

DOCUMENT RESUME

ED 363 106

FL 021 548

AUTHOR Hok-shing, Brian Chan  
 TITLE Code-Mixing in Hongkong Cantonese-English Bilinguals: Constraints and Processes.  
 PUB DATE Feb 93  
 NOTE 25p.; For the serial publication as a whole, see FL 021 547.  
 PUB TYPE Reports - Research/Technical (143) -- Journal Articles (080)  
 JOURNAL CIT CUHK Papers in Linguistics; n4 pl-24 Feb 1993  
 EDRS PRICE MF01/PC01 Plus Postage.  
 DESCRIPTORS \*Bilingualism; Cantonese; \*Code Switching (Language); English; Language Processing; Language Research; Linguistic Theory; \*Morphology (Languages); \*Syntax; Uncommonly Taught Languages

ABSTRACT

This study focuses on the morphosyntactic aspects of Cantonese-English code-mixing as commonly spoken by bilinguals, most of whom are Cantonese. A corpus of Cantonese-English code-switching collected from informal conversations is analyzed in terms of structural properties, followed by a critique of the major constraints or principles that have been proposed in the literature. It is suggested that none of these constraints or principles are descriptively adequate in view of the data collected, and three alternative constraints--the category equivalence constraint, the bound morpheme constraint, and the specifier constraint are proposed. The possible implications of the constraints on bilingual processing are also discussed. (Author/JL)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

ED 363 106

**Code-mixing in Hongkong Cantonese-English Bilinguals:  
Constraints and Processes<sup>1</sup>**

Brian Chan Hok-shing  
City Polytechnic of Hong Kong

**Abstract**

This study focuses on the morphosyntactic aspects of Cantonese-English code-mixing as commonly spoken by bilinguals most of whom are native speakers of Cantonese. A corpus of Cantonese-English code-mixing collected from informal conversations is analysed in terms of structural properties, followed by a critique of the major constraints or principles which have been proposed in the literature. It is suggested that none of these constraints or principles are descriptively adequate in view of the data collected, and three alternative constraints, namely, the category equivalence constraint, the bound morpheme constraint, and the specifier constraint, are proposed. The possible implications of the constraints on bilingual processing are also discussed.

1. Introduction

By "code-mixing"(CM) I am referring to cases of intrasentential alternation of linguistic elements from two languages, to follow such researchers as Disciullo, Muysken and Singh (1986), Sridhar and Sridhar (1980), Kamwangamalu (1989) and Bokamba (1989), etc.

One of the major issues in the study of code-mixing is whether there are syntactic constraints on code-mixing. The predominant view nowadays is that there are, as suggested by a number of researchers who base their claims on empirical data of different varieties of code-mixing. (i.e. Poplack(1980) on Spanish-English, Kamwangamalu(1989) on Bantu-English/French, Sridhar and Sridhar(1980) on Kannada-English, etc.) Among the constraints which have been proposed, some of them are claimed to be language-universal and applicable to typologically diverse varieties of code-mixing. For instance, the Free Morpheme constraint is reported to be observed in Spanish-English (Sankoff and Poplack 1981), and to a large extent in Arabic-French (Bentahila and Davis 1983) and Spanish-Hebrew (Berk-Seligson 1986) amidst a few violations. The Government constraint is observed in the data of Hindi-English and French-Italian-English code-mixing as collected by Disciullo, Muysken and Singh(1986). The Equivalence constraint, on the other hand, is affirmed in Spanish-English (Sankoff and Poplack 1981) and German/Dutch-English (Clyne 1987).

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

Gladys  
Tang

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)"

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

This document has been reproduced as  
received from the person or organization  
originating it.  
Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OERI position or policy.

BFC/12/548

Nevertheless, others have cast doubts on whether a "constraint-oriented" approach towards the study of code-mixing is justified. The alleged drawbacks of such an approach include the following: First of all, nearly all the universally postulated constraints proposed so far have counterexamples (Bokamba 1989). It leads researchers like Bokamba to suspect that the "constraint-oriented" approach is misguided in the first place. Secondly, in some cases, it is difficult to identify with consistency the grammatical code-mixing patterns as against other ungrammatical ones. Clyne (1987) notes a certain degree of variation of syntax in his corpus of German-English and Dutch-English code-mixing, which may be due to variation of "standard" and "sub-standard" forms and syntactic transfer of English syntax to Dutch and German structures. He also comments that the so-called "ungrammatical" patterns of code-mixing are "nothing more than a tendency" (p.762)

This article focuses on Cantonese-English code-mixing, which has seldom been studied with reference to constraints or other notions of current linguistic theories.<sup>2</sup> This does not mean that Cantonese-English is not subject to any constraints whatsoever. For example, the following sequences with asterisks hardly appear in spontaneous Cantonese-English code-mixing behaviour.

(1)<sup>3</sup>

- a. néih yiu *seriously* tái yī go mahn tàih  
 You MOD seriously look DET CL question  
 You have to seriously look at this question.  
 \* néih yiu yihng-jān-ly tái yī go mahn tàih

- b. gó go *lecturer* séuhng òhng *prepare* dāk m̀h hóu  
 DET CL lecturer attend lessons prepare COMP NEG good  
 That lecturer doesn't prepare well for his lessons.  
 \* *That* góng sī séuhng òhng *prepare* dāk m̀h hóu

Nor does the lack of grammatical studies on Cantonese-English imply that the variety is uninteresting. Quite the contrary, it exhibits many interesting phenomena. Mixing is possible within phrases the structure of which is vastly different in Cantonese and English. (ref. section 3) Even more fascinating are the rich, distinct Cantonese morphological processes which can be applied to English words with English phonological forms retained. For instance, in 2a, the English verb "run" is incorporated into the distinctly Cantonese "A-not-A" structure. In 2b and 2c, the English preposition "for" and "Where's" appear on their own, while in other varieties prepositions and conjunctions rarely alternate in code with other elements in a code-mixed sentence. (ref. Kachru 1978; Joshi 1985)

(2)

- a. go *program* run-m̀h-run dóu a ?  
 CL program A-NOT-A ASP  
 Can the (computer) program run?
- b. yī dī cháan báu haih *for* sī yahn yuhng òuh  
 DET CL product COP for private use  
 These products are for private use.

- c. *Whereas kéuih fun gán gaau...*  
Whereas s/he sleep PROG  
Whereas she is sleeping...

Among these few studies, Gibbons(1987) thinks that Cantonese-English is syntactically governed in some ways (ref. section 4), but he does not refer to those constraints or principles which are postulated to be universal as mentioned above. Leung(1988:37) thinks that Cantonese-English is bound by the Free Morpheme constraint and the Equivalence constraint\*.

