the syllabic duration, a specific procedure is proposed for each speaker from the database:

1) Randomly choose i periods of proper length ($10s$) audio;
2) For the first period, find out the range of D_s^1 and D_l^1 respectively;
3) Repeat step 2) for the period i ($i = 2,3,4 \dots$) chosen in step one, find out the range of D_s^i and D_l^i;

4) Calculate the integrated range of duration for the speaker based on D_s^i and D_l^i $(i = 1,2,3 \ldots)$; for example, the range of a ***speaker A*** is $D_s \in [0.06s,\ 0.09s]$ and $D_l \in [0.30s,\ 0.40s]$;
5) Calculate the critical value[19] of the duration of each speaker; for example, the critical value of ***speaker A*** is $D_s = 0.075s$ and $D_l = 0.35s$.

When processing the audio data of ***speaker A***, for any syllable D_x, when $D_x < D_s = 0.075s$, it is defined as a short syllable; when $D_x > D_l = 0.35s$, it is defined as a long syllable; for a length in between, it is defined as a normal syllable.

The syllabic duration reflects the speaking rate of an individual. The longer the syllabic duration, the slower the individual speaks. Since every individual speaks at a different rate, it would be impossible to set an absolute number on D_s or D_l for everybody. Although it is possible to find a range for a database on how long a D_s and a D_l should be. This duration range offers a global and direct overview of the length of different syllables within this database. So after applying the procedure to all 16 speakers of the current database, the duration range is presented in Figure 18.

The red and blue rectangles represent the range of short and long duration. For the current database, the ranges of of D_s and D_l are respectively between 0.06s and 0.15s, and between 0.30s and 0.79s. Specific examples of how the syllabic duration reflects the speaking rate of individuals are given as followed. Take the syllabic durations of ***speakers A, B*** and ***C*** as an example, illustrated with the gray, green and orange lines on the rectangles. ***Speaker A*** is represented by the gray line, $D_s^A = 0.075s$, $D_l^A = 0.35s$. ***Speaker B*** is represented by the green line, $D_s^B = 0.11s$, $D_l^B = 0.55s$. ***Speaker C*** is represented by the orange line, $D_s^C = 0.15s$, $D_l^C = 0.62s$. This illustrates that D_s and D_l are indeed individually specific. It can be observed from the syllabic durations that ***speaker A*** is relatively a fast speaker, whereas ***speaker C*** has a relatively slow speaking rate.

[19] Critical values are essentially cut-off values that define regions where the test statistic is unlikely to lie. Here the critical value of the duration of different types of syllables is set to be the mean value of the maximum and the minimum values.

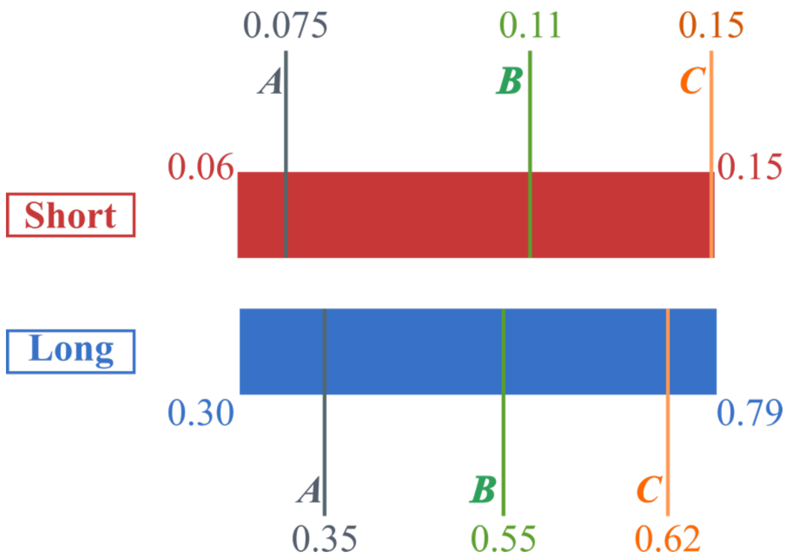


Figure 18 Overview of the duration range for the database of the current study

Note that this syllabic duration range parameter suits solely this study because it is established based on an investigation of the speakers from the current database. For any other database, a new investigation would be required.

(3) Pause

The information-based approach is proposed by Chafe, then applied by many other linguists in this field. Section 2.2.2 introduces that Chafe (1994) pointed out the importance of taking speaking rates into consideration when applying the *change in syllabic duration* feature. However, he did not mention whether or not speaking rates would affect the pause length of different individuals. To the best knowledge of the current study, other studies did not discuss this matter either. Two possibilities exist: 1) speaking rates have no impact on the pause length, Chafe confirmed this and defined directly the length of different types of pauses, in order to provide better reference for later researchers. 2) speaking rates have certain influence on the pause length, but no researchers have considered this. No matter which situation, note that together with another feature, any types of pauses can always define an IU boundary. Defining different types of pauses by length makes the transcription clearer (see Appendix B). Since this matter does not influence the analysis and the conclusion of the current study, I do not focus on this detail. Future studies are encouraged to conduct an investigation, and confirm whether speaking rates impact the pause length.

(4) *Change in voice quality (creaky voice)*

Section 2.2.2 presents that Chafe (1994) defined a *creaky voice* to be a laryngealization or a “fry”. Other previous studies stated that a creaky voice possesses “obvious difference” from a regular voice. The obvious problem is pointed out in Section 2.3.2, without a specific criterion, the identification of this feature relies on the judgement of the researcher. Thus, the current study conducts a further investigation to provide a criterion for the identification of a *creaky voice*.

A creaky voice usually has lower F0 than a modal voice (Garellek, 2012; Gerfen & Baker, 2005; Keating et al., 2015). Figure 19 illustrates a creaky voice indicated by low F0 value, the waveform of the modal voice and the creaky voice stand in clear contrast. Although in the practical analysis, pointed out by Keating et al. (2015), a low F0 can be difficult to estimate, for two reasons:

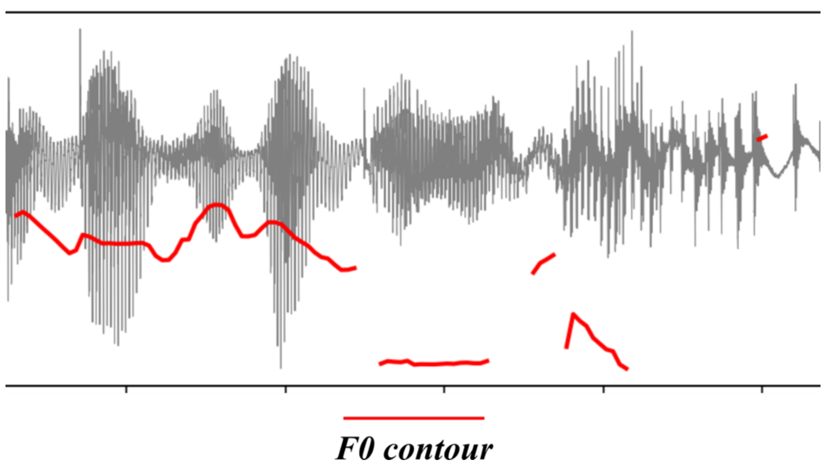


Figure 19 Example of creaky voice indicated by low F0 retrieved from speaker ***03NC05FAX***

1) A *low F0* is identified subjectively, based on the comparison with the F0 value of the modal voice.
2) Usually, when creaky voice occurs, the irregular F0 value follows, sometimes there is even no F0 to be found. Thus, a solid *low F0* may not always be present when needed.

Based on these two reasons, the feature *Low F0 value* serves only as the secondary feature in the identification of a *creaky voice*, which researchers turn to only when running out of better choices. The primary feature is the *irregular F0 value*. According to Keating et al. (2015), a creaky voice usually has less regular voicing than a modal voice. As a deviation from the regular F0, a creaky voice with such voicing irregularity is usually perceived as noise, which is then measured by the HNR. The definition and the preliminary parameter setting up of the HNR are discussed in Section 4.2.2. The irregular F0 determined by the low HNR value is a correlate of a creaky voice (Garellek, 2012, 2014, 2015; Garellek & Keating, 2011; Miller, 2007; Pan et al., 2011).

For a healthy voice, when producing a sustained vowel [a] or [i], the HNR value is around 20dB. For a hoarse voice, the HNR value is much lower than 20dB when producing the vowel [a] (Boersma, 1993). For natural discourse, based on the database of the current study, the mean HNR value for a modal voice is between 15dB and 25dB. Teixeira et al. (2013) and Boersma (1993) suggested that a value of less than 7dB in the HNR should be considered pathological. However, in the current study, a slight change has been made on the regulation. Since every individual possesses different voice quality, just as the F0 and speaking rate are individually specific. The mean HNR value also differs for each speaker. In that case, setting a specific value as 7dB as the low-value indicator for everyone would be inappropriate. Considering the fact that the suggested 7dB is roughly 20% of the average mean value. For the current study, when analyzing each speaker, the low HNR indicator is set as 20%.

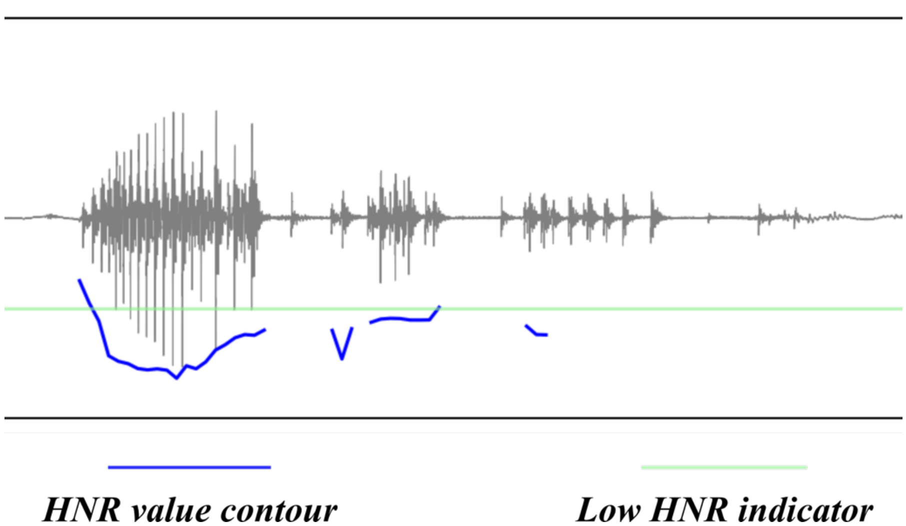


Figure 20 Example of creaky voice indicated by low HNR value retrieved from speaker ***05NC10MAY***

Figure 20 is an example of a creaky voice, the waveform represents the English word "ah". With the blue line representing the HNR value contour, and the green line indicating the

20% of the mean HNR value, one can observe from Figure 20 that the HNR value for this period is quite low, which further indicates the existence of a creaky voice.

4.3.1.2 Application of features

The workflow in Section 4.1 illustrates the application of the features. For each speaker, the complete audio data and the preliminary parameter are processed by Python script in order to obtain audio parts at suitable analyzing length, introduced in Section 3.2.1. Then, the audio parts are processed by both the Parselmouth package and Praat. The Parselmouth package plots result figures, which present critical information for the identification of features *change in F0* and *creaky voice*. Praat allows researchers to read the waveform and listen to the corresponding audio, with which researchers obtain critical information for the identification of *change in syllabic duration* and *pause*. After this, all the feature information is post-processed with the following procedures: 1) ***Applicability evaluation***, with which the features are graded as *Good*, *Okay* or *Bad*. Features with a clear presentation are graded as *Good*, and those that have a very vague presentation or do not exist at all are graded as *Bad*. Those that stand somewhere in between are graded as *Okay*. 2) ***Feature selecti*on**, based on the grading in step 2, features with the highest grading are selected to identify the IU. Note that it is not necessary for all the features to be clearly present in each IU. To be more specific, on an arbitrary basis, certain features can turn out to have a less obvious presentation for some IUs. This is not a problem because, based on the observation during the data processing of the current study, two *Good* features are enough to identify an IU. Most of the time, three or four *Good* features are observed. 3) ***IU identification***, IU is identified by applying the selected two features. After the post processing, we conduct the analysis on IU. With the help of the corresponding transcription, all the IUs are classified as either monolingual IU or bilingual IU. Then, all the CSs are classified as either internal CS (INCS, CS inside an IU) or external CS (EXCS, CS outside the IU). The two examples in Section 4.4 explain this procedure in a practical way.

Based on consultation to previous work presented in Section 2.2.2 and further investigation conducted in this study (presented in this section), a systematic methodology is developed. The next section discusses how this methodology reduces the subjectivity problem as much as possible.

4.3.2 Discussion on subjectivity problem

Section 2.3.2 presents that both the identification and the application of the features come with subjectivity problem. In order to avoid the impression of the subjectivity, some issues need to be clarified.

4.3.2.1 Identification of features

Concerning the identification, features *change in F0* and *pause* suffer little impact from the subjectivity problem, even for tonal language Mandarin (clarified in Section 4.3.1). On the other hand, features *change in syllabic duration* and *creaky voice* are influenced more by this problem.

Section 2.3.2 discusses the problem that during the identification of *change in syllabic duration*, subjectivity problem mainly affects the duration categorization and the delimitation of syllables. Considering the duration categorization in previous studies, solely Chafe (1994) proposed a referential duration for each type of syllables, pointing out that it should be adjusted according to the speaking rate of different speakers. Based on the database, the current study develops a systematic syllabic duration study procedure, with which the syllabic duration is adjusted to distinguished speakers (clarified in Section 4.3.1). However, one can still encounter problems during the practical identification process. Take ***speaker A*** ($D_s \in [0.06s,\ 0.09s]$, $D_l \in [0.30s,\ 0.40s]$; $D_s = 0.075$, $D_l = 0.35$) from Section 4.3.1 as an example, Figure 21 illustrates the distribution of the duration.

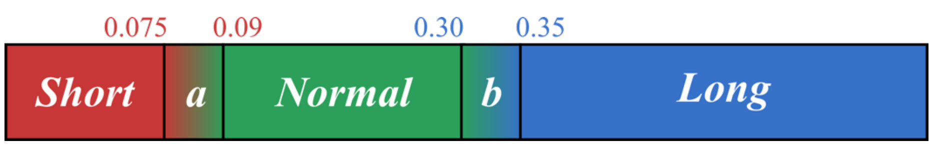


Figure 21 Example of duration distribution of ***speaker A***

As is concluded in the previous section and what could be observed in Figure 21, for ***speaker A***, one can easily define a syllable (D_x) as a short syllable $D_x < D_s\ (0.075s)$, $0.09s < D_x < 0.30s$ as a normal syllable, or $D_x > D_l\ (0.35s)$ as a long syllable. However, certain situations may cause a problem in the decision making. Figure 21 presents that in *Zone a* [0.075s, 0.09s] and *Zone b* [0.30s, 0.35s], it is difficult to clearly decide the type of syllable, it could be a short or a normal one in *Zone a*, and it could be a normal or a long one in *Zone b*. These *Zones* are named *Fuzzy Zones*. When a syllable falls into one of the *Fuzzy Zones*, it might (but not necessarily) affect the identification of an IU, and further affects the final results. Several situations are discussed:

First of all, for a phrase with several syllables (e.g., more than five), such as the phrase in Figure 10, one or two *Fuzzy Zones* syllables wouldn't affect the grading. It might still be graded as *Good*. Then, if most of the syllables in one phrase fall into the Fuzzy Zone, the influence is significant enough. This feature should be graded as *Bad*. In that case, one needs to rely on other *Good* features to identify an IU. Finally, if this feature is graded as *Bad* and all the other features are also graded as *Bad*, it will truly affect the identification of an IU. This kind of situation is rarely encountered. In this case, it is suggested to categorize this IU as unanalyzable and eliminate it from the analysis.

As for the delimitation of syllables, Section 2.3.2 points out that under certain circumstances, the boundary between two syllables can be difficult to locate. The exact same example is illustrated in Figure 22.

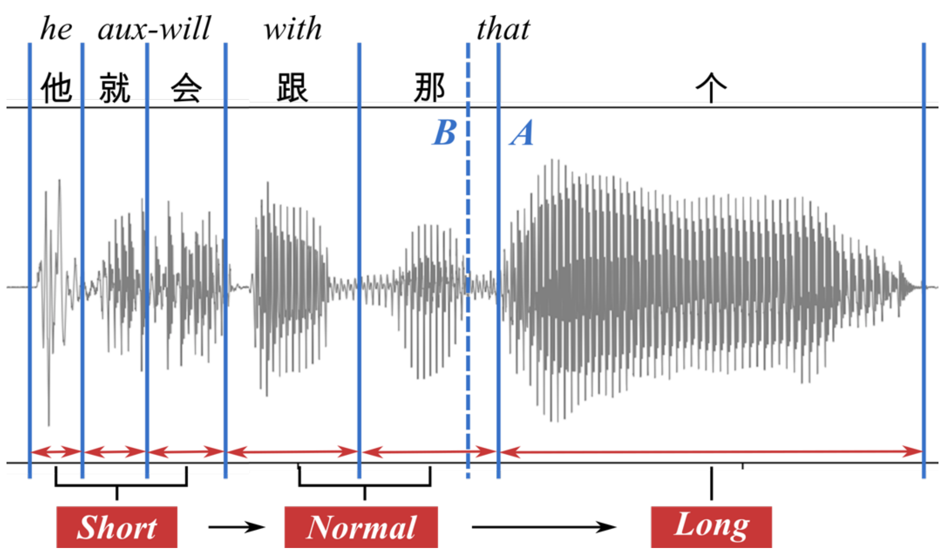


Figure 22 Example of syllabic duration change in an IU retrieved from speaker ***NI20MBP***

In terms of boundaries between syllables *那 (nà, **/na4/**)* and *个 (gè, **/kɤ4/**)*, different researchers could make slightly different decisions solely from listening to the audio and reading the waveform. Any boundary point between A and B is possible and this does not affect the categorization of a syllable. Note that the duration between A and B is shorter than 0.02s. Section 4.3.1 illustrates that for the current database, the ranges of D_s and D_l are respectively between 0.06s and 0.15s, and between 0.30s and 0.79s. Compared to the length of an actual syllable, the influence brought by this uncertainty is neglectable. In other words, even with this difference of 0.02s, the categorization of a short, a normal or a long syllable hardly differs. As a result, the IU identification is not affected.