However, in this article, I will show that the patterns of Cantonese-English are so diverse that they are not constrained by many of the universal constraints proposed in the literature. Instead, it will be argued that the data suggest a number of alternative constraints. I will also suggest that these constraints provide insights into the psycholinguistic processes involved in code-mixing production, and help to identify the role of L1 and L2 in code-mixing behaviour. Thus, "constraints" are very useful tools in understanding some of the psychological aspects of code-mixing.

## 2. Collection of Data

Before I proceed, let me provide some background information to the variety of code-mixing under study. Code-mixing with Cantonese and English is a common feature in conversations of Hongkong Cantonese-English bilinguals nowadays. Their L1, or mother tongue, is Cantonese, but they acquire English as L2 in schools. There is no evidence that code-mixing is learnt in any manner, but, of course, a certain proficiency level in English is assumed, as code-mixing is recognized as a distinct ability of bilinguals.<sup>4</sup>

The data used in this article contain 500 utterances which are taken from recorded speech and written transcriptions of spontaneous conversations in situations ranging from tutorial discussions, fellowship sharing, informal conversations, TV and radio interviews, etc. The recorded data were taken from radio phone-in programmes, so that a natural setting was ensured without the speakers knowing the recording process. The inclusion of the transcribed data is also intended to ensure the natural setting, and, in addition, to elicit the production of a wider spectrum of speakers.

## 3. A Linguistic Description of the Cantonese-English Data

A distinction is made between the major patterns, in which an English word is surrounded by a predominantly Cantonese discourse, and the minor patterns, in which more than one English word alternate with Cantonese. Such a distinction, which is based purely on the frequency of the patterns found in my corpus as well as Gibbons' (1987) corpus, is meant to serve descriptive purposes only, without any implications on the linguistic properties of these patterns.

As for the major patterns, I describe them in terms of the word class to which the English words belong. There are four major classes, namely, "noun-mixing", "verb-mixing", "adjective or adverb-mixing" and "preposition or conjunction-mixing", which mirror the N, V, A, P taxonomy in the generative framework. In the generative framework, N, V, A, P are postulated to be the four major categories in languages (Radford 1988). The minor patterns refer to the fragments which contain more than one English words.

The following statistics show the number of utterances which fall into the sub-groups of "noun-mixing", "verb-mixing", "adjective or adverb-mixing", "preposition or conjunction-mixing" and "fragment", and the percentage of the utterances these sub-classes contain out of the total of 500 utterances.<sup>5</sup>

	<u>No. of utterances</u>	<u>Percentage</u>
1. Noun-mixing	260	52.0%
2. Verb-mixing	148	29.6%
3. Adjective or Adverb-mixing	84	16.8%
4. Preposition or conjunction-mixing	11	2.2%
5. Fragment	127	25.4%

### 3.1 The major patterns

#### 3.1.1 Verb-mixing

By "verb-mixing" I am referring to the pattern of Cantonese-English code-mixing in which an English verb is surrounded by a predominantly Cantonese discourse. The distribution of the English verb is, as can be observed from the examples below, in positions where a Cantonese verb is distributed; that is, after the subject NP and before the object NP in declarative sentences.

Some of the examples of verb-mixing are cited below:

(3)

a. néih deih hó yíh *ignore* kéuih  
You PL MOD ignore him  
You can ignore him.

b. kéuih deih *plan* jō yī go syú ga heui ōu jōu léuhk hàhng  
They PL plan ASP DET CL summer vacation go Europe travel  
They have planned to go traveling to Europe this summer vacation.

c. go *program* run-mh-run dóu a  
CL program A-NOT-A ASP Q  
Can the (computer) program run?

- d. *gó go taai tái sèhng yaht séung show off*  
 DET CL lady always want show off.  
 That lady always want to show off.

In 1a, a single English verb "ignore" is distributed after the Cantonese subject NP "*néih deih*(you)" and before the Cantonese object NP "*kéuih*(him)". In 1b, the English verb "plan" is followed by a Cantonese perfective aspect marker "*jó*". In 1c, the English verb "run" is incorporated into a distinctly Cantonese morphological structure which appears in questions, "A-not-A". In 1d, a complex verb, "show off", is mixed with Cantonese.

There are some characteristics of the verb-mixing data which deserve attention. Firstly, in most cases the root form of the English verb appears. Tense and agreement are not marked by overt morphological markers, contrary to English syntax but conforming to Cantonese syntax. An even more obvious case in which the English verb is adapted to Cantonese syntax would be its appearance in such distinctly Cantonese structures as A-not-A and reduplication.

### 3.1.2 Noun-mixing

By "noun-mixing" I am referring to the pattern of code-mixing in which an English noun is surrounded by a predominantly Cantonese discourse. The distribution of an English noun is in positions where the Cantonese nouns are distributed in Cantonese; that is, at the head of an NP after a Cantonese determiner and/or a classifier, if any.

Some examples of noun-mixing are cited below:

- (4)
- a. *gó go lecturer séuhng òhng prepare dāk m̀h hóu*  
 DET CL lecturer attend lessons prepare COMP NEG good  
 That lecturer doesn't prepare well for his lessons.
- b. *kéuih deih heui jó mahk jing òuh sihk lunch*  
 They PL go ASP MacDonalds eat lunch  
 They have gone to MacDonalds for lunch.
- c. *William chàhm yaht máaih ge walkman haih sán model*  
 William yesterday buy RCL walkman COP new model  
 marker  
 The walkman which William bought yesterday is a new model.

In 2a, the English noun "lecturer" appears in the subject position, and it is marked by the Cantonese determiner and classifier "*gó go*". In 2b, the English noun "lunch" appears in the object position after the Cantonese verb "*sihk*(eat)". In 2c, the English noun "walkman" is modified by a prenominal relative clause.

From the data, one can easily observe signs of the English noun being adapted to Cantonese syntax. Firstly, sequences of CL(Cantonese) N(English) are common, which in fact violates English syntax (as English does not possess a rule as NP-> CL N). Secondly, if a relative clause is to appear, it is found to be prenominal rather than postnominal, an order which conforms to Cantonese syntax but violates English syntax.

### 3.1.3 Adjective or adverb-mixing

By "adjective or adverb-mixing" I am referring to the pattern of code-mixing in which an English adjective or adverb is surrounded by a predominantly Cantonese discourse. The distribution of an English adjective or adverb is in positions where the Cantonese adjectives and adverbs can also be distributed in Cantonese. For the English adjectives, they either occur at predicative positions after the grammatical subject or the attributive position before the head noun and after the determiner.

Let's now turn to some examples of adjective-mixing.