As to the feature *creaky voice*, previous studies did not provide a solid criterion to help identify a creaky voice. The identification relies on the researcher's judgement. Section 4.3.1 clarifies that the current study provides a criterion for the identification by introducing the concept of HNR. This helps identify a creaky voice in a quantitative way, which reduces the impact of the subjectivity problem.

4.3.2.2 Application of features

Section 2.3.2 discusses the fact that the application of features suffers possible impact of the subjectivity problem too. Previous studies in this field only provided rather vague instructions on this matter. In the current study, a three-step procedure is proposed, this is thoroughly introduced in Section 4.1 and Section 4.3.1. This procedure solves a great deal of the remaining few problems. First, based on the observation of the data processing of the current study, two features can efficiently denote an IU boundary. If one or two of the features read really poorly, researchers can simply grade it as *Bad*, then seek the other *Good* features to help make the identification. If a rare situation occurs, that is three or four features read poorly and are graded as *Bad*, then this would truly affect the judgment of an IU, researchers would run out of solutions. They could make a judgement with the limited information at hand, or eliminate this unit from the analysis since it is judged to be unanalyzable. After processing over 4 hours of data, this study determined that this is such low probability event (much less than 1%) that it won't affect the conclusion of the whole research.

4.4 Identification of an IU

In this section, two practical examples retrieved from speaker ***NI20MBP*** and speaker ***UI12FAZ*** are given in order to offer a more detailed description of the workflow presented in

Section 4.1. These examples also present further details on the methodology introduced in Section 4.3 and explain more practically how the identification work is performed.

4.4.1 Example of speaker *NI20MBP*

Speaker ***NI20MBP***, male, age 22, Singaporean, speech type *Interview*. Preliminary input parameter for the pitch contour and the HNR value are listed below. Figure 23 is an example of an audio part plotted by the Parselmouth package:

For pitch contour:

1) Maximum pitch: 280
2) Minimum pitch: 60

For HNR value:

1) Time step: 0.01
2) Minimum pitch: 60
3) Silence threshold: 0.1
4) Number of periods per window: 4.5

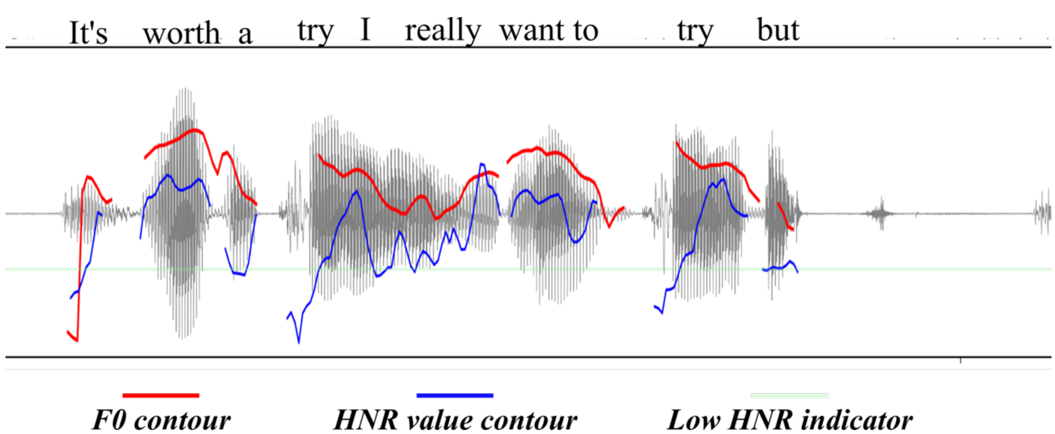


Figure 23 Example of an analyzable period retrieved from speaker ***NI20MBP***

The first step in the post-processing is to analyze all the features, and then evaluate their applicability:

1) ***Change in F0***

Represented with the red line, three clear global F0 contours are observed in Figure 23. Two pitch resets are observed, the first one between the words “a” and “try”, and the second one between the words “to” and “try”. All three contours present a clear declination tendency. Four boundaries are identified with the *change in F0* feature, and it is graded as *Good*.

2) Change in syllabic duration

According to the procedure introduced in Section 4.3.1, the syllabic duration range for this speaker is determined as $D_s \in [0.11s,\ 0.14s]$ and $D_l \in [0.44s,\ 0.60s]$. Then the critical value for this speaker is $D_s = 0.125s$ and $D_l = 0.52s$. Thus for any syllable D_x, when $D_x < D_s = 0.125s$, it is defined as a short syllable; when $D_x > D_l = 0.52s$, it is defined as a long syllable. By processing the waveform and the audio in Praat, in this period, no syllables can be directly defined as D_s or D_l, they all fall into the range of a normal duration or into the *Fuzzy Zone*, which is then hard to define. Based on the above analysis, no boundaries are identified with the *change in syllabic duration* feature, and it is graded as *Bad*.

3) Pause

Four visible pauses are observed in the waveform in Praat. The first one is at the beginning before the word “it’s”. It lasts 0.3s and is defined as a medium pause. The second one is between the words “a” and “try”, lasts 0.1s and is defined as a short pause. The third one is between the words “to” and “try”, lasts 0.1s and is defined as a short pause. The fourth one is between the word “but” and the beginning of the next period. It lasts 0.7s and is defined as a long pause. All four pauses have a clear presentation and are easy to observe. Four boundaries are identified with the *pause* feature, and it is graded as *Good*.

4) Change in voice quality (creaky voice)

4-1) As the red line indicates, a low F0 value is observed at the beginning of the first contour, a creaky voice is suggested based on the word “it’s”. Although the *low F0 value* is considered the secondary feature in the identification of a creaky voice, more evidence is needed.

4-2) As the blue line indicates, the low value indicated by the green line, five low HNR values are observed. The first one is identified at the word “it’s”, which supports the one identified by the low F0 value; the second one is identified at the word “a”; the third one is at the first “try”; the fourth one is at the second “try”; and the fifth one is at the word “but”.

Since creaky voice often appears at the boundary (often at the end and sometimes at the beginning) of an IU, four boundaries are identified with the *creaky voice* feature, and it is graded as *Good*.

The second step in the post-processing is to select the two most applicable features (those that are graded as *Good*), and then **step three** is to make the identification of the IUs. For this audio part, based on the analysis of **steps one and two**, choose any two of the features among *change in F0*, *pause* and *creaky voice*, or with the help of all of them, four boundaries are confirmed, thus three IU are identified. The IU is presented with respect to the transcription convention (Du Bois et al., 1993) introduced in Chapter 3 (speaker ID ***NI20MBP*** represented by N in the transcription):

(4-2) N: ... it's worth a
.. try I really want to
.. try but (SNORT)

4.4.2 Example of speaker ***UI12FAZ***

With the same procedure applied to speaker ***NI20MBP***, a brief description is given for the audio part of speaker ***UI12FAZ***. Figure 24 is an example of an audio part:

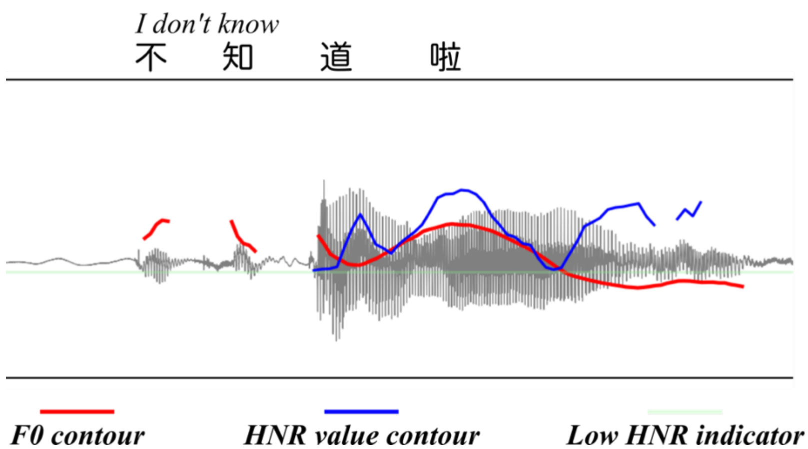


Figure 24 Example of an analyzable period retrieved from speaker ***UI12FAZ***

The feature *change in F0* is graded as *Okay*, for a relatively clear pitch contour with declination tendency is observed. The feature *change in syllabic duration* is graded as *Good*, for a clear accelerate – decelerate change is observed: first there are two short syllables, followed by a normal syllable, then finally a long syllable. The feature *pause* is graded as *Good*, for two clear pauses are observed, one long pause before the first syllable and one short pause after the last syllable. The feature *creaky voice* is graded as *Bad*, for no creaky voice is observed in this period. Choose at least two *Good* features, and an IU is identified, transcribed as below:

(4-3) U: …(0.9s) 不知道啦 =.

Translation:

Characters:	不	知	道	啦
Pinyin:	bù	zhī	dào	la
IPA:	**/pu4/**	**/ʈʂɚ1/**	**/tɑʊ4/**	**/la5/**

'I don't know.'

4.5 Syntactic workflow

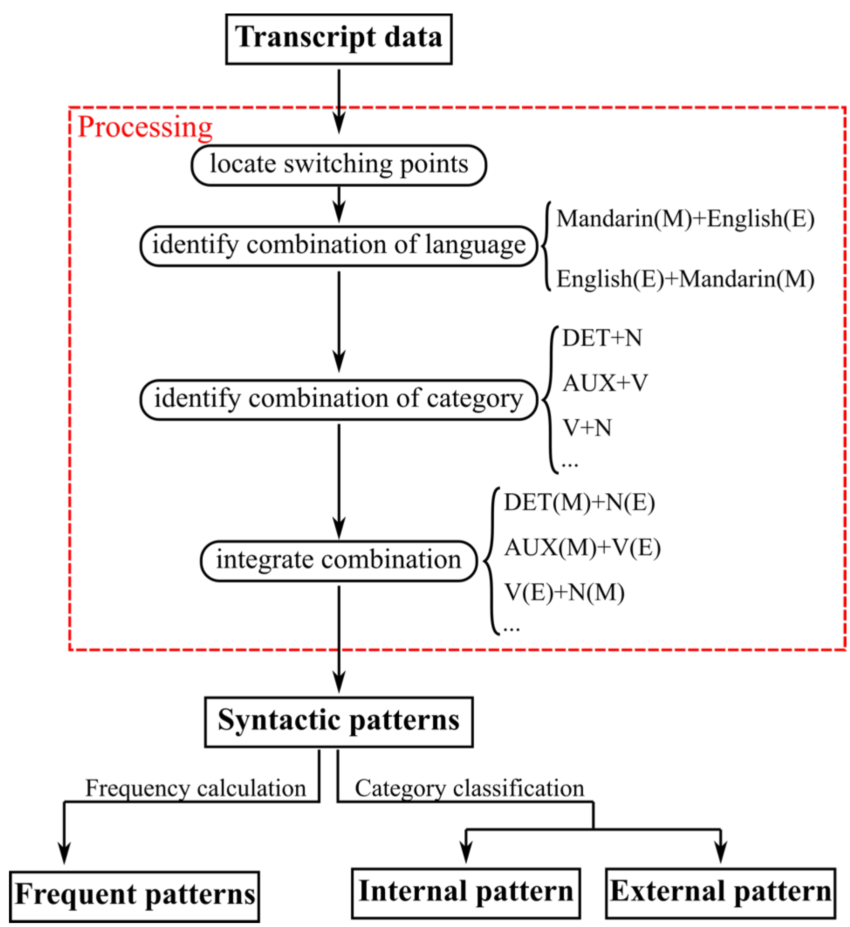


Figure 25 Workflow of finding syntactic patterns

For all speakers in the database, once the analysis of the audio data is done, the corresponding transcription data is analyzed, in order to examine the frequent syntactic patterns at switching points. Four steps are included in the processing: 1) locate all the CS points; 2) identify the combination of language, that is, the languages before and after the switching point, either *M+E* or *E+M*; 3) identify the combination of the syntactic category; 4) integrate the combinations of language and syntactic categories. The patterns are first classified into either internal patterns (internal to IU) or external patterns (external to IU), then calculated to find the frequent patterns.

Section 4.6 introduces the notation of the analysis of syntactic patterns. Three practical examples are given in Section 4.7, to offer a better and more detailed understanding of the workflow. Results are discussed in Chapter 5.

4.6 Notation

With consultation to previous work on syntactic patterns of Mandarin-English CS (Lu, 1991; Ong & Zhang, 2010; Tan, 1988), the coding procedure that denotes switch sites and distinguishes different types of CS patterns are shown in Table 3.

The syntactic categories preceding and following the switching points are specified with the coding notation indicated in Table 3. For example, at a switching point, if the preceding category is a verb, and the following category is a noun phrase, then the pattern at this switching point is noted as *V+NP*. Besides, as is indicated in the workflow, the specific language of each category is also noted, placed in the parentheses right after the category. The languages are noted as either *(M)*, indicating Mandarin, or *(E)*, indicating English. For example, the notation *V(E)+NP(M)* refers to a switch from a verb in English to a noun phrase in Mandarin.

Syntactic category	Coding notation
1. Noun	N
2. Noun phrase	NP
3. Adjective	ADJ
4. Adjective phrase	AP
5. Verb	V
6. Verb phrase	VP
7. Adverb	ADV
8. Auxiliary	AUX
9. Preposition	PREP
10. Determiner	DET
11. Determiner phrase	DP
12. Particle	Particle

Table 3 Notation used for the coding procedures

4.7 Identification of syntactic patterns

Three practical examples retrieved respectively from speakers ***NI18MBP***, ***NI20MBP,*** and ***UI12FAZ***, with which a more detailed description is given, transcribed with respect to the convention proposed by Du Bois et al. (1993).

Example of speaker *NI18MBP*

"… then the side hall 有什么 (/joʊ3//ʂən2//mɤ5/)***(has any)*** sports subcommittee, 还有 (/xaɪ2//joʊ3/)***(and there is also)*** business group. 我 (/uɔ3/)***(I)*** take electives 啊 (/a5/)***(Particle)***, take electives 要看嘞 (/jɑʊ4/ /kʰan4//leɪ5/)***(it depends)***. I think 我会 (/uɔ3//xweɪ4/)***(I will)*** take 那些 (/na4//ɕjɛ1/)***(those)***, 嗯就是 ((/en1//tɕjoʊ4//ʂɚ4/)***(well just those)***, 我 (/uɔ3/)***(my)*** senior

跟我讲很容易 (/kən1//uɔ3//tɕjɑŋ3//xən3//zʊŋ2//i4/)***(told me that it is really easy to)*** score 的 (/tɤ5/)***(PARTICLE)***."

Number	Pattern
(1)	**DP(E)** *'the side hall'*+**V(M)** *'有, has'*
(2)	**PRO(M)** *'什么, any'*+**NP(E)** *'sports subcommittee'*
(3)	**NP(E)** *'sports subcommittee'*+**VP(M)** *'还有, also has'*
(4)	**VP(M)** *'还有, also has'*+**NP(E)** *'business group'*
(5)	**PRO(M)** *'我, I'*+**VP(E)** *'take electives'*
(6)	**N(E)** *'electives'*+**PARTICLE(M)** *'啊, PARTICLE'*
(7)	**AUX(M)** *'会, will'*+**V(E)** *'take'*
(8)	**V(E)** *'take'*+**DET(M)** *'那些, those'*
(9)	**DET(M)** *'我（的）, my'*+**N(E)** *'senior'*
(10)	**N(E)** *'senior'*+**VP(M)** *'跟我讲, told me'*
(11)	**ADV(M)** *'容易, easily'*+**V(E)** *'score'*
(12)	**V(E)** *'score'*+**PARTICLE(M)** *'的, PARTICLE'*

Table 4 Syntactic patterns of speaker ***NI18MBP***

The paragraph above is extracted from speaker ***NI18MBP***, for the content in Mandarin, the translation in English is provided in bold italic letters. The first CS occurs in the first line, between the English determiner phrase “the side hall” and the Mandarin verb “有 (/joʊ3/)***(has)***”, thus this CS is noted as ***DP(E)+V(M)***. The second CS also takes place in the first line, between the Mandarin pronoun “什么 (/ʂən2//mɤ5/)***(any)***” and the English noun phrase “sports subcommittee”, thus this CS is noted as ***PRO(M)+NP(E)***. Then, taking the CS in the second line as an example, since it is between the English noun “electives” and the Mandarin particle “啊 (/a5/)***(Particle)***”, thus it is noted as ***N(E)+PARTICLE(M)***. Sometimes a switch contains no categorical patterns. For example, a switch takes place between a main clause and a subordinate clause, then it will not be noted down. The same procedure applies to all the CS points in this paragraph, and all the syntactic patterns are noted. Then, with the same procedure, the transcription and the identification of the next two speakers are listed.

Example of speaker *NI20MBP*

Transcription:

I am hoping that will happen and probably I will 认识多一点 (/ʐ̩ən4//ʂɚ4//tuɔ1//i1//tjɛn3/)***(know more)*** Japanese, and next time I can go there for free subcom.

Identification:

Number	Pattern
(1)	**AUX(E)** *‘will’*+**VP(M)** *‘认识多一点, know more’*
(2)	**VP(M)** *‘认识多一点, know more’*+**N(E)** *‘Japanese’*

Table 5 Syntactic patterns of speaker ***NI20MBP***

Example of speaker *UI12FAZ*

Transcription:

然后只是我们 (/ʐ̩an2//xoʊ4//ʈʂɚ3//ʂɚ4//uɔ3//mən5/)***(then it’s just we)*** 这边因为太 (/ʈʂɤ4//pjɛn1//ɪn1//weɪ2//tʰaɪ4/)***(it’s just too)*** stressed 了 (/lɤ5/)***(PARTICLE)***. 老板还是叫你 (/l ɑ ʊ3//pan3//xaɪ2//ʂɚ4//tɕj ɑ ʊ4//ni3/)***(the supervisor still asks you)*** 不要去

(/pu2//j ɑ ʊ4//tɕʰy4/)***(not to)*** 管那些 (/kwan3//na4//ɕjɛ1/)***(mind those)*** noise 咯 (/lɔ5/)***(PARTICLE)***. Then 你 (/ni3/)***(you)*** concentrate on your work.

Identification:

Number	Pattern
(1)	**ADV(M)** *'太, too'*+**ADJ(E)** *'stressed'*
(2)	**ADJ(E)** *'stressed'*+**PARTICLE (M)** *'了, PARTICLE'*
(3)	**DET(M)** '那些, those'+**N(E)** 'noise'
(4)	**N(E)** *'noise'*+**PARTICLE(M)** *'咯, PARTICLE'*
(5)	**ADV(E)** *'then'*+**PRO(M)** *'你, you'*
(6)	**PRO(M)** *'你, you'*+**VP(E)** *'concentrate on your work'*

Table 6 Syntactic patterns of speaker ***UI12FAZ***

Chapter 5 Discussion

Chapter 1 and Chapter 2 introduce that for years, linguists have been searching for syntactic constraints behind CS to explain this natural bilingual behavior. However, there are still a number of examples that cannot be explained by them. Shenk (2006) was among the first ones to explore solutions from prosodic aspects. By applying the information-based approach in the data processing, she found that bilinguals massively code-switch at IU boundaries. Thus, she proposed that prosodic constraints also control CS behaviors. This proposition is corroborated by studies on certain language pairs. With support from more future studies in various other language pairs, linguists shall reach a consensus on the existence of prosodic constraints. This will provide possibilities to understand CS behaviors in a more comprehensive way. For example, it will help explain some of the counterexamples provided by the critics of the syntax-only approach. To this day, few studies have been conducted on the Mandarin-English language pair concerning prosodic constraints. Therefore, the current study means to provide contribution in this area. With the current Mandarin-English CS corpus, this study finds evidence to support Shenk's (2006) proposition. Furthermore, with the additional information provided by the investigation on syntactic patterns at switching points, the results of this study are even more well-founded.

Following Shenk (2006), this study applies the information-based approach to conduct data processing. Based on the methodology in previous studies (proposed by Chafe (1994), then applied by Shenk (2006) and many other linguists as introduced in Chapter 2), the current study optimizes it and develops a systematic methodology (introduced in Chapter 4). The current database contains data from 16 distinguished speakers, 8 from the conversation group and 8 from the interview group. For the conversation group, 21min of audio data and the data from the corresponding transcription are extracted to conduct the analysis. As for the interview group, 15min of audio data and the data from the corresponding transcription are extracted. The audio data is divided up into IUs and transcribed using the transcription convention introduced in Section 3.3.

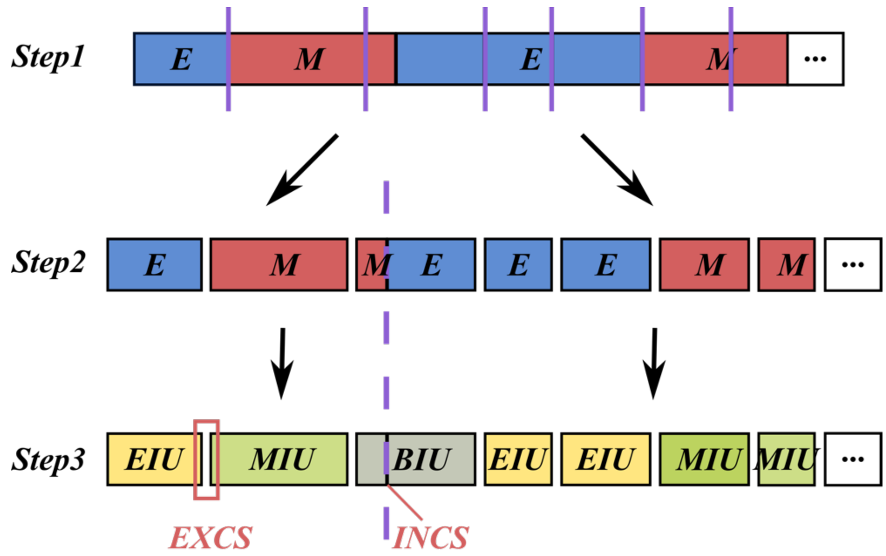


Figure 26 Illustration of the identification and classification of IUs

As in the review of Shenk's study in Section 2.2.3, Figure 26 illustrates that the IUs are identified and classified in a similar way in the current study. Step 1: the four features (introduced in both Chapter 2 and Chapter 4) are applied to identify the IUs. Step 2: the original data is divided up into different IUs. Step 3: the IUs are categorized into monolingual English IUs (EIU, as in English Intonation Unit), monolingual Mandarin IUs (MIU, as in Mandarin Intonation Unit), or bilingual IUs (BIU, as in bilingual Intonation Unit). To determine if the CS takes place at IU boundaries or within an IU, as indicated in Step 3, they are categorized as either EXCS (indicating CS external to IUs, i.e., at IU boundaries), or INCS (indicating CS internal to IUs).

In this chapter, the results of the data processing are presented, then a thorough discussion is developed after each presentation. First, in Section 5.1, the IU distribution and the CS distribution are presented. The low BIU percentage and the low INCS percentage support Shenk's (2006) conclusion and the results are proved to be generally applicable to all speakers. Thus, one of the objectives of this study is fulfilled. In Section 5.2, first the syntactic patterns at switching points are investigated, then certain frequent patterns are concluded. Among the frequent patterns, some coincide with previous studies on Mandarin-English CS, others are the novel ones observed from the current database. This study offers initial explanation for the patterns' existence, this fulfills another objective is of the study. Then, in Section 5.3, a synthetic discussion is conducted on the results of the of the previous two sections. It is observed that a specific pattern brings interference to the final results. With this pattern removed, different results are obtained. Following this, based on the modified results, a thorough discussion on the individual differences in IU and CS distribution is developed in Section 5.4. Finally, in Section 5.5, this study makes an effort to explore the issue with respect to the absence of a database appropriateness parameter discussed in Section 2.3.3. Based on the results and analysis of the current database, a quantitive parameter is proposed. The third objective of this study is fulfilled thanks to the quantitive parameter and the systematic methodology developed in Chapter 4.

5.1 IU distribution and CS distribution

The current section presents the results of IU distribution and the CS distribution, this concerns the main objective of the current study. The IU categorization and its percentage calculation method follow the steps of previous work in the field (Chafe, 1994; Shenk, 2006; Tao, 1996; Urrea, 2012). Figure 27 illustrates the distribution of the IUs of all 16 speakers. A total number of 14,163 IUs are identified. Within which, the total number of EIU is 5303, which takes up 37.4% of the IUs. The total number of MIU is 7356, and that makes it take the biggest percentage out of the three types, that is 51.9%. Finally the total number of BIU is 1504, which makes it the smallest type of IU and takes up 10.6% of the total number.

This result agrees with Shenk (2006) and many later studies focusing on the prosodic constraint in CS behaviours (Cacoullos & Travis, 2010; Manfredi et al., 2015; Mettouchi, 2008; Myslin & Levy, 2015; Travis & Cacoullos, 2013; Urrea, 2012; Vargas, 2008). In her study of Spanish-English CS, Shenk (2006) obtained a mere 4% of BIU. This led her to conclude that based on her corpus, bilinguals mostly code switch at IU boundaries, which then could be taken as clear evidence that prosody plays a crucial role in the CS behavior. Apart from the existing syntactic constraints, prosodic constraints also control CS behavior. She then proposed that the most robust unit in CS behavior is prosodic, rather than syntactic. However, the result of 10.6% BIU in the current study is not as small as the 4% in Shenk's (2006) study, but its results are similar to other previous studies. Section 2.2.3 introduces that the mean value of their BIU percentage is approximately 10%. It is concluded from this distribution that Chinese-English bilinguals in this database also prefer to code switch at IU boundaries. This leads to the conclusion that CS behavior is indeed controlled by prosodic constraints.

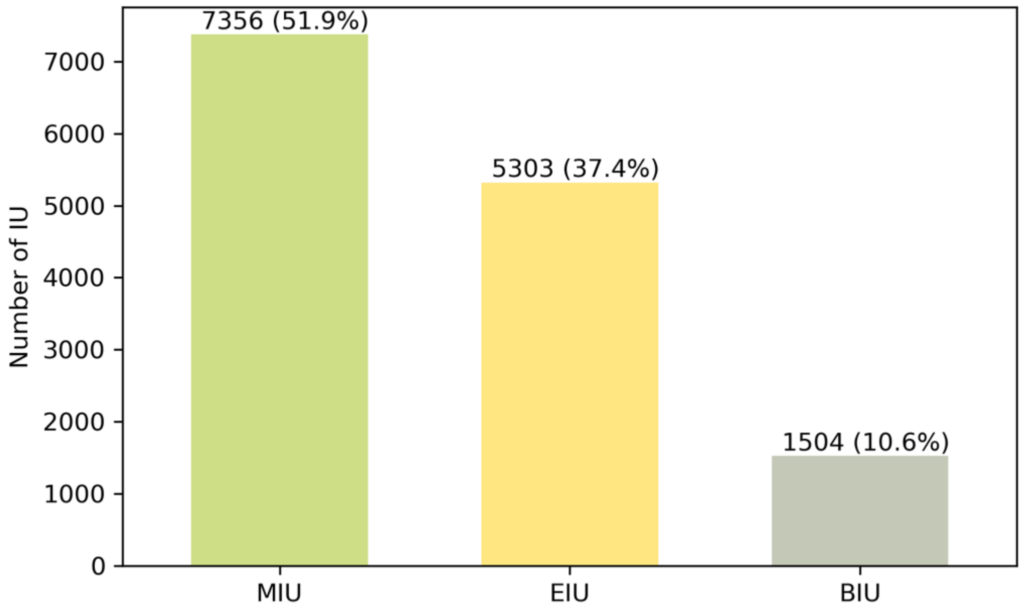


Figure 27 Distribution of the three types of IU of all 16 speakers

Section 2.2.3 presents that Shenk (2006) was among the first to apply the information-based approach in CS studies. Her conclusion has received support from several studies on various language pairs. However, this is the first time that Shenk's proposition receives support from Mandarin-English language pair. In fact, unlike the Spanish-English language pair or the French-English language pair, on which many studies are conducted (introduced in Section 2.1 and 2.2), previous studies on this language pair mainly focused on syntactic aspects. The current study is the first to explore prosodic constraints on this language pair. The result and conclusion require support from future studies. Besides, it is introduced in Section 2.3.2 that the methodology in this field suffers from certain subjectivity problem. In Chapter 4, the current study tries to optimize the methodology, which has reduced the problem as much as possible. Even so, the systematic methodology developed in the current study requires examination from many future studies, especially in this language pair.

The above results are evidence that prosodic constraints control CS behavior. However, it is possible that significant differences exist among individuals. In order to verify that the conclusion is applicable to every individual, other than exploring the global IU distribution, the current section proposes to investigate the individual IU distribution. This helps make the conclusion more reliable.

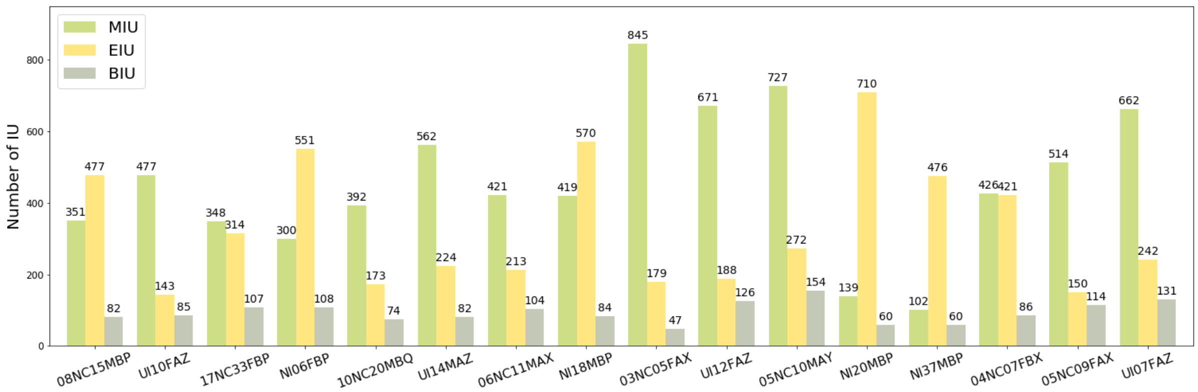


Figure 28 Individual distribution of different types of IU

Figure 28 presents the individual IU distribution of the current database. It is observed that the percentage of MIU and EIU between different individuals can be significantly different. For example, speaker ***UI12FAZ*** creates a very large amount of MIU and a very small amount of EIU, whereas the result is reversed for speaker ***NI20MBP***. Generally speaking, the BIU percentages of all speakers stay at a low level. However, the MIU and EIU percentages differ greatly among individuals. This distribution testifies that, generally speaking, CS behaviour is controlled by prosodic constraints. In fact, previous studies might have conducted an analysis on individuals concerning sociolinguistic factors. For example, Urrea (2012) conducted a thorough discussion on how social factors influence the language usage of different speakers, such as age, education, occupation etc. Travis & Cacoullos (2013) focused on how the language dominance of different speakers had an impact on the BIU percentage in their Spanish-English database. They all came to interesting conclusions based on these factors. It is the first time an explicit individual distribution like Figure 28 is presented. This distribution testifies that this conclusion is generally applicable to all speakers.

Even though the IU distribution is well established and discussed, according to the observation of data processing and results analysis, it might bring issues to the results if solely the IU distribution is investigated. Further exploration is obligatory. The above results illustrate that bilinguals prefer to code switch at IU boundaries, and that prosodic constraints do control CS behaviors. With the individual distribution, the conclusion is testified to be applicable to all individuals, which helps make it more reliable.

However, there is still some underlying uncertainty. The conclusion that bilinguals prefer to switch at IU boundaries is based on the low BIU percentage. This value is calculated based on the number of different types of IU. With more monolingual IUs, the percentage of BIU naturally gets smaller. Note that so far, the actual number of CS is not considered. Under most circumstances, there is one switch of language in a BIU. In a very limited number of cases, there is more than one switch in a BIU. Consequently, the numbers of INCS and BIU are practically the same, with the former slightly larger than the latter. Thus, the amount of BIU reflects rather directly the number of INCS, illustrated in Figure 26. Yet, a large amount of MIU+EIU do not necessarily denote a large amount of EXCS. There exists a small possibility that even though the amount of MIU+EIU is large, the amount of CS is still small. In that case, even though a small BIU percentage is obtained, the inference about the preferred switching points can no longer be made. For instance:

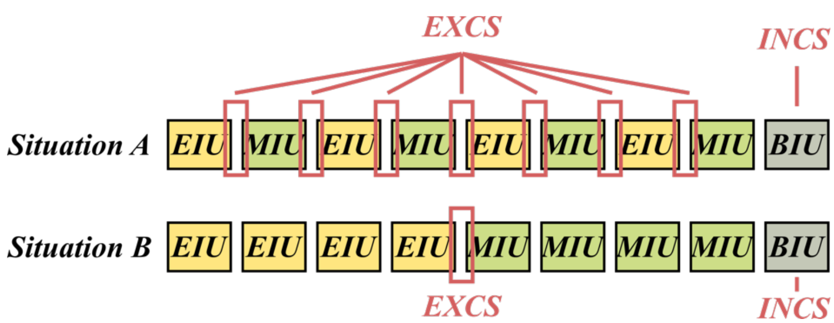


Figure 29 Two hypothetical extreme situations of the distribution of different types of IU

Figure 29 illustrates two hypothetical extreme situations: in Situation A, there is a large amount of CS happening at IU boundaries, thus EXCS is way higher than INCS; whereas in Situation B, the number of EXCS equals to that of INCS. It is not hard to spot that in these two situations, the BIU percentages in terms of total IU is the same. Obviously, based on Situation A, we could come to the conclusion that bilinguals mainly code switch at IU boundaries, and that the prosodic constraint controls CS behavior. As to Situation B though, no such conclusion can be obtained since bilinguals create as many internal CS as external CS. No such prosodic constraint even exists. As a consequence, the low BIU percentage is not caused by the constant switching at IU boundaries (Situation A), but rather, it is simply caused by a high amount of monolingual content (Situation B). This would then make the results in the previous sections unreliable. Even though the above discussion makes this matter seem quite straightforward, to the best knowledge of the author, it has not been considered in previous work in the field (Cacoullos & Travis, 2015; Shenk, 2006; Urrea, 2012; Vargas, 2008). The relationship between IU and CS has not been thoroughly discussed, the possible extreme situation as in Figure 29 has not been considered. It might be one of the reasons that cause the issue presented in Section 2.3.3. According to the database of the current study, the extreme situation discussed above is a low probability event. But it is still necessary to conduct a quantitative analysis to prevent doubts in an objective way.