(5)

a. kéuih jòuh yéh hóu *serious*  
 He do things EMP serious.  
 He is very serious to his work.

b. kéuih haih yāt go hóu *critical* ge yāhn  
 He COP NUM CL EMP critical ADJ person  
 marker

He is a very critical person.

c. néih tīng yaht *free-mh-free* a  
 you tomorrow free(A-not-A) Q  
 Are you free tomorrow?

While an English adjective may appear in the predicative position (i.e. after subject NP within a predicate; e.g. 3a) or the attributive position (i.e. before head noun within an NP; e.g. 3b), it may take on the distinctly Cantonese structure of A-not-A (e.g. 3c) when distributed predicatively.

The following are some examples of adverb-mixing:

d. néih yiv *seriously* tái yī go mahn làih  
 You MOD seriously look DET CL question  
 You have to seriously look at this question.

e. *Honestly*, ngóh gok dāk kéuih ge léuih pāng yóuh  
 Honestly I feel ASP he GEN girlfriend  
 haih hóu ngok ge yāhn  
 COP EMP unkind ADJ person  
 marker

Honestly, I feel his girlfriend is a very unkind person.

- f. kéuih hóu *sensible* gám wah...  
 s/he EMP sensible ADV say  
 marker  
 He/She says very sensibly that...

For the "ly-adverb", it may appear before the VP (e.g. 4d), or before the clause (e.g. 4e). In 4f, the adverb is formed by an English adjective "sensible" and a Cantonese adverb marker, "gám", and it also occurs at the preverbal position.

### 3.1.4 Preposition or conjunction-mixing

By "preposition-mixing" I am referring to the pattern of code-mixing in which an English preposition is surrounded by a predominantly Cantonese discourse. The distribution of an English preposition is prenominal, whereas in Cantonese there are both prepositions and postpositions. For instance,

- (6)  
 a. gaau yuhk hohk yún haih *under* gaau yuhk sī chyúh  
 Education College COP under Education Department  
 The Education colleges are under the Education Department.  
 b. *After* yī go review jī houh...  
 After DET CL review P  
 After this review...

In both 6a and 6b, the English prepositions "under" and "After" are distributed prenominally. A special thing about these patterns is that unlike other patterns the Cantonese equivalents of these English prepositions are not distributed in the same way. The Cantonese equivalents are, in these contexts, postnominal. (i.e. "hah"(under) and "houh"(After)) In 6b, the postnominal "houh" co-occurs with "After". This may well indicate that the English prepositional rule is actively accessed in the production of such code-mixing patterns.

By "conjunction-mixing" I am referring to cases in which an English conjunction is surrounded by a predominantly Cantonese discourse. The distribution of these conjunctions is compatible with that of equivalent Cantonese conjunctions, that is, clause-initial positions.

- c. *Whereas* kéuih fun gán gaau...  
 CONJ he/she sleep PROG  
 Whereas he/she is sleeping...  
 d. néih sīn jòuh vùhn yī yeuhng, *and then* ngóh wúih béi daap ngon néih  
 you first do PRT this CONJ I MOD give answer you  
 You complete this (exercise) first, and then I give  
 you the answer.



### 3.2 The minor patterns

Let's look at the minor patterns, that is, the fragments in which more than one English word alternate with Cantonese elements in a sentence.

In describing the fragments, Gibbons (1987:59) observes that "in the great majority of cases where the fragment of English consists of two words or more, it retains English grammar internally, while not disrupting the surrounding Cantonese grammar".

Two points can be made from Gibbons' observation. Firstly, the internal distribution of English elements must be in conformity with the grammar of English. Secondly, in terms of external distribution, the fragment of English is distributed in places where the surrounding Cantonese syntax is not disrupted. The overall sentence structure of the code-mixed utterance thus conforms to Cantonese syntax. These two points specify the structure of fragments.

Besides, Gibbons also made the observation that "where two or more English words are introduced, one will be innermost in a phrase, while the other will be next innermost". (Gibbons 1987:62) Gibbons does not define what "innermost parts" are formally. Judging from the examples he gives, "the most innermost parts" are referred to as heads of a phrase. The "next innermost parts" are referred to as modifiers which are immediately adjacent to the corresponding heads, and those modifiers further adjacent to these "next innermost parts" would be "the *further next* innermost" parts. Gibbons(1987:62) further specifies that "intrusion of code A into base code B takes place at the innermost parts of the syntactic structure".

In other words, by Gibbon's observation on Cantonese-English, the head of a phrase may appear in English on its own. This is because the head itself is the most innermost constituent anyway. The modifier, however, cannot appear in English without its head and, if any, other modifiers which are closer to the head also appearing in English. This is so because the head and these modifiers would be the more "innermost" constituents.

In the light of these, a fragment consists of a head and its modifiers. This point specifies the form of the fragments.

The above specifications of fragments in Cantonese-English work well for my data. Below are some examples of these fragments:

- (7)  
NP(Eng.)  
a. kéuih go jái jing yāt haih *naughty boy*  
PRON CL son ADV COP *naughty boy*  
Her son is really a naughty boy.

VP(Eng.)

b. hóu sām kéuih *keep his mouth shut*  
 good heart he keep his mouth shut  
 I hope he keeps his mouth shut.

ADJP(Eng.)

c. gó dī gaau sī haih *properly train(ed)*  
 DEM CL teacher COP properly trained  
 Those teachers are properly trained.

PP(Eng.)

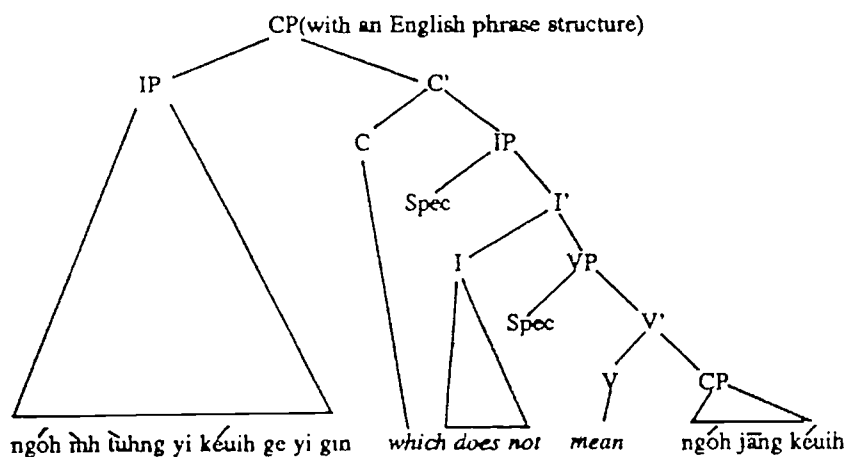
d. m̄h jī dī jē wúih-m̄h-wúih *out of stock* ge nē  
 NEG know CL umbrella MOD(A-not-A) out of stock PRT  
 I don't know if the umbrellas are out of stock.