Most researchers tend to look at one phenomenon, either the IU distribution or the CS distribution. Shenk (2006) investigated IU distribution. Then, Section 2.3.3 reviews that in her doctorate dissertation, Urrea (2012) followed Shenk (2006) and applied the information-based approach to investigate the role played by prosodic constraints. Instead of looking into the IU distribution, Urrea focused on the CS distribution. The 21% INCS percentage led her to conclude that only a small part of the CS takes place inside the IU, bilinguals tend to code switch at IU boundaries. Based on this result, she supported Shenk by concluding that the prosodic constraint plays an important role in controlling CS behaviors. Following Urrea's work, the current study also investigates the CS distribution, in order to testify that the extreme situations in Figure 29 are not an issue in this database. The results are presented below.

Figure 30 illustrates that during the 288 min worth of audio of the 16 speakers, a total number of 6933 switches are created. 5323 of them are external to IU (i.e. EXCS), which takes up 76.8 % of the total CS; the other 23.2% is taken by CS internal to IU (i.e. INCS). Since the result of 23.2% is quite close to the result of 21% in Urrea (2012), the current study follows Urrea and concludes that speakers create only a small part of CS inside the IU. The low percentage of BIU is caused by the constant switches at IU boundaries, and not by the massive monolingual content. It is worth pointing out that the total number of INCS is 1610. This is quite close to the 1540 total BIU number presented in Figure 27. The relation between them is discussed above.

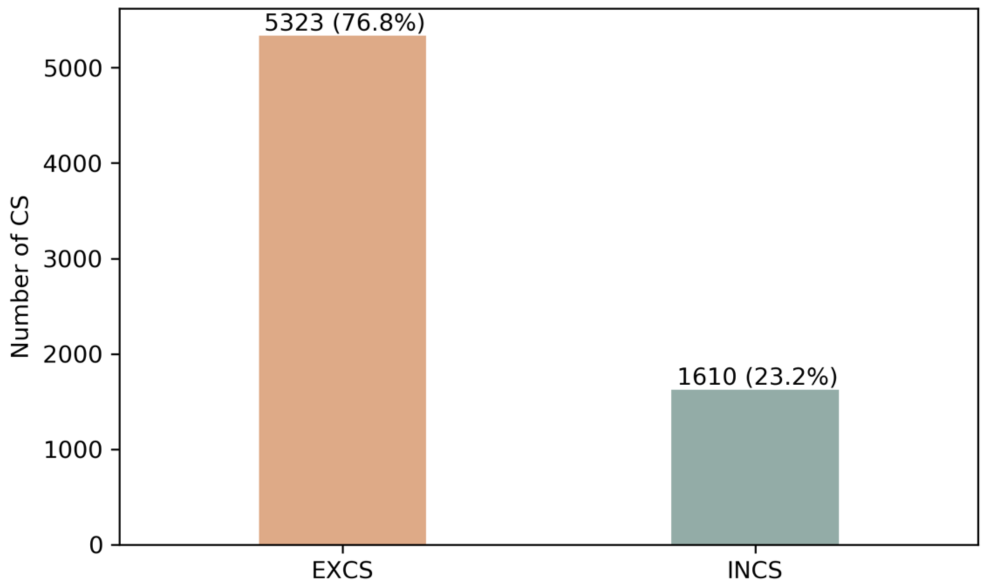


Figure 30 Distribution of INCS and EXCS of all 16 speakers

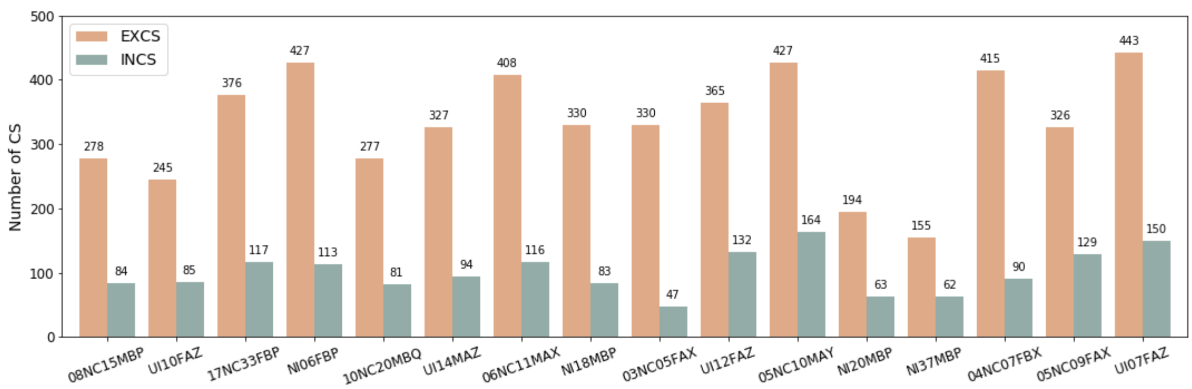


Figure 31 Individual distribution of INCS and EXCS

For the same reason the individual analysis on IU was conducted, the individual CS distribution is also investigated. Figure 31 gives a clear presentation of the CS distribution of each individual. The difference between the number of EXCS and INCS differs significantly from individual to individual. For the EXCS, speakers like ***NI06FBP*** and ***UI07FAZ*** create more than 400, whereas speakers like ***NI20MBP*** and ***NI37MBP*** create lower than 200. Even though the numbers of INCS remain at a low level for all the speakers, the individual difference is still not negligible. Speaker ***05NC10MAY*** creates the highest amount (164) and speaker ***NI37MBP*** creates the lowest amount (47). Based on the distribution, it is observed that the conclusion above is generally applicable to all individuals. All speakers create only a small part of CS inside the IU. Thus, it is clear that the extreme situation presented in Figure 29 is not an issue for any speaker of the database.

Based on the above discussion, the direct evidence of prosodic constraints is found in the current database, and thus it provides support to Shenk's (2006) proposition from Mandarin-English language pair. This fulfills one of the objectives of the research.

5.2 Syntactic patterns discussion

In order to obtain a more profound understanding of a linguistic phenomenon, it should be explored in the most comprehensive way possible. From the development of the field introduced in Chapter 2, we see that linguists are trying to conduct comprehensive research with respect to CS behaviour: at first, they mainly focused on the syntactic aspect (introduced in Section 2.1). However, when they encountered problems, they turned to the prosodic aspect for solutions (introduced in Section 2.2). Thus, even though the main objective of the current study is to provide support for Shenk's (2006) proposition about prosodic constraints based on Mandarin-English language pair, the syntactic patterns are still investigated. Previous studies on Mandarin-English CS focused on investigating syntactic patterns at switching point, applying the notion of syntagm. They found similar frequent patterns (introduced in Section 2.1.3.2). Considering that the current corpus was collected from a rarely explored bilingual community, this study means to follow previous research, and investigate the syntactic patterns at switching points. The result illustrates that in this corpus, not only the frequent patterns in previous studies are found, but also new frequent patterns appear. This offers additional information that helps make the results from the prosodic aspect more solid. The current section conducts a discussion based on them.

It is observed in Figure 32 that, first the syntactic patterns at the switching points are identified, and then they are categorized as Pn (n=1,2,3…) using the methodology presented in Chapter 4. Then, with respect to different types of CS, the patterns are corresponded with either EXCS or INCS. This is how the connection between the patterns and the CS is established. For each of the 16 speakers, generally over 20 types of patterns are created. Four patterns are most frequent for all speakers. The coding notation is presented in Table 3 in Chapter 4. Then, Table 7 illustrates the individual details of the four patterns, which provides a closer look at their distribution. The percentage of the four patterns is listed in Table 7. As a comparison, the percentage of all other patterns combined (generally over 20 types, named *Pattern Rest*, PR) is also presented for each speaker. This provides a clearer illustration of how P1-P4 significantly outnumbers the other patterns, which is also the reason why these four patterns are discussed in the current section.

(1) Pattern 1 (P1): V/VP(M)+N/NP(E)

(2) Pattern 2 (P2): DET(M)+N/NP(E)

(3) Pattern 3 (P3): AUX(M)+V/VP(E)

(4) Pattern 4 (P4): N(NP)/V(VP)/ADJ(AP)/ADV(AdvP)/DET(DP)(E)+Particle(M)

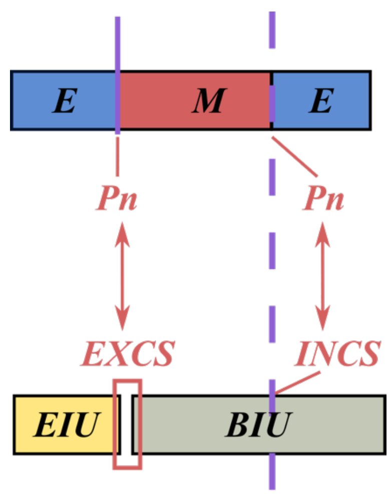


Figure 32 Connection between patterns and CS

P1 and P2 coincide with previous studies. It is introduced in Section 2.1.3.2 that, Lu (1991) and Tan (1988) found that bilinguals tend to let the CS fall on certain patterns such as P1 and P2. Lu (1991) provided initial assumption on the reason behind, that is due to the ease of expression and the influence of interlocuters. Tan (1988) did not provide an explanation. Ong & Zhang (2010) focused on discussing the existence of P2. According to them, bilinguals tend to use Mandarin determiners because of the *Lemma versatility filter*. People subconsciously prefer simpler forms of speech from either lexicon. It is reviewed in Section 2.1.3.1 that, Mandarin determiners allow for more economy of speech. Then, the *Grammatical feature filter* explains why bilinguals tend to use English nouns. This filter directs people to choose the language that possesses more grammatical features. Section 2.1.3.1 presents that English nouns inflect for number whereas Mandarin nouns do not. In fact, the current study thinks that the two filters discussed by Ong & Zhang (2010) can explain not only P2, but also P1. In the introduction of the syntactic and morphological characteristics of Mandarin in Section 2.1.3.1, it is mentioned that compared to inflectional languages such as English, Mandarin verbs do not change forms according to tense, subject-verb agreement, case, gender or number. They maintain the same form under all circumstances. Thus, the *Lemma versatility filter* explains why people tend to use Mandarin verbs. Furthermore, as with P2, the choice of English nouns is motivated by the *Grammatical feature filter*. It is also observed in Table 7 that a significant difference exists between the number of INCS and EXCS in P1 and P2. Bilinguals tend to create this pattern at IU boundaries, rather than within the IU.

Speaker	P1			P2			P3			P4			PR
	EXCS	INCS	%	EXCS	INCS	%	EXCS	INCS	%	EXCS	INCS	%	%
NI18MBP	23	9	13.3	23	13	14.9	20	9	12.0	0	21	8.7	51
03NC05FAX	42	4	22.4	29	3	15.6	32	3	17.1	0	18	8.8	36.1
UI12FAZ	33	9	12.8	49	4	16.1	21	12	10.0	0	55	16.7	44.4
05NC10MAY	20	5	7.0	58	9	18.7	28	9	10.3	0	66	18.4	45.7
NI20MBP	26	2	15.9	22	2	13.6	26	0	14.8	0	43	24.4	31.3
NI37MBP	29	1	17.5	34	1	20.5	24	0	14.0	0	43	25.1	22.8
04NC07FBX	23	5	13.0	28	3	14.4	29	1	14.0	0	52	24.2	34.4
05NC09FAX	34	1	12.1	30	4	11.7	32	5	12.8	0	78	26.9	36.6
UI07FAZ	32	9	12.6	51	13	19.6	19	5	7.4	0	55	16.9	43.6
08NC15MBP	33	3	19.1	28	1	15.4	13	4	9.0	0	53	28.2	28.2
UI10FAZ	27	2	14.1	41	1	20.4	13	1	6.8	0	52	25.2	33.5
17NC33FBP	39	6	17.2	34	4	14.5	21	3	9.2	0	78	29.8	29.4
NI06FBP	17	2	6.7	32	4	12.7	21	2	8.1	0	81	28.5	44.0
10NC20MBQ	52	6	21.6	38	6	16.4	23	3	9.7	0	44	16.4	36.1
UI14MAZ	23	6	11.8	32	12	18.0	25	3	11.4	0	59	24.1	34.7
06NC11MAX	34	4	12.0	38	7	14.2	23	2	7.9	0	83	26.2	39.7

Table 7 Individual distribution of patterns. The percentages marked in red indicate the percentage of EXCS and INCS combined.

P3 and P4 have not been mentioned in previous studies, they are novel patterns observed in the current study. Previous sections introduces that the current database is collected in a rarely explored bilingual community in Singapore and Malaysia. Consequently, it is possible that these two patterns are specific to this community, more studies are needed to confirm this. Due to the limitation of analyzing data and method, a fair explanation of P3 has not been provided yet. However, this study makes an attempt to explain P4.

P4 is really a special case that is worth mentioning. Particles or grammatical particles are a special group of words that exist in Mandarin and usually appear at the end of a phrase or a sentence. Prosodically, Section 2.1.3.1 introduced five lexical tones in Mandarin, particles take the fifth neutral tone. There is no specific tone set to this fifth one, they are free to be pronounced with the tendency of the phrasal intonation. Syntactically, they belong to the category of function words. No lexical meaning is set to them and they are usually applied to express certain grammatical information. Table 8 provides a detailed list of some of the particles in Mandarin that appear in the database of the current study, with more specific information, including examples extracted from the database.

The most peculiar thing about P4 is that, for all speakers, this is observed from Table 7, they only appear inside the IU and never at IU boundaries. Syntactically, for Mandarin-English bilinguals, under most circumstances, Mandarin particles are used at the end of a phrase solely as an auxiliary word, or a modal particle that indicates a mood. They may not even express anything, they occur just because of habit. Prosodically, as is already outlined, no particle possesses any lexical tones, they are pronounced as the phrasal intonation goes. The prosodic and syntactic peculiarity gives them this pattern and why they only appear inside IUs. Compared to other bilingual IUs, an English IU with a Mandarin particle in the end is apparently less "bilingual". It is more of an expression out of customs and convenience. Other than this, it is observed in Table 7 that P4 is very recurrent for all the speakers.

It is introduced above that previous studies did not mention the existence of P4 (Lu, 1991; Ong & Zhang, 2010; Tan, 1988; Zheng, 2005). The current study speculates three possible reasons. The first reason is that P4 does not exist in their data. The second reason is that P4 exists in their data, but researchers chose to not regard it as CS, perhaps due to the syntactic and prosodic characteristics of Mandarin particles introduced above. The third reason is that P4 exists in their data, researchers found it but chose to exclude it from their results because they might deem that it possesses no theoretical value or practical meaning. No matter the reasons, P4 never appeared in the results and discussion of previous studies. However, if integrated with the concept of IU, this study thinks that P4 possesses great importance. According to Table 7, P4 appears within IU 100% of the time with very high frequency, but never at IU boundaries. P4 is a fine example of how syntactic and prosodic characters work together to influence people's language usage. Unfortunately, the current study could not find out whether P4 exists generally in Mandarin-English CS, or it only exists in the specific bilingual community of the current corpus. Thus, researchers are encouraged to present their findings in the most comprehensive and objective way possible. It is likely that one outcome seems meaningless in one study, but if more concepts are introduced in future studies, the same outcome could become meaningful.

Particle	Pinyin	IPA	Word preceding	Functional information	Example
啦	la	/la5/	N/ADJ/V…	Auxiliary word	*I quite enjoy the course 啦.*
啊	a	/a5/	N/ADJ/V…	Auxiliary word	*Then I 参加的 (participate in) basketball 啊.*
嘛	ma	/ma5/	N/ADJ/V…	Auxiliary word	*I was transferred to the Chinese high school 嘛.*
咯	lo	/lɔ5/	N/ADJ/V…	Auxiliary word	*Okay 咯.*
了	le	/lɤ5/	N/ADJ/V…	Expressing the perfect	*I'm going to graduate 了.*
的	de	/tɤ5/	N/ADJ/V…	Auxiliary word	*It's interesting 的.*

Table 8 Examples of Mandarin particles

However, due to the lack of further syntactic theoretical support, a solid explanation for P3 is not provided. Since the current study is based on one database, the generalization of P1-P4 still requires support from other studies. Previous studies could provide additional information on their results, future studies could examine the existence of the patterns. For example, do their results back up P3 and P4? Is there a solid explanation on why P3 exists? Are there counterexamples to P1 and P2 (for example, V/VP(E)+N/NP(M) are more frequent)? With more evidence, we could learn more about people's language usage.

Based on the discussion above, the syntactic patterns at switching points are investigated and presented. P1 and P2 coincide with previous studies. A comparison is conducted between them and the current study and further discussion is provided. P3 and P4 are novel patterns observed in the current study and some explanations are provided. A discussion is then conducted on the theoretical and practical meaning, as well as the limitations of this point in the current study. Thus, one of the objectives of this study is fulfilled.

5.3 Modified IU distribution and CS distribution

Based on the discussion in Section 5.2, P4 is peculiar. First, it only appears within IUs, and it takes up a big percentage among the patterns for all speakers. Second, the syntactic and prosodic characteristics of the Mandarin particles in P4 make this pattern less bilingual than other patterns. Thus, P4 is a special case and is considered to interfere with the final results. For the global and the individual distribution of IU, P4 is taken out of the final results. The results of comparison are illustrated in Figure 33 and Figure 34, followed by the individual distribution of IU after P4 is excluded in Figure 35.

Figure 33 illustrates the comparison of the distribution of different IUs before and after taking out P4. The number of MIU and EIU has not changed, yet their percentages have changed because, as is discussed in previous sections, P4 only takes place within the IU and never at IU boundaries. Taking it out therefore only affects the number of BIU. Since P4 takes up more than half of the total BIU in the previous result, taking it out has caused the percentage of BIU to drop from 10.6% to 4.7%. Note that in the current study, the conclusion remains the same with or without the interference. With consultation to previous work (Cacoullos & Travis, 2010; Manfredi et al., 2015; Mettouchi, 2008; Myslin & Levy, 2015; Shenk, 2006; Travis & Cacoullos, 2013; Urrea, 2012; Vargas, 2008), both 10.6% and 4.7% are small enough to lead to the conclusion that bilinguals mainly code switch at IU boundaries. Support for Shenk's proposition is obtained either way. However, the significance of this should not be ignored. By taking this interference into consideration, effects are brought to the results.

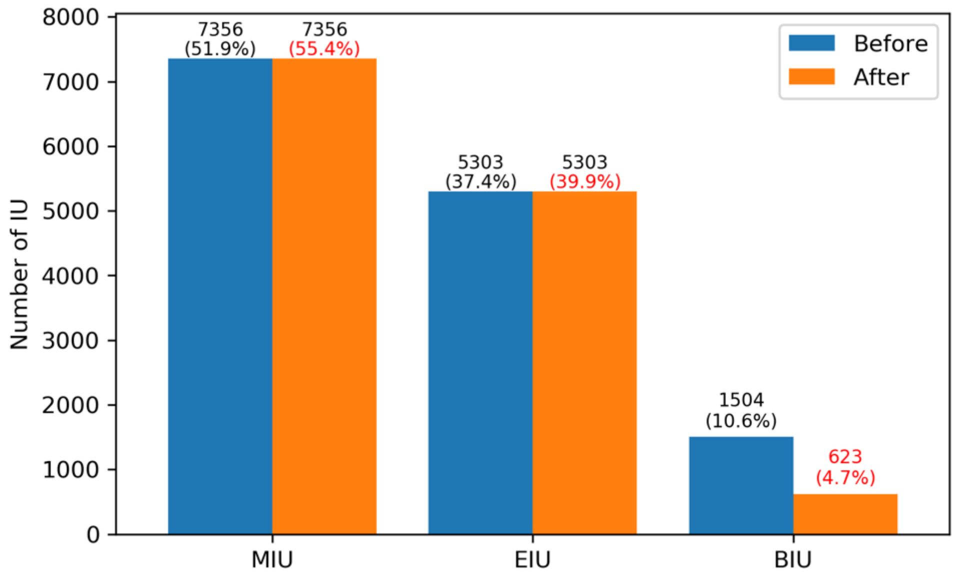


Figure 33 Comparison of IU distribution before and after omitting P4

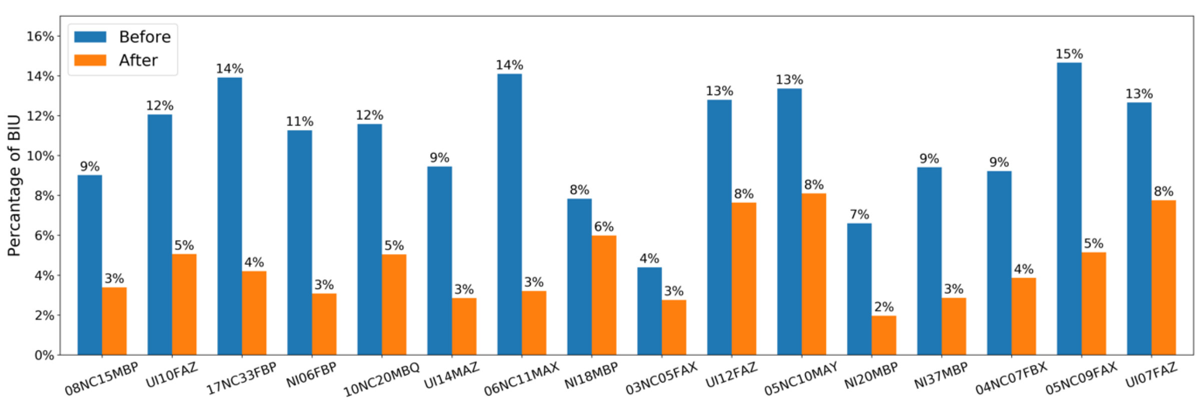


Figure 34 Individual comparison of BIU distribution before and after omitting P4

Figure 34 presents the individual comparison of BIU before and after taking out P4, in order to make sure that this phenomenon applies generally to all individuals. P4 does influence every single individual of the database, with slight individual differences. For example, the BIU percentage of speaker ***06NC11MAX*** has dropped greatly from 14% to 3%. Yet for speaker ***NI18MBP***, the influence of the interference is not that significant, the BIU percentage drops only from 8% to 6%. Besides, when taking into account all 16 speakers before removing the interfering results, the personal deviation is rather big. BIU percentages range from 4% to 15%, with big variations among individuals. After removing the interference, the personal deviation is significantly reduced. BIU percentages range from 2% to 8% and the majority of them stay around 3% and 4%.

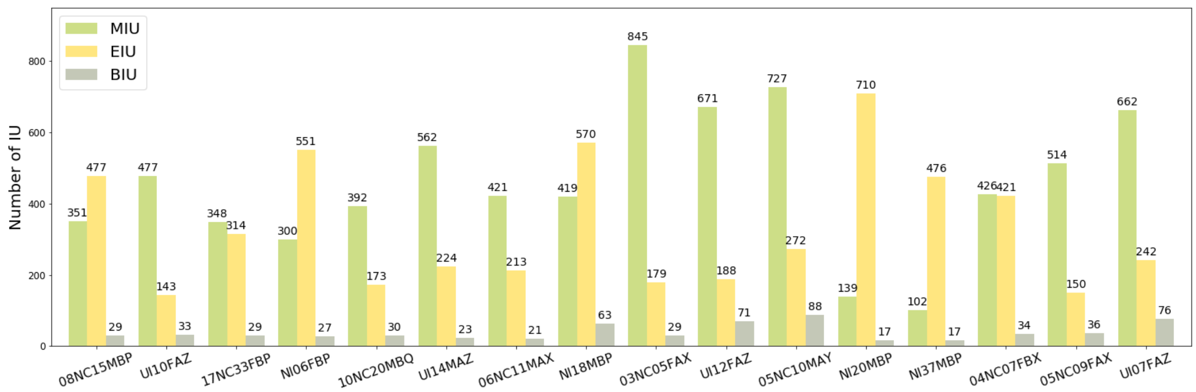


Figure 35 Individual distribution of different types of IU after omitting P4

Figure 35 presents the individual distribution of IU after omitting P4. Compared to Figure 28 in Section 5.1, it is observed that the number of MIU and EIU does not change, yet the number of BIU decreases rather significantly for all speakers.

Since P4 takes place within IUs 100% of the time, then it should only take place in INCS. Thus, this pattern should also interfere with the CS distribution. Therefore, for the global and the individual distribution of CS, P4 is taken out of the final results. The results of comparison are illustrated in Figure 36 and Figure 37, followed by the individual CS distribution after P4 is excluded in Figure 38.

Figure 36 illustrates the comparison of EXCS and INCS before and after removing P4. Just as explained previously, since P4 only happens within IUs, all of them are categorized as INCS. Thus the number of EXCS has not changed, yet the number of INCS has, just like the number of BIU was halved. The percentage of INCS drops from 23.2% to 12.0%, which causes the percentage of EXCS to raise from 76.8% to 88.0%. The individual comparison of the percentage of INCS illustrates that the existence of P4 has a great effect on every individual. For certain individuals such as ***06NC11MAX*** and ***NI20MBP***, the majority of INCS created by them is P4.

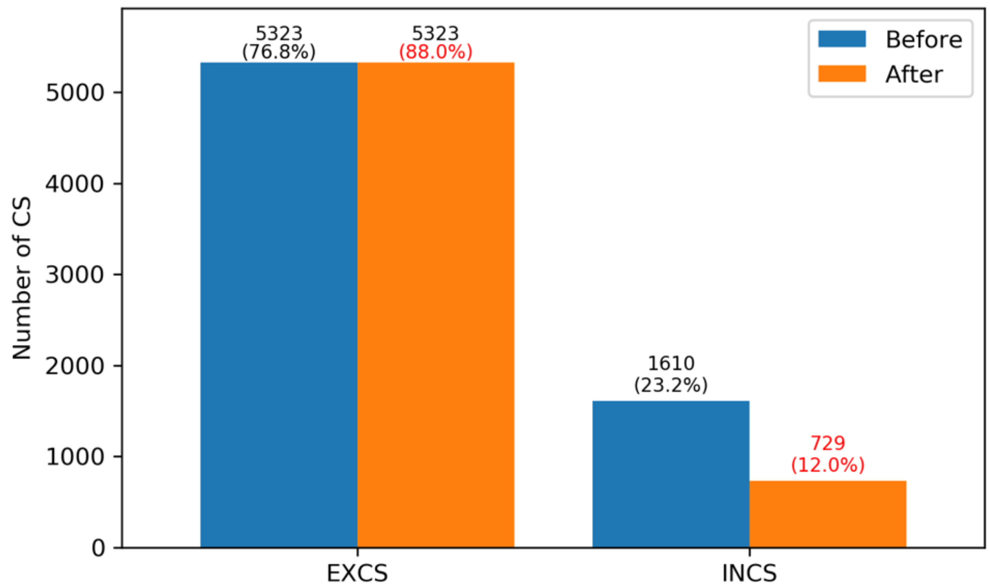


Figure 36 Comparison of CS distribution before and after omitting P4

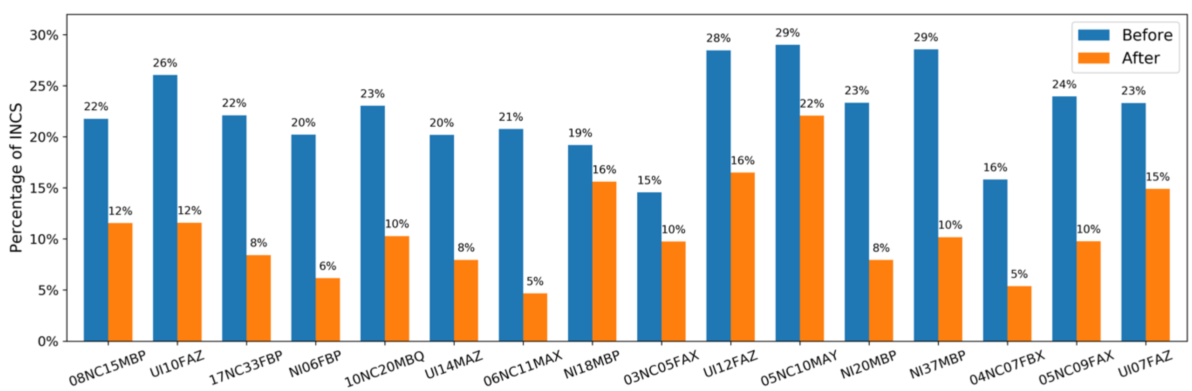


Figure 37 Individual comparison of INCS distribution before and after omitting P4

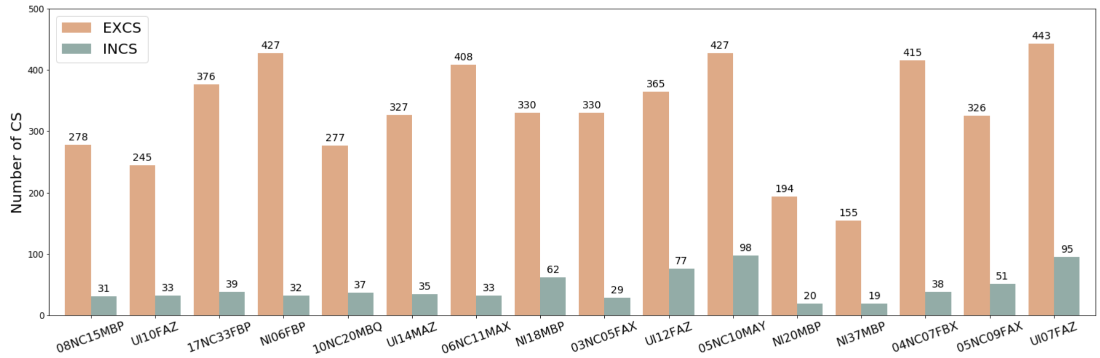


Figure 38 Individual distribution of CS after omitting P4

Figure 38 presents the individual distribution of CS after omitting P4. Compared to Figure 31 in Section 5.1, it is observed that the number of EXCS does not change, yet the number of INCS decreases rather significantly for all speakers. Just as the discussion on the effect of P4 on IU distributions, in terms of the CS distributions, certain effects are obtained by removing P4 from the final results.

Based on the result, an assumption is provided. Similar patterns like P4 might exist in other language pairs. If there are more than one interfering pattern, or if the number of similar patterns is too large, it might affect the IU percentage. This might bring more obstacles to the research of prosodic constraints. This means that sole observations on prosodic aspects might not help find certain phenomena. If additional information is provided from the syntactic aspects, some interferences might be located. The truth behind certain linguistic phenomena could then be found. Then again, for the current study, the conclusion remains the same with or without the interference. Future studies are encouraged to take the possibility of interference into consideration. For certain studies, it might have a significant impact on the outcome and turn bad results into good ones. For instance, if one encounters a bad outcome from the prosodic aspect, one could turn to the syntactic aspect to look for the possible existence of the interference. To understand any kind of linguistic phenomenon, the research should be as comprehensive as possible. So far, very few studies have explored both the prosodic and syntactic aspects in this field, especially in Mandarin-English CS studies. The existence of interference (such as P4 observed in the current study) requires support from future studies on various language pairs.

5.4 Individual influence discussion

Based on the discussion in previous sections, analysis for the rest of the chapter is conducted on results after P4 is removed. Other than solely presenting the IU and CS distribution, this study tends to make the most of the database and looks into the reason behind the individual differences, which concerns the personal information and the speech type etc.

5.4.1 IU distribution discussion

Table 9 illustrates the specific IU distribution of each speaker, as well as the individual information. The speakers are sorted by BIU percentage in an ascending order. Speaker ***NI20MBP*** holds the lowest percentage of BIU, 1.96%, which is quite low compared to the global result obtained by Shenk (2006); the highest percentage of BIU 8.10% is created by speaker ***05NC10MAY***. Even at a relatively low level, the BIU percentages do not remain at an absolutely stable value. BIU percentages vary within a small range. As for the MIU and EIU percentages, they change significantly. Although very few previous studies looked at the reason behind the changes (Travis & Cacoullos, 2013; Urrea, 2012), the current research makes an attempt to explore this.

Various factors could lead to the individual differences in the results. Specific to the current study, two major factors could be the possible reason: 1) The differences among distinguished speakers. For instance, age reflects the person's life experience and the educational level. These are the critical factors that form one's language habits. The SEAME team provided four personal categories of information: gender, nationality, age and speech type. Granted, the database is not big enough to conduct a proper sociolinguistic analysis. Despite that, this study makes the most of all the information at hand and tries to make a potential inference of the reasons behind the results. 2) The session progress. Considering this factor, the change of topics, the participant's attitude and interest etc. could all be potential influences on the results. A detailed analysis is conducted as shown below.

SPEAKER	EIU		MIU		BIU		Total IU	Gender	Nationality	Age	Speech type
	N	%	N	%	N	%					
NI20MBP	710	81.99	139	16.05	17	1.96	866	M	S	22	I
03NC05FAX	179	17.00	845	80.25	29	2.75	1053	F	M	21	C
UI14MAZ	224	27.69	562	69.47	23	2.84	809	M	M	30	I
NI37MBP	476	80.00	102	17.14	17	2.86	595	M	S	22	I
NI06FBP	551	62.76	300	34.16	27	3.08	878	F	S	20	I
06NC11MAX	213	32.52	421	64.27	21	3.21	655	M	M	23	C
08NC15MBP	477	55.66	351	40.96	29	3.38	857	M	S	21	C
04NC07FBX	421	47.79	426	48.35	34	3.86	881	F	S	20	C
17NC33FBP	314	45.44	348	50.36	29	4.20	691	F	S	23	C
10NC20MBQ	173	29.08	392	65.88	30	5.04	595	M	S	21	C
UI10FAZ	143	21.90	477	73.05	33	5.05	653	F	M	27	I
05NC09FAX	150	21.43	514	73.43	36	5.14	700	F	M	22	C
NI18MBP	570	54.18	419	39.83	63	5.99	1052	M	S	19	I
UI12FAZ	188	20.22	671	72.15	71	7.63	930	F	M	27	I
UI07FAZ	242	24.70	662	67.54	76	7.76	980	F	M	30	I
05NC10MAY	272	25.02	727	66.88	88	8.10	1087	M	M	23	C

Table 9 Distribution of the IUs of each speaker, with personal information included. (In the *Gender* column, *F* = Female, *M* = Male; in the *Nationality* column, *M* = Malaysian, *S* = Singaporean; in the *Speech Type* column, *I* = Interview, *C* = Conversation.