In other cases in which the fragments are not projections of a major word category, it seems plausible that such sentences are generated by phrase structure rules from the English grammar, accompanied by a process in which the Cantonese phrases are inserted into places where an English phrase of an equivalent category is distributed. Below is an example of these cases:

(8)

a. ngóh m̄h t̄hng yi kéuih ge yi gin, *which does not mean* ngóh jāng kéuih  
 I NEG agree he GEN opinion I hate hi-m  
 I do not agree to his opinions, which does not mean I hate him.

Evidence of the claim that these sentences are generated from the English grammar can be found in the realization of I, ("Inflection" in the X-bar framework) in 11a (i.e. "does"), and the postmodifying relative clause, which are absent in Cantonese. Besides, the absence of verb "to be" and wh-elements in other intrasentential code-mixing patterns, which, as discussed, result from an English word/phrase occupying a place where a corresponding Cantonese category is distributed, also suggests that the cases in question stem from the English grammar.<sup>6</sup>



#### 4. Critique of the major constraints and principles

In this section, a number of constraints and principles which have been proposed in the literature are evaluated with reference to my data of Cantonese-English. Among them are "language-universal" ones, which are based on studies of different varieties of code-mixing, and "language-specific" ones, which are based on Gibbons' observations on his Cantoese-English data (Gibbons:1987).

##### 4.1 Language-universal constraints

###### 4.1.1 The Free Morpheme constraint

The Free Morpheme constraint stipulates that mixing is not allowed between a free morpheme and a bound morpheme. (ref. Poplack 1980:585-586; Sankoff and Poplack 1981:5)

e.g. (Spanish-English) \*run-eando  
PRCG

The Free Morpheme constraint fails to cover the following Cantonese-English patterns:

(9) Violations (Cantonese-English):

a. verb-mixing V(Eng.) ASP(Cant.)  
kéuih deih *plan* jǒ yī go syú ga heui ʔu jǒu léuhk hǎhng  
They PL plan ASP DET CL summer vacation go Europe travel  
They have planned to go traveling to Europe this summer vacation.

b. adverb-mixing: ADJ(Eng.) ADV marker(Cant.)  
kéuih hóu *sensible* gám wah...  
s/he EMP sensible ADV say  
marker  
He/She says very sensibly that...

c. "A-not-A": ADJ(Eng.) NEG(Cant.) ADJ(Eng.)  
néih ũng yaht *free-mh-free* a  
you tomorrow free (A-not-A) Q  
Are you free tomorrow?

In 9a, an English verb, a free morpheme, is inflected by a Cantonese aspect marker, a bound morpheme. In 9b, an English adjective, a free morpheme, is followed by a Cantonese adverb marker, which is a bound morpheme. In 9c, an English adjective, a free morpheme, alternates with the Cantonese negative marker, which is a bound morpheme. All these cases obviously violate the Free Morpheme constraint.

#### 4.1.2 The Equivalence constraint

The Equivalence constraint stipulates that mixing is not allowed between sentence constituents the order of which is different in the two languages concerned. (ref. Poplack 1980:586; Sankoff and Poplack 1981:5-6) As shown below, the order of an adjective is different in a Spanish noun phrase (i.e. postnominal) and that in an English noun phrase (i.e. prenominal) Thus, by the Equivalence constraint, there is no mixing or code-alternation between a noun and an adjective in Spanish-English.

e.g. (Spanish) NP-> DET N ADJ  
(English) NP-> DET ADJ N

By the Equivalence constraint, code-mixing can take place between sentence constituents the order of which is the same in the two languages concerned. This also implies the sentence constituents under comparison are of the same categories. The presence of equivalent categories as a condition for the observance of the Equivalence constraint is explicitly stated in Sankoff and Poplack (1981:5-6).

The Equivalence constraint fails to cover the following patterns in Cantonese-English code-mixing.

(10)

a/ [RCL(Cantonese)N(English)]NP

*William chahm yaht maaih ge walkman haih san model*  
William yesterday buy RCL walkman COP new model  
marker

The walkman which William bought yesterday is a new model.

b/ [CL(Cantonese)N(English)]NP

a *Paul gei jo jeung postcard bei ngoh*  
AFFIX Paul send ASP CL postcard to me  
Paul sent a post card to me.

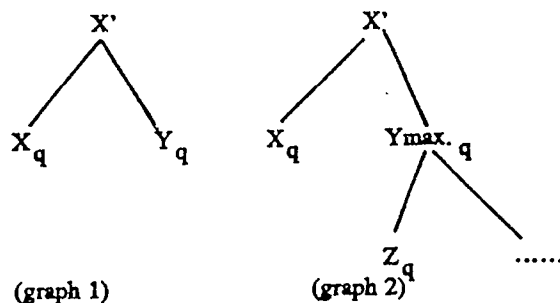
In 10a, the English noun "walkman" is premodified by a relative clause. However, the order of a relative clause in noun phrases is different in English (i.e. postnominal) and Cantonese (i.e. prenominal) By the Equivalence constraint, there should be no mixing between a relative clause and a head noun in Cantonese-English, but there is in reality.

#### 4.1.3 The Government constraint

The Government constraint stipulates that code-mixing is not allowed between elements bearing the same language index, the assignment of which is based on government relations, with "government" defined as:

" X governs Y if the first node dominating X also dominates Y, where X is a major category N, V, A, P and no maximal boundary intervenes between X and Y." (ref. Disciullo, Muysken and Singh 1986:6)

By the Government constraint, if X governs Y, X and Y should receive the same index "q" and there is no mixing between them (ref. graph 1). In case Y is a maximal projection which in turn dominates more than one lexical element, the "highest" element which assymmetrically c-commands the other lexical elements should receive the same index as X. (i.e. "q" ;ref. Disciullo, Muysken and Singh 1986:6) Therefore, there is no mixing between that "highest" element and X (ref. graph 2).



In many examples of verb-mixing, a single English verb may appear without the "highest" element of the noun phrase it governs also appearing in English. Similarly, in patterns of preposition-mixing, a single English preposition may appear without any element of the noun phrase it governs also appearing in English. These cases violate the Government constraint. (ref. 3.1.1, 3.1.4)

#### 4.1.4 The Matrix Code Principle

The Matrix Code principle stipulates that the mixed elements must conform to the morphosyntactic structure of the matrix code (i.e. the native language, or L1) irrespective of any possible violations of that of the embedded code. (i.e. the foreign language, or L2) (ref. Kamwangamalu 1989)

It is a descriptively powerful principle, covering most patterns whereby the distribution of the English words or fragments is compatible with that of the Cantonese words or fragments of the same category. (i.e. except preposition-mixing cases and sentences bearing an English structure)

Besides, it describes a number of phenomena in which an English word is incorporated to the distinct morphosyntactic structure of Cantonese, such as an English verb or adjective which undergoes "A-not-A" structure or reduplication, or an English noun preceded by the Cantonese classifier, or an English verb inflected by a Cantonese aspect marker.

However, the Matrix code principle is not free from weaknesses. It does not really specify the distribution of the mixed elements. Besides, as the principle only states that the morphosyntax of L1 (in this case, Cantonese) is preserved, it cannot describe the fact that the internal constituency of the English fragments conforms to the English phrase structure rather than the Cantonese one. For instance, in 11a, the English preposition phrase "out of stock", contains a complement which is another preposition phrase "of stock". Such a sequence of [P [P NP] pp]pp hardly appears in Cantonese.

In addition, it erroneously rules out cases in which a L2 phrase structure is introduced, (e.g. 11b; ref. 3.2) as it rules that only the morphosyntax of L1 (the matrix code) is preserved.

(11)

a/ Fragments:

m̄h jī dī jē wúih-m̄h-wúih out of stock ge nē?  
 NEG know CL umbrella MOD(A-not-A) out of stock PRT  
 I don't know if the umbrellas are out of stock.

b/ Cases with English phrase structure

ngóh m̄h t̄hng yī kéuih ge yī gin, which does not mean ngóh jāng kéuih  
 I NEG agree he GEN opinion I hate him  
 I do not agree to his opinions, which does not mean I hate him.

#### 4.1.5 The Dual Structure Principle

The Dual Structure Principle stipulates that "the internal structure of the guest constituent need not conform to the constituent structure rules of the host language, so long as its placement in the host language obeys the rule of the host language. This can be clearly illustrated by the following example of Kannada-English. The Dual Structure principle states that the English fragment, which is an NP, is distributed where the Kannada NP is distributed in an otherwise Kannada sentence. Besides, the principles also predicts that the internal constituency of the English fragment conforms to English grammar rather than Kannada grammar. The English fragment contains a postmodifying phrase "of considerable courage", while in Kannada, according to Sridhar and Sridhar (1980), the modifiers are prenominal.

e.g. (Kannada-English)

avanu abba man of considerable courage.  
 (He is a man of considerable courage)  
 (Sridhar and Sridhar 1980:412)

The Dual Structure Principle predicts that the internal structure of English fragments in Cantonese-English sentences as conforming to the grammar of English. As for its weaknesses, the principle does not address the properties of the single-word cases, the major patterns in Cantonese-English. It does not specify the distribution of the mixed elements; that is, it does not predict the places in a Kannada sentence where English words or phrases are allowed to appear. Furthermore, it states that an element of L2 (i.e. the guest language) is mixed under a phrase structure of L1 (i.e. the host language). However, in some cases, the reverse phenomenon can be found. (e.g. 10a.)

## 4.2 Language-specific constraints (Gibbons:1987)

### 4.2.1 The "closed-class" item constraint

This constraint stipulates that "closed-class" or "structure" words ( such as English determiners, conjunctions and auxiliary verbs) do not appear as single-word in Cantonese-English CM, unless in conjunction with the "open-class" or "content" words.

e.g. *néih hó yih go and tái háh*  
you MOD go and look ASP  
you can go and have a look  
\* *néih hó yih heui and tái háh*  
(Gibbons 1987:61)

However, examples of preposition and conjunction-mixing (ref. 3.4) are obvious violations of this constraint.

### 4.2.2 The "fragment" constraint

Gibbons(1987:59) makes the following observation on fragments: The longer elements of English must be fitted "at the same point as the equivalent elements of Cantonese". So, in the following case, the English fragment, which is an NP, is distributed in the position where a Cantonese NP is also distributed in an otherwise Cantonese sentence. I rename Gibbons's observation the "fragment" constraint.

e.g. *m̀h sái joi wán part-time job àh*  
NEG MOD again.find part-time job Q  
No need to find a part-time job? (Gibbons 1987:59)

Obviously, the constraint does not address the major patterns in Cantonese-English code-mixing, the single word cases.

### 4.2.3 The "innermost" constituent constraint

Another observation made by Gibbons(1987) concerns the relative priority of different constituents that may appear in Cantonese-English. As elaborated in 3.2, Gibbons (1987) observed that in Cantonese-English, the head of a phrase, being the most innermost parts, may appear in English on its own. A modifier cannot appear in English on its own if the head or other modifiers closer to the head, being the "more" innermost parts, do not appear in English.

I rename Gibbons' observation the "innermost" constituent constraint. By the constraint, a constituent may appear in English in Cantonese-English if it is the head. If it is a modifier, the corresponding head and, if any, the modifiers closer to the head must also appear in English. Accordingly, the following sequence with asterisk is ruled out as "the", being a modifier, cannot appear in English without the head "*yuhn yāng*(reason)" appearing in English.

kéuih déih giu ngóh *explain the reason*  
 They PL ask me explain the reason  
 They ask me to explain the reason.  
 \*kéuih deih giu ngóh *explain the yùhn yāng.*  
 (Gibbons 1987:62)

In terms of the most recent X-bar theory on phrase structure, what was generally known as the "modifier" in traditional grammar falls into three kinds, namely, the complements, which expands a constituent from zero bar to single bar; the adjuncts, which expand a constituent from zero/single/double bar to the same bar level; and the specifiers, which expands a constituent from single bar to double bar. Let's now refer to the X-bar framework to examine the innermost constituent constraint (ref. Radford 1988).

In the light of the X-bar framework, however, it is found that the following examples violate the innermost constituent constraint:

- (12)
- a. néih yiu *seriously* tái yī go mahn t'aih  
 You MOD seriously look DET CL question  
 You have to seriously look at this question.
- b. kéuih haih yāt go hóu *critical* ge y'ahn  
 He COP NUM CL EMP critical ADJ person  
 marker  
 He is a very critical person.
- c. kéuih go jái jing yāt haih *naughty boy*  
 PRON CL son ADV COP naughty boy  
 Her son is really a naughty boy.

In 12a, the adverb "seriously", an adjunct, appears in English without the head verb "t'ai(look at)" also appearing in English. In 12b, the attributive adjective "critical", a complement, appears in English without the head noun "y'ahn(man)" also appearing in English. In 12c, the noun phrase "naughty boy", a verb phrase complement, appears in English but the head verb "haih(COP)" does not.