5.4.1.1 Personal information

Among the four personal categories of information provided by the SEAME team, *nationality* could be a possible reason behind the individual differences in results. Table 9 indicates that all eight of the Malaysian speakers possess a relatively high percentage of MIU, all higher than 60%. Among them, speaker ***03NC05FAX*** holds the highest MIU percentage, which is 80.25%; whilst speaker ***06NC11MAX*** creates the lowest MIU percentage of 64.27%. On the other hand, the eight Singaporean speakers possess a relatively high percentage of EIU, other than the 29.08 % created by speaker ***10NC20MBQ***, the majority of them create more than 50% of EIU. The highest is 81.99%, possessed by speaker ***NI20MBP***.

The above observation basically reflects the fact that in the current database, Malaysians prefer to use Mandarin and Singaporeans use English a bit more. In other words, it is reasonable to infer that for this database, Malaysians are Mandarin dominant, while Singaporeans are English dominant. Considering the history and development of the two nations, Malaysians and Singaporeans are proficient in both languages. In Singapore, starting from the year of 1979, there has been a *Speak Mandarin Campaign* initiated by the government. Prime Minister Lee Kuan Yew launched this campaign to encourage the use of Mandarin in this country (Leong, 2014). From then on, Singaporeans become more and more proficient in Mandarin. In Malaysia, Malaysian Chinese have always been trying to keep their Chinese culture and Chinese language, thus education in this area has never stopped in this community. Then with English being the official language in both countries, it is no wonder that people there handle both languages well. Thus, the BIU percentage is supposed to reflect to certain level the impact of a nation's culture and history. It is observed in Table 9 that most of the Malaysians tend to create a rather high percentage of BIU. In comparison, the Singaporeans in the database gather more around the low BIU percentage area. To conclude, Malaysian nationality has a positive influence on the usage of Mandarin which leads to a higher BIU percentage. On the other hand, the Singaporean nationality has caused a higher percentage of EIU which leads to a lower BIU percentage.

It is quite interesting to explore social factors that affect individual differences. However, the current corpus provides rather limited information on this matter. Thus, a more prolific sociolinguistic discussion could not be conducted. Previous studies mainly focused on certain specific CS phenomena, such as IU percentage (Shenk, 2006; Vargas, 2008) and syntactic patterns (Lu, 1991; Ong & Zhang, 2010). Only a few studies discussed the social factors (Travis & Cacoullos, 2013; Urrea, 2012). Among those who conducted a sociolinguistic discussion, researchers did the data collecting themselves. They were able to work up a specific data collection plan according to their own needs, such as the personal information and the length of the speech recordings. For example, before data collecting, Urrea (2012) planned specifically the speech recording, as well as the personal information such as age, sex, occupation, level of bilingualism, education level and community environment. As a result, other than supporting Shenk's (2006) proposition with the results from her data processing, Urrea (2012) was also able to conduct a comprehensive sociolinguistic analysis on her database. She obtained many observations from this analysis, and this has enriched her whole work. Compared to Urrea (2012), the database of the current study is formed from SEAME, a large-scale corpus purchased from Linguistic Data Consortium. While the SEAME corpus did a fine job on the collection and the presentation of the data, one has to admit that it provides very limited sociolinguistic information. In this study, solely the *nationality* is worth a discussion. Clearly a self-collected database offers more possibilities concerning the sociolinguistic analysis, whilst a purchased database has to make use of whatever provided information. In fact, a sociolinguistic discussion is of great importance. It provides the possibility to understand the CS behaviour in a profound and comprehensive way. Rather than simply obtaining a pattern in certain phenomena, this study looks into the reason behind it.

5.4.1.2 Session progress

The session progress could be another factor that influences the language usage of individuals. Since the current study means to investigate this matter, procedures are taken during the data extraction with respect to this phenomenon (for further details, refer to Section 3.2.1). Thus for each speaker, three periods of equal length are extracted at the beginning, in the middle and at the end of the audio. This makes it possible to observe more clearly the influence brought by the session progress.

For the majority of the speakers, the influence of the progress is not too significant. Thirteen of them show the same tendency between MIU and EIU for all three stages: MIU outnumbers EIU, EIU outnumbers MIU, or the numbers of the two types are very close. For example, in the results of speaker ***NI06FBP*** illustrated in Figure 39, the EIU outnumbers the MIU for all three stages, in which the difference between the number of MIU and EIU remains stable. The BIU number is not influenced by the session progress.

However, the stages seem to have a significant influence on three of the speakers. They do not follow the same tendency throughout the three stages. For example, Figure 40 illustrates the distribution of MIU and EIU during the three stages for speaker ***08NC15MBP***. For this speaker, MIU outnumbers EIU greatly in Stage 1, then EIU outnumbers MIU a little in Stage 2, yet in Stage 3 the number of EIU is significantly bigger than the number of MIU. The session progress has a big impact on the language usage of this speaker. Again, the BIU number is not significantly influenced by the session progress.

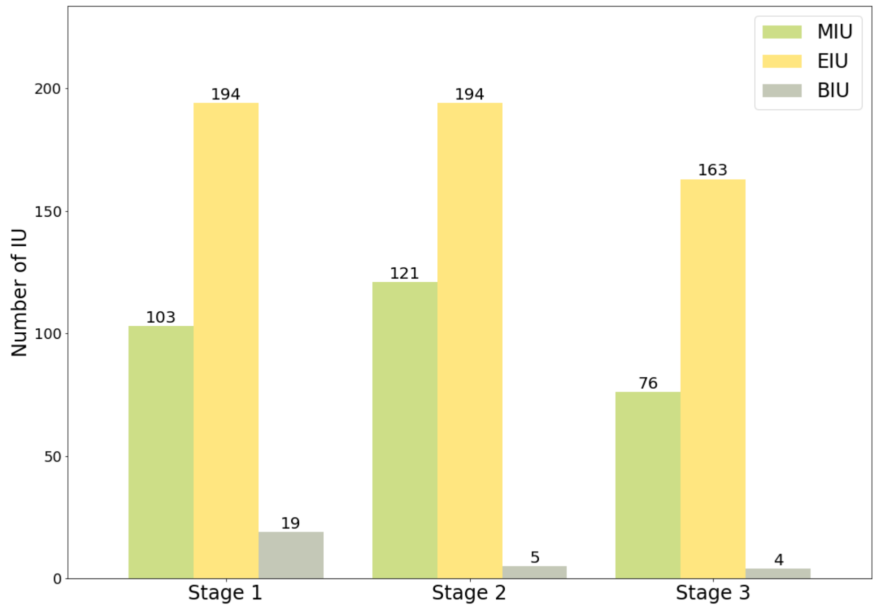


Figure 39 Distribution of MIU, EIU and BIU of all three stages of speaker ***NI06FBP***

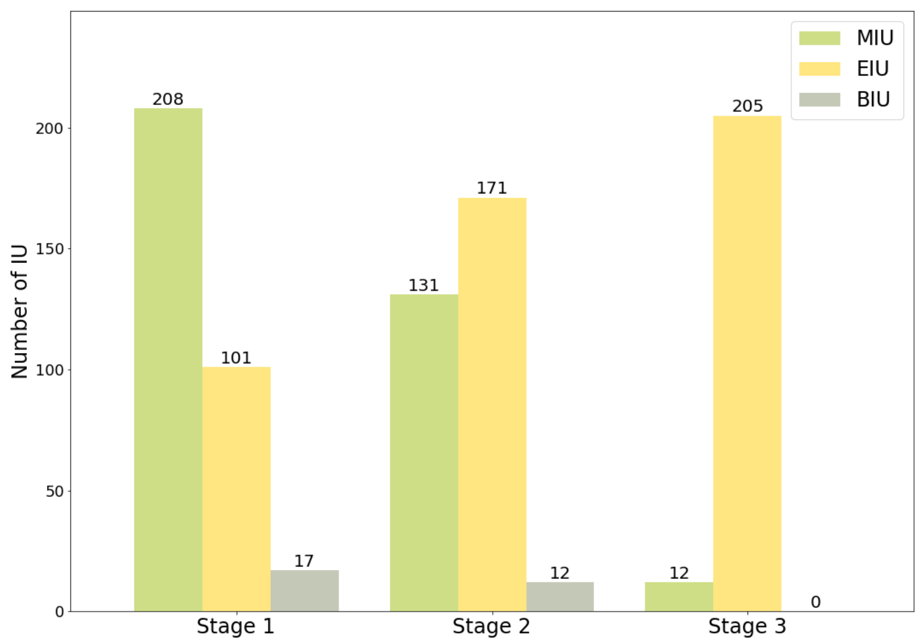


Figure 40 Distribution of MIU, EIU and BIU during all three stages of speaker ***08NC15MBP***

Influenced by either the topic change, or the language usage of the interlocutor, or any other possible reasons, this speaker gradually changes from using Mandarin more to basically speaking only English. This natural change is truly reflected by the IU distribution of the three stages. In order to make more general and accurate observation of the language usage, we expect as much CS as possible in the discourse. To be more specific, a situation like Stage 3 of speaker ***08NC15MBP*** might induce interference to the analysis. Because 0 BIU indicates that in this stage, there is no INCS. Then, 12 MIU and 205 EIU indicate that there is a big number of monolingual English utterances, so the number of EXCS is extremely small. The BIU percentage in this stage is 0%, yet it is not because this speaker only performs CS at IU boundaries. It is due to the fact that there are simply not many CS. A low BIU percentage like this does not prove the existence of prosodic constraints[20]. With this considered, if one simply extracts a continuous period of audio, rather than spreading out the audio extraction to three stages, it is possible to encounter extreme situations like in Stage 3. Then, this would affect the results. In fact this situation is caused by an individual peculiarity during this period of time, and not the general speaking convention of this speaker. Thus, the extraction method of the current study helps reduce this kind of extreme situation as much as possible (Section 3.2.1).

From the results above, it is observed that under some circumstances, the progress session would affect the language usage (such as MIU and EIU distribution) of some speakers. Previous studies seldom discussed this matter because they usually collected the data themselves and analyzed the whole audio they collected. They were not able to make a selection from a large-scale corpus like in the current study. One major reason is that it is extremely time-consuming to design, collect and transcribe the data, within a limited time length. It is introduced in Section 2.2.3 that, Shenk (2006) conducted her research on a sell-collected database, which is 1h audio of 4 Spanish-English bilinguals. While Shenk was one of the pioneers in a new field, the universality of her conclusion is not quite enough since her study only looked at 4 speakers. Compared to Shenk, the database of the current work comes from 178h of 156 bilinguals. It took a professional team 3 years to collect and transcribe the data. This makes it possible for the current study to make a sufficient and reasonable selection from it. It is introduced in Section 3.2.1 that this study conducts a convergency study on the total 6h audio of 6 randomly selected speakers, and finds that 7min (for conversations) and 5min (for interviews) are able to represent the speaker's language usage. This observation significantly increases the efficiency of data processing. Within a limited time length, analyzing as many speakers as possible can help

[20] More detailed discussion on this matter is conducted in Section 5.1.

increase the universality of the results. Of course, for the sake of cautiousness, the study takes the possible impact of the session progress into consideration, and triples the representative length, introduced in Section 3.2.1. The discussion in the current section shows that this consideration is quite necessary because the session progress has certain impact for some speakers. Thanks to the large-scale SEAME corpus, the current study is able to conduct the convergence study and to consider the impact of the session progress. This helps the results and conclusion of this study to be more universal and more objective.

5.4.2 CS distribution discussion

The current section also makes an attempt to explore the reason behind individual differences in CS distribution. Table 10 illustrates the distribution of INCS and EXCS, sorted in an ascending order of INCS. Generally speaking, the INCS percentages stay at a relatively low level. Speaker ***NI06FBP*** possesses the lowest percentage of INCS, 6.97%; whereas the highest percentage of 18.67% is created by speaker ***05NC10MAY***. It is observed that speakers in interviews seem to create more INCS percentages compared to those in conversations. Malaysian speakers possess bigger INCS percentages compared to Singaporean ones. Other than this, there might exist some influences on the INCS percentage at a more profound level, although due to the limitation on the analyzing ability, those influences cannot be observed from the current study.

SPEAKER	INCS		EXCS		R_{CS}	Total CS	Gender	Nationality	Age	Speech type
	N	%	N	%						
NI06FBP	32	6.97	427	93.03	0.50	459	F	S	20	I
06NC11MAX	33	7.48	408	92.52	0.64	441	M	M	23	C
03NC05FAX	29	8.08	330	91.92	0.32	359	F	M	21	C
04NC07FBX	38	8.39	415	91.61	0.49	453	F	S	20	C
NI20MBP	20	9.35	194	90.65	0.23	214	M	S	22	I
17NC33FBP	39	9.40	376	90.60	0.57	415	F	S	23	C
UI14MAZ	35	9.67	327	90.33	0.42	362	M	M	30	I
08NC15MBP	37	10.03	278	89.97	0.34	315	M	S	21	C
NI37MBP	19	10.92	155	89.08	0.27	174	M	S	22	I
10NC20MBQ	37	11.78	277	88.22	0.49	314	M	S	21	C
UI10FAZ	33	11.87	245	88.13	0.40	278	F	M	27	I
05NC09FAX	51	13.53	326	86.47	0.49	377	F	M	22	C
NI18MBP	62	15.82	330	84.18	0.33	392	M	S	19	I
UI12FAZ	77	17.42	365	82.58	0.42	442	F	M	27	I
UI07FAZ	95	17.66	443	82.34	0.49	538	F	M	30	I
05NC10MAY	98	18.67	427	81.33	0.43	525	M	M	23	C

Table 10 Individual distribution of INCS and EXCS, with personal information included. (In the *Gender* column, *F* = Female, *M* = Male; in the *Nationality* column, *M* = Malaysian, *S* = Singaporean; in the *Speech Type* column, *I* = Interview, *C* = Conversation.)

5.5 Database appropriateness discussion

It is discussed in Section 2.3.3 that, Urrea (2012) followed Shenk's (2006) work and applied the information-based approach to investigate the role played by prosodic constraints in CS behaviors. Shenk calculated different IU percentages and the 4% BIU percentage led her to conclude that bilinguals mainly code switch at IU boundaries. Urrea, on the other hand, calculated different CS percentages, and the 21% INCS percentage led her to conclude that most CS happen at IU boundaries, she thus supports Shenk's conclusion. Urrea then claimed that Shenk's result is community specific, because the 4% is much lower than her 21%. Section 2.3.3 pointed out two reasons why this comparison might be problematic. The first reason is that they applied different calculating methods on different objects, thus their results are bound to be essentially different. In fact, results presented in Section 5.1 and Section 5.3 are able to directly explain this: even for the same database, IU distribution and CS distribution are still very different. These are two distinguished kinds of results obtained from two different calculating methods, a direct comparison between them offers little useful information. The second reason why this comparison might be problematic concerns database appropriateness.

Section 2.3.3 points out that for CS studies that apply the information-based approach, there lacks an adequate parameter to help measure the appropriateness of the database. Without providing a parameter, researchers are not eligible to claim that their database is appropriate while others' is *special* (claimed by Urrea). Thus based on the results presented in the previous sections, this section makes an attempt to propose a novel parameter to help measure the appropriateness of a database for this specific analysis.

Section 2.3.3 named two critical requirements that a database should meet to conduct a CS study applying the information-based approach: 1) it needs to be a truly natural bilingual speech; 2) it needs to contain enough CS. The first requirement is straightforward, it is not hard to meet by collecting the data from a truly bilingual community. As for the second requirement, it is discussed in Section 5.1 that, a database needs to contain fewer extreme situations like *Situation B*, and more situations like *Situation A* in Figure 29. In other words, for research focusing on CS issues, a database needs to contain more CS. It is pointed out in Section 5.1 that, an enough amount of EXCS is needed among the monolingual IU (MIU+EIU), the more the better. Thus the parameter in the following is proposed, as the parameter to help measure the number of EXCS among the MIU+EIU:

$$R_{CS} = \frac{EXCS}{\text{MIU} + \text{EIU}}$$

The formula illustrates that using the total number of MIU and EIU to divide the number of EXCS, the value R_{CS} (which represents "the ratio of CS") is obtained. For each individual, the R_{CS} value is calculated with the same formula: use the total number of MIU and EIU of this individual to divide the number of EXCS. For the same amount of MIU+EIU, a bigger R_{CS} value indicates a larger amount of EXCS. For instance, when a speaker produces a great number of monolingual phrases, illustrated in *Situation B* in Figure 29, this makes the number of MIU+EIU large. If the number of EXCS is extremely small, in that case, the R_{CS} value is then very small. At the same time, it is discussed in Section 5.1 that, the speaker with *Situation B* creates a small BIU percentage. For this speaker, even a small BIU percentage cannot lead to any conclusion about the switching point preference, nor about the existence of prosodic constraints. Thus, the data provided by this speaker is considered not to be appropriate enough. The result of BIU percentage needs the support from an adequate R_{CS} value. The following discussion on the R_{CS} of the current database provides a practical example for this parameter.