## 5. Revised Constraints

Considering the inadequacies of the constraints or principles when applied to the Cantonese-English code-mixing data, I propose a number of alternative constraints in this section.

### 5.1 The category-equivalence constraint

Two generalizations can be drawn from the patterns of Cantonese-English code-mixing we have looked at. Firstly, for the single-word cases (except some preposition-mixing cases<sup>7</sup>) and the fragments, the elements of English are embedded in positions where the corresponding Cantonese categories are distributed. Secondly, for the code-mixed sentences and the preposition-mixing cases with an English phrase structure, elements of Cantonese are embedded in positions where the corresponding English categories are distributed.

To cover these two generalizations, the following constraint is formulated:

#### (13) The category equivalence constraint

"In code-mixing, an element from the embedded code is distributed in a position where an element of the same category from the matrix code is distributed in the matrix code. The matrix code is the language from which the sentence structure of the code-mixed sentence is derived, and the embedded code is the language from which the mixed elements are derived. Such an element range from a morpheme to a phrase."

Here, I must say there have been many researchers who also view code-mixing as involving a sentence constituent being substituted by a constituent of the same category from another language. For instance, the Dual Structure principle proposed by Sridhar and Sridhar (1980) stipulates that in code-mixing a fragment of the guest language be inserted into a place where the rule (i.e. sentence structure) of the host language is not violated. That is possible only when the guest constituent is distributed in a place where a host constituent of the same category is distributed. In his parsing model for code-mixing, Joshi (1985) explicitly states a switch rule which allows all categories except closed class words to switch from the matrix language to the embedded language. In Nishimura's study (1986) of Japanese-English, the switch is not unidirectional, and constituents from either Japanese or English may be switched to constituents of the same category in another language. In their study of Singaporean Chinese-English, Kamwangamalu and Lee (1991:251-255) argue that the code-mixed sentences, which are assigned "Chinese" as the matrix code in most cases, involve "lexical" and "structural" substitution. That is, a Chinese constituent (i.e. lexical or phrasal) of an otherwise Chinese sentence is substituted by an English constituent of the same category in a code-mixed sentence.

Here, the category equivalence constraint is proposed mainly because it is descriptively most powerful compared to the other constraints and principles for Cantonese-English. (ref. section 4) Apart from its empirical adequacy, the constraint also explains the absence of certain English items, such as the auxiliary verbs, the articles and the clause complementizer "that", in Cantonese-English as single-words, since these items obviously do not have equivalents in Cantonese.

There is an apparent shortcoming of the category equivalence constraint. It cannot explain the non-switchability of pronouns, modals, quantifiers and possessive as single-word cases, even though these items have equivalents in Cantonese as well. My position concerning this is that some other constraints are accountable for the non-switchability of these items. The category equivalence constraint still works over most patterns of Cantonese-English code-mixing.

## 5.2 The bound morpheme constraint

The second constraint, the bound morpheme constraint, originates from the observation that in code-mixing only a bound morpheme of the matrix code can combine with a free morpheme of the embedded code, but not vice versa. Such a possibility can be illustrated by the obvious violations of the Free Morpheme constraint found in the Cantonese/English data, that is, the sequences of:

<u>Free</u>	<u>Bound</u>	
i/ V(English)	ASP (Cantonese)	(e.g. 3b)
kéuih deih <i>plan</i> jǒ	yī go syu ga	heui ɔu jǒu léuhk hǎhng
They PL	plan ASP DET CL	summer vacation go Europe travel
		They have planned to go traveling to Europe this summer vacation.

ii/ ADJ(English)	ADV marker(Cantonese)	(e.g. 5f)
kéuih hóu <i>sensible</i> gám wah...		
s/he EMP	sensible ADV	say
	marker	
		He/She says very sensibly that...

iii/ V/ADJ(English)	NEG(Cantonese)	V/ADJ (English) (e.g. 3c,5c)
go <i>program run</i> -m̃h-run dǒu a		
CL	program A-NOT-A	ASP Q
		Can the (computer) program run?

néih tīng yaht <i>free</i> -m̃h-free a	
you tomorrow	free(A-not-A) Q
	Are you free tomorrow?
(ref. 3.1.1, 3.1.3)	..

Besides, most of the alleged violations of the Free Morpheme constraint also assume the form of a free morpheme of the native language and a bound morpheme of the foreign language. (i.e. Bokamba on Arabic-English and Nairobic Swahili/English 1980; Bentahila and Davis on Arabic-French 1983; Nartey on Adanme-English 1982). In other studies, (i.e. Kamwangamalu on Bantu-English and Bantu-French 1989, Romaine on Punjabi-English 1989) it has been mentioned that roots of the foreign language are inflected by the native language morphology in code-mixing. The bound morpheme constraint is aimed at capturing these cases.

**(14) The bound morpheme constraint**

▪ A bound morpheme of the embedded code does not occur in code-mixing unless it is attached to the corresponding roots from the embedded code, according to the word-formation rules of the embedded code.<sup>3\*</sup>

With reference to the remaining problems of category equivalence constraint, the bound morpheme constraint may help solve the puzzle. If one considers the bound morphemes as markers of grammatical categories such as tense and agreement in a language, the bound morpheme constraint generally prohibits the grammatical categories of the embedded code from entering the matrix code. As regards the case of possessives, modals and pronouns, the first is automatically ruled out by the bound morpheme constraint. The non-switchability of the other two may be explained by the fact that their morphological forms are inextricably linked with grammatical categories like tense and agreement, and case respectively.

**5.3 The specifier constraint**

The third constraint, the specifier constraint, is in fact partly inspired by Gibbon's innermost constituent constraint, which stipulates that the innermost constituent, that is, the head of a phrase, must appear in English in order for its modifiers to appear in English as well. (ref. 4.2.3) Such a constraint is able to explain the non-switchability of quantifiers/articles as single-word in Cantonese-English code-mixing. (ref. Gibbons 1987:58,62) However, if one considers this constraint more carefully, one can easily find a lot of violations as well, which involve mixing of the following categories: attributive adjectives, adverbs, and object NP. (ref. 4.2.3 sentences 12a,b,c.) Yet, if one compares the quantifier or the article with these violations, the difference would be that the former is clearly a specifier while the latter items are only complements or adjuncts in terms of X-bar framework. In the light of these, the following hypothesis is proposed.

**(15) The specifier constraint**

▪ In Cantonese-English code-mixing, the specifier of a phrase does not appear in English without other constituents of the phrase, the head and, if any, the complements and adjuncts, appearing in English. The constraint holds as long as the Cantonese serves as the matrix code.<sup>4</sup>

The predictions of the specifier constraint would be such that the following items do not appear in English if the other constituents of the phrase do not appear in English:

- i/ quantifiers- NP specifier
- ii/ degree adverb- ADJP specifier
- iii/ primary auxiliary verbs - VP specifier
- iv/ subject NP (except Proper names)- IP specifier
- v/ wh-words -CP specifier

It is found that most of the predictions of the constraint conform to the data, except for a few cases in which the subject NP appears in English. Yet, their status as genuine violations of the specifier constraint is in doubt. There are two reasons: first, the nouns may be seen as syntactically adapted to Cantonese. As illustrated in the following examples, the nouns which appear in subject position are either proper names or generic noun phrases.

(13)

a. *Christine* jui gahn hóu chíh móuh mūt mood gám  
 Christine recently seem NEG CL mood PRT  
 Christine seems not to be in good mood recently.

b. *computer* hó yih tau gwo *keyboard* taih gōng yāt dĩ *feedback*  
 computer can through keyboard provide NUM CL feedback  
 The computer can provide some feedback through the keyboard.

The proper names, like "Christine" or "Rosalie", may be treated as NPs which are adapted to Cantonese, as they can be preceded by the Cantonese affix "a". Syntactic adaptation to Cantonese is even more obvious with the generic NP's as the article or the plural suffix are both omitted, which are otherwise needed to mark the generic usage of the common noun in English.

Secondly, the occurrence of these cases is rare. (i.e. 6 out of 260 cases of noun-mixing)

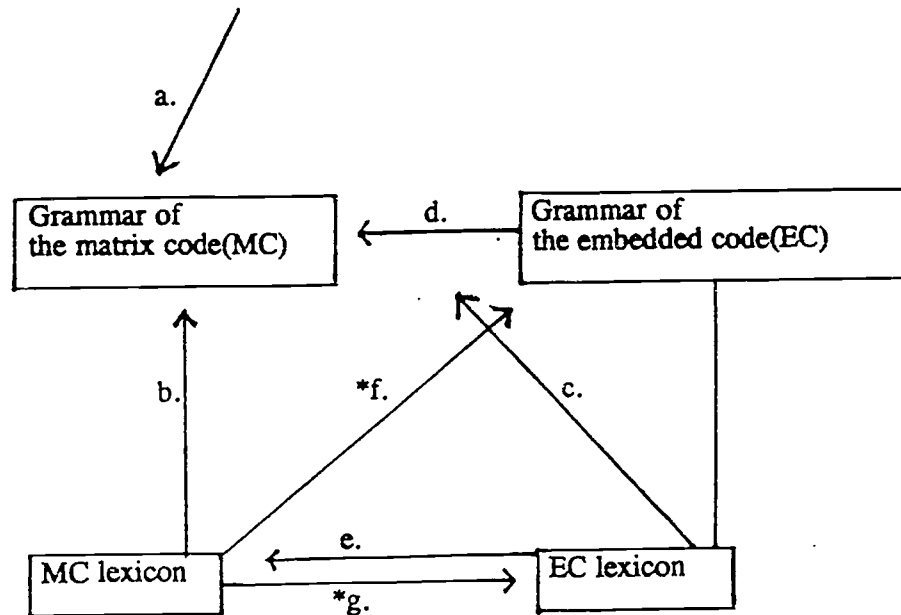
## 6. Psycholinguistic Processes of Code-mixing Production

This section attempts to investigate into the psycholinguistic processes that are involved in the production of code-mixing. Based on the earlier description of Cantonese-English and the three constraints, my own proposal for a production model of code-mixing would be something like this:

The code-mixer only has access to the matrix code(MC) in deriving the sentence structure of the code-mixed utterance. In my model, the matrix code may be either Cantonese or English. While the former is often the case, the latter is also possible in code-mixed utterances in which an English phrase structure is introduced. (ref. 5.1) The code-mixed utterance is produced with lexical items drawn from either the matrix code lexicon or the embedded code(EC)lexicon.

For the single-word cases, the lexical items are drawn directly from the embedded code lexicon to fit in a sentence structure generated by the matrix code grammar. For those elements of EC which take on the morphological characteristics of MC, they are drawn to the MC lexicon first, considering that the mental lexicon contains the morphological rules. (ref. Aitchison 1984, Ch10) This may capture cases in which the English lexical elements are incorporated into such distinct Cantonese morphological structures as "A-not-A" or reduplication. (ref. 3.1.1, 3.1.3) For those fragments the internal constituency of which observes EC grammar, the lexical elements of EC are drawn to the EC grammar from the EC lexicon before they fit in the sentence structure generated by the MC grammar. This may capture the case of English phrases in Cantonese-English. (ref. 3.2)

## THE CODE-MIXING BILINGUAL



### "A Psycholinguistic Model of Code-Mixing Production"

Grammar- PS rules;

Lexicon- lexical items, affixes and word-formation rules

a./ The code-mixing bilingual accesses the matrix code, which may be his/her L1 or L2, in deriving the sentence structure.

b./ Lexical items are drawn from the MC lexicon.

c./ Lexical items are drawn from the EC lexicon.

d./ Fragments are drawn from the EC lexicon through the EC grammar.

e./ Lexical items are drawn from the EC lexicon through the MC lexicon which are morphologically adapted

\*f./ The EC grammar, however, cannot take the items from the MC lexicon, either directly or through the MC grammar by the specifier constraint.

\*g./ Lexical items of the MC lexicon cannot enter the EC lexicon by the bound morpheme constraint.

The category equivalence constraint serves as a filter on interfaces c.-d., so that the elements from EC are fitted in the appropriate slots of the sentence structure generated by MC grammar. It also acts as a filter on interface e., so that morphological rules of MC apply to the lexical items of only certain categories from EC lexicon.