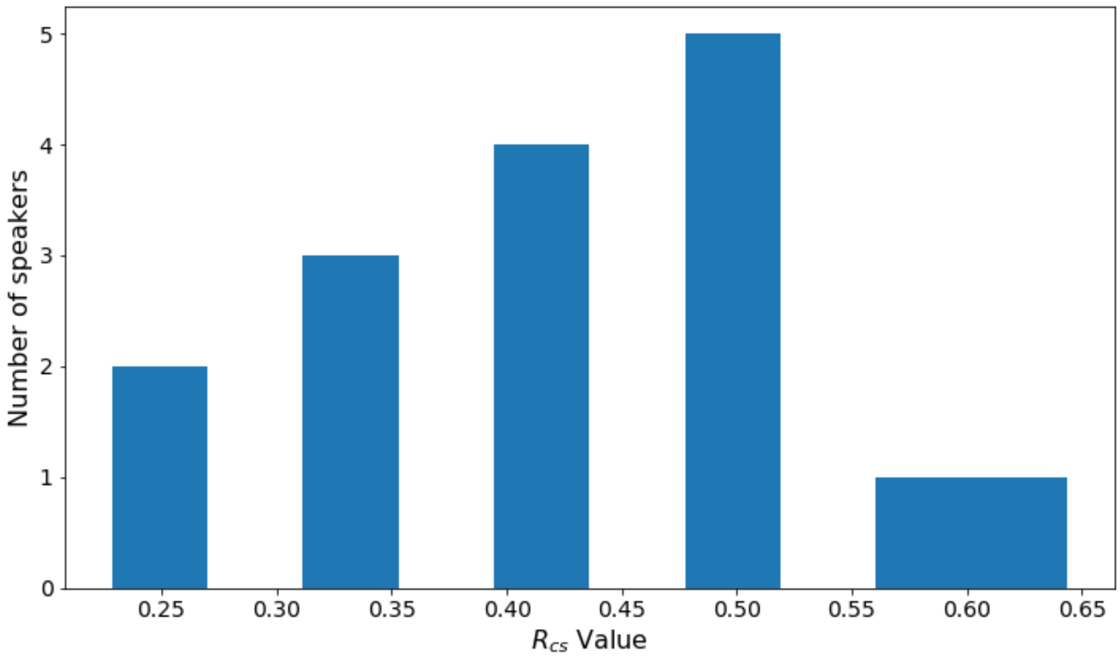


Figure 41 Distribution of R_{CS} value of 16 individuals

For the 16 speakers of the database, a total number of 5323 EXCS is created, and the total number of MIU and EIU combined is 12,659. Thus, the R_{CS} for the current database is 0.42. Figure 41 illustrates the R_{CS} distribution of the 16 individuals. The biggest R_{CS} is 0.64, whereas the smallest R_{CS} is 0.22. The majority of the speakers (14 out of 16) possess an R_{CS} over 0.30.

Since the R_{CS} parameter is proposed in the current research, for the time being the appropriate critical R_{CS} value[21] could only be obtained based on the distribution of the current

[21] Section 4.2.1 points out that critical values are essentially cut-off values that define regions where the test statistic is unlikely to lie. Here the critical R_{CS} value decides whether a database is appropriate: a database with a R_{CS} higher than the critical R_{CS} value is determined to be appropriate, whereas a database with a R_{CS} lower than the critical R_{CS} value is determined to be inappropriate.

study. It is discussed in Section 5.4 that, for some speakers, the session progress has a rather significant impact on the language usage. For example, the IU distribution is influenced. According to the definition of R_{CS} presented above, it correlates highly with the IU distribution, thus the influence of the progress on the R_{CS} value of the individuals is examined.

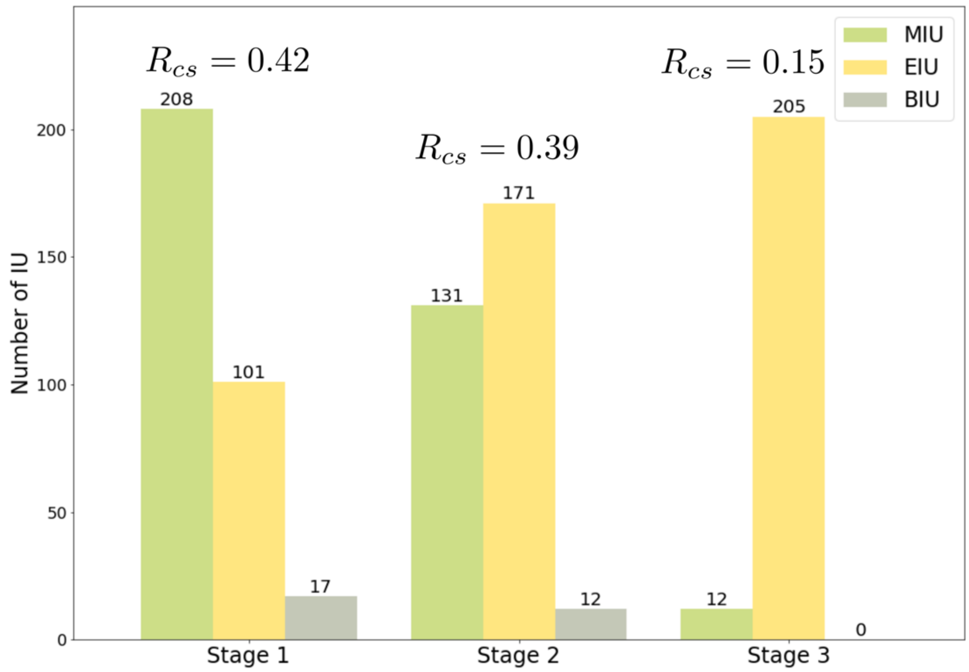


Figure 42 Distribution of MIU and EIU of speaker ***08NC15MBP*** at stages 1, 2 and 3

Figure 42 illustrates the IU distribution of speaker ***08NC15MBP*** at each stage, with the R_{CS} indicated accordingly. R_{CS} values at Stage 1 and stage 2 are relatively high, at respectively 0.40 and 0.32. However, at Stage 3, the number of MIU differs significantly from the number of EIU. This leads to an extremely small number of EXCS, which finally results in an extremely small R_{CS} value of 0.15. The consequence is that the R_{CS} value for all three stages combined together of this speaker is relatively low, which is 0.33. Note that the BIU percentage of Stage 3 is 0%, which seems to be solid evidence of the existence of the prosodic constraint that CS mostly (or only, under this specific circumstance) occurs at IU boundaries. Although since there are only 12 MIU and 205 EIU, the number of EXCS is extremely low at this stage. This is an example of the extreme situation (Situation B) discussed in Section 5.1. Clearly the low BIU percentage of this stage is caused by the huge amount of monolingual IU, and there are not many EXCS, which is reflected by a very low R_{CS} value of 0.15. This further proves that sometimes, solely the BIU percentage isn't enough to illustrate the actual situation of language usage. It is thus necessary to consult the R_{CS} value to make sure that there are enough CS created, and that the database is appropriate.

Situations of speaker ***NI20MBP*** and ***NI37MBP*** are quite similar to speaker ***08NC15MBP***. An extremely low R_{CS} value is observed for one of the stages, and thus causes the total R_{CS} value to be lower than other speakers, at respectively 0.23 and 0.27. They are precisely the three speakers whose IU distribution suffers a significant impact from the session progress, mentioned in Section 5.4. The situation for speaker ***03NC05FAX***, on the other hand, is rather different from the three above mentioned. Although this speaker also possesses a relatively low R_{CS} value of 0.32, it is not caused by one extremely low value at one of the stages. Instead, the R_{CS} values for all three stages maintain at a relatively low level, illustrated in Figure 43.

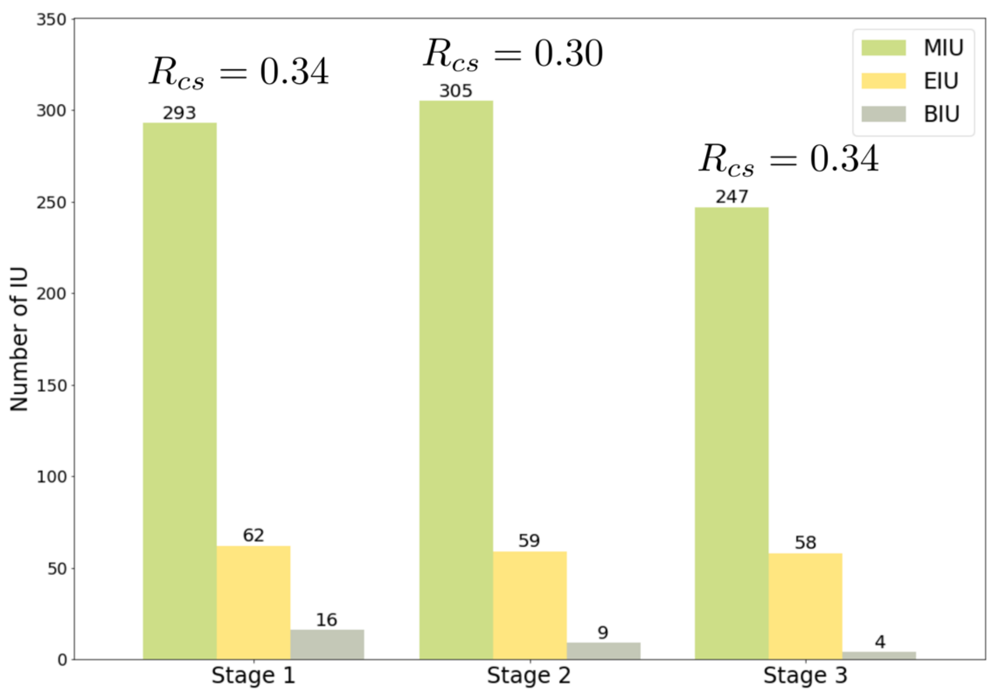


Figure 43 Distribution of MIU and EIU of speaker ***03NC05FAX*** at stages 1, 2 and 3

Other than the four speakers mentioned above, all the other speakers possess a rather stable IU distribution with a reasonable difference between EIU and MIU, which keeps the R_{CS} value in a stable range. For instance, the distribution of MIU and EIU at different stages of speaker ***05NC10MAY*** is presented in Figure 44. This speaker possesses a total R_{CS} value of 0.43, respectively 0.47 at Stage 1, 0.38 at Stage 2 and 0.43 at Stage 3.

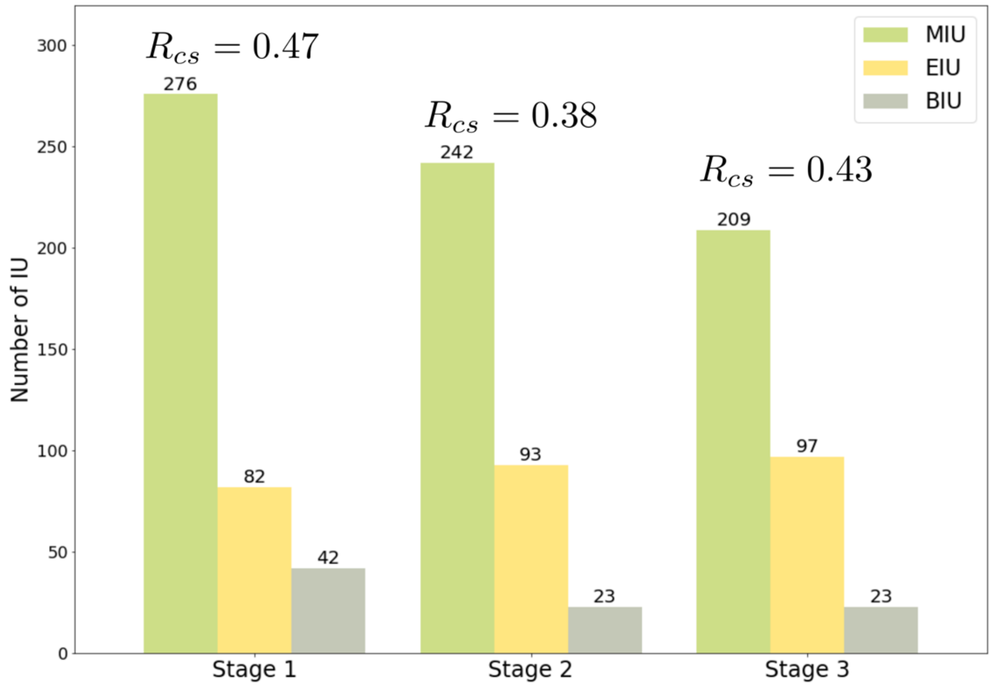


Figure 44 Distribution of MIU and EIU of speaker ***05NC10MAY*** at stages 1, 2 and 3

The difference between speaker ***08NC15MBP*** and speaker ***05NC10MAY*** is quite interesting. It is observed from Figure 42 and Figure 44 that, these two speakers possess very close R_{CS} values at both Stage 1 and Stage 2. But in Stage 3, R_{CS} of speaker ***08NC15MBP*** is 0.15 and R_{CS} of speaker ***05NC10MAY*** is 0.43. This causes the total R_{CS} of the former to be smaller than the latter. Besides, it is observed that the number of MIU and EIU of speaker ***08NC15MBP*** ($MIU = 131, EIU = 171$) is quite different from that of speaker ***05NC10MAY*** ($MIU = 242, EIU = 93$) in Stage 2. Yet, the R_{CS} values are still quite close. Thus, a high R_{CS} value does not rely solely on the definite number of MIU and EIU, it concerns more the number of EXCS in terms of the number of MIU+EIU. However, if the difference between MIU and EIU is too significant (as in Stage 3 of speaker ***08NC15MBP***), a high number of EXCS cannot be obtained. This then results in a low R_{CS} value.

Based on the discussion above, and the fact that $R_{CS} > 0.30$ for the majority of the speakers (illustrated in Figure 41), the critical R_{CS} value is suggested to be 0.30, which indicates an enough amount of EXCS among MIU+EIU. To sum up, the current section proposes the R_{CS} as a parameter to help measure the appropriateness of a database and suggests the critical value of 0.30.

It is pointed out in Section 2.3.3 that, a database could be suitable for all sorts of research, the R_{CS} parameter concerns solely studies that apply the information-based approach, it cannot be applied as a measure of appropriateness for other research. To this day, there is no parameter like this to help measure the database appropriateness, which has caused certain misunderstandings and problematic claims in previous studies (such as Urrea (2012)). This is what encouraged the current study to try and develop a parameter for this specific kind of research. To a certain degree, this parameter reveals the relationship between the number of monolingual IU and CS in the database, which is seldom studied by other researchers. Based on the results of the current study, this parameter reveals how monolingual IU and CS changes with the session progress. Should this parameter be applied in practical research, it could help researchers increase the persuasiveness of their results. In addition, the platforms that sell corpora (such as the Linguistic Data Consortium) could provide the information of the R_{CS} parameter on their databases. This could help researchers make more adequate choices when making a purchase. However, since it is the first time the parameter is developed, and it is based on a single database, its applicability still requires further validation from many future studies on various language pairs in this field. Researchers are encouraged to take this parameter into consideration in their work, in order to find out whether this is able to truly measure different databases. Furthermore, future studies are encouraged to work on optimizing the specific formula of this parameter, to help it provide more accurate information on the databases.

Based on the discussion above, a solution is provided for the key issue proposed in Section 2.3.3. Finally, together with the systematic methodology proposed in Chapter 4, which deals with subjectivity problem mentioned in Section 2.3.2, the third objective of this study is fulfilled.

Chapter 6 Conclusion

The study of CS has been a popular subject among linguistic research over the past few decades during which linguists have been exploring the constraints behind this behaviour. Many syntactic constraints have been proposed during these years, although practically all of them have encountered challenges and counterarguments. Then, to provide a complementary explanation, prosodic constraints were proposed recently, which has attracted a certain amount of attention. However, as a rather minor language pair, studies on Mandarin-English CS were rare until much later due to the limited Mandarin-English bilingual communities. Thus, research on syntactic aspects of this language pair remains rather limited, those that focus on prosodic aspects are few to be found. Three objectives are proposed in the current research to explore both aspects.

The first objective concerns the syntactic patterns. Studies on Mandarin-English language pair have mainly focused on investigating the syntactic patterns at switching points. This study follows previous studies in the field and investigates the patterns. Note that the current database is collected from a bilingual community that has rarely been explored. Thus, other than testing the patterns discovered in previous studies, new patterns have also been found. It is presented in Section 5.2 that, P1 and P2 coincide with previous studies, whereas P3 and P4 are novel and observed in the current study. P4 contains peculiar Mandarin particles. It appears within the IU 100% of the time for all the individuals. Also, for some individuals, it counts for the majority of the BIU. Initial discussion is provided on why these patterns exist. For instance, the Mandarin particles are function words that possess a neutral tone, this characteristic might be the reason why Mandarin-English bilinguals add them at the end of a phrase constantly, out of customs and convenience. Based on the prosodic and syntactic results, the peculiar P4 disrupts the results and should be excluded from the final results and conclusion. A comparison of the results before and after taking out P4 is conducted. The percentage of BIU is reduced from 10.6% to 4.7%. The influence brought by this interference is not significant in the current study, since with or without it, the final conclusion is not affected. It offers a novel way of thinking for future studies: if a future study conducts an analysis from the prosodic perspective and obtains bad results, should it consider the possible existence of the interference from the syntactic perspective before coming to a conclusion? In the current study, the interference is a specific syntactic pattern. Although for other syntactic aspects in different language pairs, the interference could be various. No matter what the specific form, it could cause significant influence for the results. For example, if the result is reduced from around 30% to 10% once the interference is taken out, bad results with no conclusion would become good ones.

The second objective concerns providing evidence of prosodic constraints. With the SEAME Mandarin-English CS database, the current study applies the information-based approach to explore evidence of the control prosodic constraints have over CS behaviour. With IU being the fundamental unit, the approach was first proposed by Chafe (1979, 1980, 1993, 1994) to conduct research on natural monolingual discourses. It was recently introduced to the studies of bilingual CS behaviour by Shenk (2006), and found solid evidence that CS is controlled by prosodic constraints. This has received support from several subsequent studies. A 4.7% of BIU (10.6% before removing P4) is obtained (discussed in Section 5.1-5.3), which led to the conclusion that bilinguals mostly code-switch at IU boundaries, and that CS behaviour is controlled by prosodic constraints. To exclude uncertainty from the results, such as the hypothetical extreme situation discussed in Section 5.1, this study has explored CS distribution in terms of IU boundaries. Based on the database, the low BIU percentage is caused by constant CS at IU boundaries. Furthermore, to strengthen the persuasiveness of this conclusion, the results are generally applicable to all individuals. This result supports Shenk's (2006) conclusion from Mandarin-English language pair, which had not been explored yet in this field, thus fulfilling the second objective.

The third objective is fulfilled by providing solutions to two problems observed in previous methodologies. The first one is the subjectivity problem. It is discussed in Chapter 4 that, with consultation to previous studies, the current study developed a novel systematic methodology to identify, classify and analyze the IU. It is discussed in Section 4.3 that, this methodology has reduced the impact the subjectivity problem has caused in previous studies, which helps this study obtain reliable results. Note that the methodology does not just apply to Mandarin-English CS studies, it should be generally applicable to other language pairs as well, even for multilingual studies. The applicability will require further support from future studies. This methodology is principled enough for future researchers to easily get familiar with and save time on getting experience. In the future, with the collaboration of the advanced research approaches and techniques in computer science such as machine learning, the methodology could be further improved: the impact of the subjectivity problem could be further reduced, and the analysis is going to be much less time-consuming.

The second problem is the lack of a parameter to help assess the appropriateness of a research database for applying the information-based approach. It is introduced in Section 2.3.3 that, in previous studies, certain misunderstandings have occurred due to misinterpretations of the database. This study has considered the appropriateness of the database and how important it is for research. Thus, a principled parameter on the appropriateness of the database is developed, that is the R_{CS} parameter, a critical value of 0.30 is also suggested. This parameter could be referred to by future studies in this field. On one hand, for researchers who plan to obtain a corpus, this parameter could be used to measure database appropriateness. On the other hand, for studies that establish their own database, this parameter can also be applied to help make adjustments to the parameters and on data collecting designs. Then, whether or not the database is for their own research or to provide to others, an R_{CS} can render the database appropriateness more persuasive. However, since the principle and the critical value of this parameter has been proposed based on the current study, its accuracy expects further examination from future studies in the field.

The results and conclusion in this study have offered support for the evidence of prosodic constraints in CS behaviour from a rarely explored language pair. This urges more research to focus on the prosodic perspective in CS studies, in order to accumulate more solid evidence from various language pairs. With the development of theoretical and practical support for these constraints, prosodic constraints could be applied to solve real problems. The scope of mind could be broadened. One cannot rely on syntax alone to explain everything concerning CS behaviour because it is also controlled by prosodic constraints. Therefore, when counterexamples are provided, linguists can search for alternatives from prosodic aspects, rather than declare their study insoluble. Only with these two aspects combined together can we offer a more comprehensive and structural understanding for this bilingual behaviour. This then paves the way for a better understanding of the first and second language acquisition, multilingual studies, language education, etc.

References

Atmaja, B. T., & Akagi, M. (2020). The effect of silence feature in dimensional speech emotion recognition. *Proceedings of the International Conference on Speech Prosody*, *5*, 26–30. https://doi.org/10.21437/SpeechProsody.2020-6

Auer, P., & Di Luzio, A. (1992). *The Contextualization of Language*. *John Benjamins Publishing Company*. https://doi.org/10.2307/415843

Backus, A. M. (1996). *Two in one: bilingual speech of Turkish immigrants in the Netherlands*. Katholieke Universiteit Brabant.

Bentahila, A., & Davies, E. (1983). The syntax of Arabic-French code-switching. *Lingua*, *59*, 301–330. https://doi.org/10.1016/0024-3841(83)90007-4

Berk-Seligson, S. (1986). Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society*, *15*(3), 313–348.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Institute of Phonetic Sciences* (Vol. 17, pp. 97–110).

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer [Computer program]. Version 6.0. 37*.

Bokamba, E. (1988). Code-mixing, language variation, and linguistic theory: Evidence from Bantu languages. *Lingua*, *76*, 21–62. https://doi.org/10.1016/0024-3841(88)90017-4

Bokamba, E. (1989). Are there syntactic constraints on code-mixing? *World Englishes*, *8*(3), 277–292. https://doi.org/10.1111/j.1467-971X.1989.tb00669.x

Brazil, D. (1985). *The Communicative Value of Intonation in English*. cambridge university press.

Bullock, B. E., & Toribio, A. J. (2009). *The Cambridge Handbook of Linguistic Code-switching*. *Journal of Sociolinguistics*. Cambridge University Press. https://doi.org/10.1111/j.1467-9841.2011.00480.x

Cacoullos, R. T., & Travis, C. E. (2010). Variable yo expression in New Mexico: English influence. *Journal of the Rivera-Mills & Villa*, *203*, 185–206. https://doi.org/10.31819/9783865278692-012

Cacoullos, R. T., & Travis, C. E. (2015). Gauging convergence on the ground: Code-switching in the community*. *International Journal of Bilingualism*, *19*(4), 365–386. https://doi.org/10.1177/1367006913516046

Chafe, W. (1979). The flow of thought and the flow of language. In *Discourse and syntax* (pp. 159–181).

Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In *Cognitive, cultural, and linguistic aspects of narrative produc- tion* (pp. 9–50). Norwood, NJ: Ablex.

Chafe, W. (1993). Prosodic and functional units of language. In *Talking data: Transcription and coding in discourse research* (pp. 33–44). Hillsdale, NJ: Lawrence Erlbaum.

Chafe, W. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press.

Chafe, W., & Tannen, D. (1987). The relation between written and spoken language. *Annual Review of Anthropology*, *16*, 383–407.

Chang, N. (1958). Tones and intonation in the Chengtu dialect (Szechuan, China). *Phonetica*, *2*, 391–413.

Chao, Y. (1930). A System of Tone Letters. In *La maître phonétique* (pp. 24–27).

Chao, Y. (1968). *A Grammar of Spoken Chinese*. Berkeley & Los Angeles: University of California Press.

Chen, H.-C., & Tzeng, O. J. (1992). *Language processing in Chinese*. Elsevier Science.

Clyne, M. G. (1991). *Community languages : the Australian experience*.

Cook-Gumperz, J., & Gumperz, J. (1976). *Papers on Language and Context*. Language Behavior Research Laboratory, University of California.

Couper-Kuhlen, E. (1986). *An Introduction to English Prosody*. Edward Arnold.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. London: Cambridge University Press.

Crystal, D. (1975). *The English Rone of Voice: Essays in Intonation, Prosody and Paralanguage*. Hodder Arnold.

De Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research*, *36*, 254–266.

Di Sciullo, A.-M., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, *22*(1), 1–24.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research* (pp. 45–89).

Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics*, *40*, 152–161.

Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics*, *45*, 106–113.

Garellek, M. (2015). Perception of glottalization and phrase-final creak. *J. Acoust. Soc. Am*, *137*, 822–831.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Mazatec. *JIPA*, *41*, 185–205.

Gerfen, C., & Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics*, *33*, 311–334.

Green, D. (1998a). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–81.

Green, D. (1998b). Schemas, tags and inhibition: Reply to commentators. *Bilingualism: Language and Cognition*, *1*, 100–104.

Green, D. (2007). Control, activation and resource: A framework and a model for the control of speech in bilinguals. *The Bilingualism Reader*, 371–383.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149.

Gumperz, J. (1977). The sociolinguistic significance of conversational code-switching. *RELC Journal*, *8*(2), 1–34. https://doi.org/10.1177/003368827700800201

Gumperz, J. (1982). *Discourse Strategies*. Cambridge University Press.

Halliday, M. (1967). *Intonation and Grammar in British English*. *The Hague: Mouton de Gruyter*. https://doi.org/10.1515/9783111357447

Halliday, M., & Matthiessen, C. M. I. M. (2013). *An Introduction to Functional Grammar*. *Routledge*.

Heller, M. (1988). *Codeswitching: Anthropological and sociolinguistic perspectives*. Berlin: Mouton de Gruyter.

Hirschberg, J., & Ward, G. (1992). The influence of pitch range, duration, amplitude and spectral features on the inerpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics*, *20*, 241–251.

Huang, C. J., Li, A., & Simpson, A. (2014). *The handbook of Chinese linguistics*. Wiley BLackwell.

Ishi, C. T., Ken-Ichi Sakakibara, H., Hiroshi, I., & Hagita, N. (2008). A method for automatic detection of vocal fry. *IEEE Transactions on Audio, Speech and Language Processing*, *16*(1), 47–56.

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001

Kachru, B. B. (1978). Toward structuring code-mixing: An Indian perspective. *International Journal of the Sociology of Language*, *16*, 27–46.

Keating, P. A., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In *ICPhS*.

Labov, W. (1971). The notion of 'system'in creole studies. *Pidginization and Creolization of Languages*, 447–472.

Lance, D. M. (1975). Spanish-English code-switching. *El Lenguaje de Los Chicanos*, 138–153.

Laver, J. (1980). The phonetic description of voice quality. *Cambridge Studies in Linguistics London*, *31*, 1–186.

Leong, N. (2014). A study of attitudes towards the Speak Mandarin Campaign in Singapore. *Intercultural Communication Studies*, *3*, 53–65. https://doi.org/10.1080/14664208.2017.1391978

Lin, H. (2001). *A grammar of Mandarin Chinese*. *Lincom Europa*. Lincom Europa. https://doi.org/10.1075/z.197

Lu, J. (1991). Code-switching between Mandarin and English. *World Englishes*, *10*(2), 139–151. https://doi.org/10.1111/j.1467-971X.1991.tb00147.x

Lyu, D., Tan, T., Chng, E., & Li, H. (2015). Mandarin–English code-switching speech corpus in South-East Asia: SEAME. *Language Resources and Evaluation*, *49*, 581–600. https://doi.org/10.1007/s10579-015-9303-x

MacSwan, & Colina, S. (2014). Some consequences of language design: Codeswitching and the PF interface. In *Grammatical theory and bilingual codeswitching* (pp. 185–210). Cambridge, MA: MIT Press.

MacSwan, J. (1999). *A minimalist approach to intrasentential code switching*. New York: Garland.

Manfredi, S., Simeone-Senelle, M.-C., & Tosco, M. (2015). Language contact, borrowing and codeswitching.

Mettouchi, A. (2008). Case-marking, syntactic domains and information structure in Kabyle (Berber). *Typological Studies in Language*, *75*(7).

Miller, A. (2007). Guttural vowels and guttural coarticulation in Ju|'hoansi. *Journal of Phonetics*, *35*, 56–84.

Murphy, P. J., & Akande, O. O. (2005). Cepstrum-based estimation of the harmonics-to-noise ratio for synthesized and human voice signals. In *International conference on nonlinear analyses and algorithms for speech processing* (pp. 150–160).

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon.

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press on Demand.

Myers-Scotton, C. (2006). Multiple voices: An introduction to bilingualism. *Wiley-Blackwell*. https://doi.org/10.1111/1467-8489.00221

Myslin, M., & Levy, R. (2015). Code-switching and Predictability of Meaning in Discourse. *Language*, *91*(4), 871–905.

Nespor, M., & Vogel, I. (1983). Prosodic structure above the word. In *Prosody: Models and measurements* (pp. 123–140). Springer, Berlin, Heidelberg.

Ong, K. K. W., & Zhang, L. J. (2010). Metalinguistic filters within the bilingual language faculty: A study of young English-Chinese bilinguals. *Journal of Psycholinguistic Research*, *39*(3), 243–272. https://doi.org/10.1007/s10936-009-9137-z

Pan, H. H., Chen, M. H., & Lyu, S. R. (2011). Electroglottograph and acoustic cues for phonation contrasts in Taiwan Min falling tones. In *Twelfth Annual Conference of the International Speech Communication Association.*

Park, H.-S. (2006). Structural characteristics of proper nouns in Korean-Swedish discourse. *International Journal of Bilingualism*, *10*(1), 17–36.

Pfaff, C. (1979). Constraints on language mixing: intrasentential code-switching and borrowing in Spanish/English. *Linguistic Society of America*, *55*(2), 291–318.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. MIT.

Poplack, S. (1978). *Syntactic Structure and Social Function of Code-switching*. Centro de Estudios Puertorriqueños,[City University of New York].

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en espanol: toward a typology of code-switching. *Linguistics*, *18*(7–8), 581–618.

Poplack, S., & Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, *24*(3), 279–315. https://doi.org/10.1017/S095439451200018X

Poplack, S., & Meechan, M. (1995). Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. In *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 199–232). New York: Cambridge University Press.

Raitio, T., Kane, J., Drugman, T., & Gobl, C. (2013). HMM-based synthesis of creaky voice. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, (August), 2316–2320.

Sankoff, D., & Poplack, S. (1981). A Formal Grammar for Code-Switching. *Papers in Linguistics: International Journal of Human Communication*, *14*(1), 2–45. https://doi.org/10.1080/08351818109370523

Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in tamil. *Language Variation and Change*, *2*(1), 71–101. https://doi.org/10.1017/S0954394500000272

Schuetze-Coburn, S., Shapley, M., & Weber, E. G. (1991). Units of Intonation in Discourse: A Comparison of Acoustic and Auditory Analyses. *Language and Speech*, *34*(3), 207–234. https://doi.org/10.1177/002383099103400301

Selkirk, E. (1984). *Phonology and Syntax: The Relationship between Sound and Structure*. Cambridge, Mass.

Shen, X. (1990). *The Prosody of Mandarin Chinese*. Berkeley & Los Angeles: University of California Press.

Shenk, P. S. (2006). The interactional and syntactic importance of prosody in Spanish-English bilingual discourse. *International Journal of Bilingualism*, *10*(2), 179–205. https://doi.org/10.1177/13670069060100020401

Singh, R. (1983). We, they and us: a note on code-switching and stratification in North India. *Language in Society*, *12*(1), 71–73.

Sridhar, S., & Sridhar, K. (1980). The syntax and psycholinguistics of bilingual code mixing. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *34*(4), 407–416. https://doi.org/10.1037/h0081105

Tan, P. (1988). A description of patterns of code-mixing and code-switching in a multilingual house-hold. *New Englishes: The Case of Singapore*, 70–99.

Tao, H. (1996). *Units in Mandarin Conversation: Prosody, Discourse and Grammar*. *Language*. John Benjamins Publishing. https://doi.org/10.2307/417887

Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal acoustic analysis – jitter, shimmer and HNR parameters. *Procedia Technology*, *9*, 1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124

Thomason, S. G. (2003). Contact as a source of language change. In *The handbook of historical linguistics* (pp. 687–712). Oxford: Blackwell.

Timm, L. A. (1978). Code-switching in War and Peace. *Aspects of Bilingualism*, 302–315.

Travis, C. E., & Cacoullos, R. T. (2013). Making voices count: Corpus compilation in bilingual communities. *Australian Journal of Linguistics*, *33*(2), 170–194. https://doi.org/10.1080/07268602.2013.814529

Treffers-Daller, J. (2005). Brussels French une fois: Transfer-induced innovation or system-internal development? *Bilingualism: Language and Cognition*, *8*(2), 145–157.

Urrea, D. E. D. (2012). *A community-based study of social, prosodic, and syntactic factors in code-switching*. The Pennsylvania State University.

Van Coetsem, F. (2000). *A general and unified theory of the transmission process in language contact*. Heidelberg: Universitätsverlag Winter.

Vargas, L. (2008). *Divers facteurs dans l'alternance des langues: étude de cas de deux enfants bilingues français - anglais*. Université Montpellier III.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: the pragmatics of fall-rise intonation. *Language*, *61*, 747–776.

Zheng, L. (2005). Grammatical structures in code-switching among second generation Chinese-Australian students. *Monash University Linguistics Papers*, *3*(2), 3–17.

Appendix A Personal information of 16 individuals

Speaker ID	Age	Gender	Nationality	Speech type	Max Pitch	Min Pitch	R_{CS} value
NI18MBP	19	Male	Singaporean	Interview	320	30	0.37
03NC05FAX	21	Female	Malaysian	Conversation	500	90	0.25
UI12FAZ	27	Female	Malaysian	Interview	600	110	0.37
05NC10MAY	23	Male	Malaysian	Conversation	360	45	0.37
NI20MBP	22	Male	Singaporean	Interview	280	60	0.23
NI37MBP	22	Male	Singaporean	Interview	400	70	0.26
04NC07FBX	20	Female	Singaporean	Conversation	600	110	0.50
05NC09FAX	22	Female	Malaysian	Conversation	550	100	0.44
UI07FAZ	30	Female	Malaysian	Interview	550	100	0.48
08NC15MBP	21	Male	Singaporean	Conversation	300	40	0.29
UI10FAZ	27	Female	Malaysian	Interview	400	90	0.39
17NC33FBP	23	Female	Singaporean	Conversation	500	100	0.58
NI06FBP	20	Female	Singaporean	Interview	600	80	0.50
10NC20MBQ	21	Male	Singaporean	Conversation	300	40	0.50
UI14MAZ	30	Male	Malaysian	Interview	300	50	0.41
06NC11MAX	23	Male	Malaysian	Conversation	280	60	0.57

Appendix B Transcription Convention

UNITS

Carriage return : New Intonation Unit

-- : Truncated Intonation Unit

Space character : Word

- : Truncated word

TRANSITIONAL CONTINUITY

. : Discourse final

, : Discourse continuing

? : Discourse appeal

= ***:*** Lengthened segment

PAUSE

…(specific duration) : Long (0.7seconds or longer)

… : Medium (0.3-0.6 seconds)

.. : Short (0.2 seconds or shorter)

VOCAL NOISES

(H) : Audible inhalation

(Hx) : Audible exhalation

@ ***:*** Laughter

% : Glottal stop

(COUGH) : Coughing sound

<@@> ***:*** Laughing quality

(SNORT) : Throat clearing sound

(THROAT) : Throat clearing sound

(GULP) : Gulping sound

(SNIFF) : Sniffing sound