## NOTES

1. This article is a shortened version of my M.A. dissertation which was done at the Chinese University of Hong Kong. I wish to acknowledge the many helpful comments made by Dr Thomas Lee Hun-tak, Dr Virginia Yip, Professor David Pollard and others on the earlier drafts of the dissertation. My correspondence at present is as follows: Mr Brian Chan Hok-shing. The English Department. City Polytechnic of Hong Kong. Tak Chee Avenue. Kowloon. Hong Kong. Tel: 7887185. Email: hkucs!CPCCVX!ENBCHAN.
2. Gibbons (1979, 1987) and Leung (1988) are the only studies on the grammatical aspects of Cantonese-English I know.
3. The transcriptions in this paper follow the Yale system.
4. For a more detailed description of the bilingual situation in Hong Kong, as well as a sociolinguistic study of the use of Cantonese-English code-mixing, please refer to Gibbons (1987).
5. Since there may be more than one instance of mixing within one utterance, the "total" of adding up all types in "No. of utterances" exceeds 500, and, by the same reason, the "total" percentages by adding up those of all types exceeds 100.
6. I would not investigate further the origin of these sentences the structure of which is generated from English. My observation is that these sentences are not common, even in the code-mixing mode, and the speaker may utter it to achieve a special rhetorical effect. (e.g. to make emphasis)
7. The preposition-mixing cases which contain the sequence of NP COP P NP (e.g. 6a in 3.1.4) are considered to bear an English phrase structure in this paper. This is because the same idea needs to be expressed in an NP COP P NP P sequence (i.e. preposition-NP-postposition) in Cantonese. Accordingly, cases in which the PP contains a preposition-NP-postposition sequence (e.g. 6b in 3.1.4) are assumed to bear Cantonese as the matrix code.
8. Although prepositions and conjunctions, which are bound, may be mixed as single-words in Cantonese-English, they are excluded by the bound morpheme constraint as they do not form words with the Cantonese element according to the word-formation rules of English.
9. This is based on the assumption that a hypothesized sequence containing a specifier of EC and other constituents of MC, which is ruled out by the specifier constraint, is formed by the EC grammar drawing constituents from the MC lexicon. Logically, such a hypothesized sequence can also be formed by the following route: a specifier of EC alone is drawn from the EC lexicon to occupy a specifier position of a phrase structure generated by the MC grammar. It is argued in this paper that such a route is not possible, assuming that the specifiers of EC must enter the EC grammar in order to form a phrase with other constituents. Such an assumption is not arbitrary, since the specifiers are always language-specific markers of grammatical categories, such as the English articles (i.e. NP specifiers which mark definiteness), the auxiliary verbs (i.e. VP specifiers which mark aspect) and the intensifiers (i.e. AP specifiers which mark degree). (ref. Radford 1988)

The three constraints which I propose earlier can be represented as filters in the above model. The category equivalence constraint can be viewed as a constraint which limits the kind of elements of the embedded code which are to be embedded into particular slots in the sentence structure generated by the matrix code grammar. The bound morpheme constraint can be represented as a filter between the MC lexicon and the EC lexicon, so that elements of MC cannot take on the morphological characteristics of the EC. The specifier constraint is a filter between the MC lexicon and the EC grammar, so that elements of MC cannot be expanded according to the PS rules of the EC grammar.<sup>9</sup>

## 7. Conclusions

This paper describes patterns of Cantonese-English code-mixing according to the syntactic categories of the English elements. Based on the description, it is argued that the major constraints and principles that have been formulated on other varieties of code-mixing cannot apply to Cantonese-English as there are violations. It is then suggested that Cantonese-English code-mixing is subject to a number of different constraints, namely, the category equivalence constraint, the bound morpheme constraint and the specifier constraint. These constraints, apart from their empirical adequacy, also bear important implications on production of code-mixing.

## References

- Aitchison, J. (1987) *Words in the Mind*. Oxford: Basil Blackwell.
- Bentahila, A. and Davis, E. (1983) The syntax of Arabic-French code-switching. *Lingua* 59,301-330.
- Berk-Seligson, S. (1986) Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society* 15,313-348.
- Bokamba, E.G. (1989) Are there syntactic constraints on code-mixing? *World Englishes* 8,277-292.
- Clyne, M. (1987) Constraints on code-switching: How universal are they? *Linguistics* 25,739-764.
- Disciullo, A.M., Muysken, P. and Singh, R. (1986) Government and Code-switching. *Journal of Linguistics* 22,1-24.
- Gibbons, J. (1979) U-gay-wa: A linguistic study of the campus language of students at the University of Hong Kong. In Lord, R. (ed.) *The Hong Kong Language Papers*. Hong Kong University Press.
- (1987) *Code-mixing and Code-choice: A Hong Kong Case Study*. *Multilingual Matters* 27.
- Joshi, A.K. (1985) Processing of sentences with intrasentential code-switching. In Dowty, D., Karthman, L. and Zwicky, A.M. (eds.) *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*. Cambridge: Cambridge University Press.
- Kachru, B. (1978) Toward Structuring Code-mixing: an Indian Perspective. *International Journal of the Sociology of Language* 16,27-46.
- Kamwangamalu, N.M. (1989) The morphosyntactic aspects of French/English-Bantu code-mixing: Evidence for universal constraints. In *CLS 25, Parasession on Language in Context*. Chicago Linguistic Society.
- Kamwangamalu, N.M. and C-L. Lee (1991) Chinese-English code-mixing: a case of matrix language assignment. *World Englishes* 10, 247-261.
- Leung, Y-B. (1988) Constraints on Intrasentential Code-mixing in Cantonese and English. In *HK Papers in Linguistics and Language Teaching: Special Issues*. University of Hong Kong.
- Nartey, J.S. (1982) Code-switching, Interference or Faddism? Language use among educated Ghanians. *Anthropological Linguistics* 24,183-192.



- Nishimura, M. (1986) Intrasentential code-switching: the case of language assignment. In Vaid, J. (ed.) *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*. Hillsdale, NJ: Lawrence Erlbaum.
- (1989) The topic-comment construction in Japanese-English code-switching. In *World Englishes* 8, 365-377.
- Poplack, Shana (1980) Sometimes I'll start a sentence in Spanish: Y TERMINO EN ESPANOL: towards a typology of code-switching. *Linguistics* 18, 518-618.
- Poplack, S., Wheeler, S. and Westwood, A. (1989) Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. In *World Englishes* 8, 389-406.
- Radford, A. (1988) *Transformational Grammar*. Cambridge: Cambridge University Press.
- Romaine, Suzanne (1986) The syntax and semantics of the code-mixed compound verb in Punjabi/English bilingual discourse. In D. Tannen (ed.), *Linguistics and Language in Contact: The Interdependence of theory, data and application*. Washington, D.C.: Georgetown University Press. pp.35-49.
- (1989) *Bilingualism*. Basil Blackwell.
- Sankoff, D. and Poplack, S. (1981) A Formal Grammar of code-switching. *Papers in Linguistics* 14,1,3-46.
- Sankoff, D., Poplack, S. and Vanniarajan, S. (1990) The case of the nonce loan in Tamil. *Language Variation and Change*, 2, 71-101.
- Sridhar, S.N. and Sridhar, K.K. (1980) The Syntax and Psycholinguistics of Bilingual Code-mixing. *Canadian Journal of Psychology*, 34,4,407-416.