

Intrasentential Codeswitching: Bilingual Lemmas in Contact*

Longxing Wei

Montclair State University

This paper studies intrasentential codeswitching (ICS) in terms of the nature and activity of the bilingual mental lexicon. It claims that the mental lexicon does not simply contain lexemes but also “lemmas”, which are abstract entries in the mental lexicon that support the surface realization of actual lexemes because they contain semantic, syntactic, pragmatic, and morphological information about lexemes. It further claims that lemmas in the bilingual mental lexicon are language-specific and are in contact during a discourse involving ICS at three levels of abstract lexical structure: lexical-conceptual structure, predicate-argument structure, and morphological realization patterns. Activation of such lemmas at any level of abstract lexical structure motivates ICS. The ICS instances provide evidence that there is differential activation of participating languages in terms of language-specific lemma activation for morphemes and morphosyntactic procedures, and activated lemmas from the Embedded Language (EL) must be congruent with the counterparts of the Matrix Language (ML) (Myers-Scotton 1993a [1997]) for permissible ICS realizations.

Keywords: mental lexicon, lemma, bilingual, content, system, morpheme, activation, congruence.

1. Introduction

Intrasentential codeswitching (ICS) involves code switches between two languages (i.e. the alternative use of two languages) within the sentence boundary. Most previous studies of ICS (e.g. Lipski 1977, Pfaff 1979, Poplack 1980, Sridhar and Sridhar 1980, Gumperz 1982, Woodford 1983, inter alia) focused on describing surface configurations of switched items in terms of the switch points (i.e. where the switching structurally possible) and the switched items (i.e. what items from another language can be switched). Different from such studies, this paper is a study of ICS at an abstract level of bilingual speech production process in relation to the nature and activity of the bilingual mental lexicon. It defines ICS as a phenomenon of languages in contact, to be more specific, as a phenomenon of lemmas from the bilingual mental lexicon in contact when the sentence contains linguistic items switched from another language.

As generally defined, the mental lexicon contains lexemes (i.e. minimal lexical units in a language, as words or idiomatic phrases, especially as distinguished from

* I would like to express my many sincere thanks to Professor Joy Wu, the other editor in chief of *Concentric*, Professor Elizabeth Zeitoun, and the anonymous reviewers of this paper for their insightful comments on its content and invaluable suggestions for its revisions. Without their professional help, the current version of the paper would be impossible. Also, I would like to thank Mr. David Ta-Chun Shen, the editorial assistant, for his tireless work and great patience with me.

their grammatical and syntactic aspects) and their meanings. Following Levelt (1989), Myers-Scotton and Jake (1995), and Wei (2001, 2002), this paper claims that the mental lexicon does not simply contain lexemes and their meanings but more abstract elements called “lemmas”. Lemmas are defined as abstract entries in the mental lexicon that support the surface realization of actual lexemes. This is because for each item, the mental lexicon contains lemma information, that is, declarative knowledge about the word’s meaning, and information about its syntax and morphology which is necessary for constructing the word’s syntactic environment. For instance, the lemmas of *know* require a subject that expresses the role of EXPERIENCER, an object that expresses the role of PERCEPT (i.e. what is known), and these elements appear in a particular order; the lemmas of *he* require the word to be used of a male and that any following present-tense main verb must have the suffix *-s* attached to it (i.e. inflectional morphology for tense marking). Lemmas also contain information about the word’s phonological structure, syllabic composition, and accent structure. In addition, lemmas may contain information about the word’s register, the kind of discourse it typically enters into, and its pragmatic function. Thus, the mental lexicon is defined as the speaker’s internal representation of language specific knowledge about the surface forms, that is, lemmas.

Researchers may differ in their views on how appropriate lemmas are activated to convey the speaker’s intended meaning. For example, de Bot and Schreuder (1993) think that lemmas are activated by pieces of conceptual structure sent by the verbalizer; Poullisse and Bongaerts (1994) believe that lemmas are not only activated by the conceptual information but also by a language cue sent by the conceptualizer; Myers-Scotton and Jake (1995) claim that lemmas are activated by language-specific semantic/pragmatic feature bundles that come from the conceptualizer; Grosjean (1997) holds that whatever the origin and the nature of information that lemmas may receive, the actual activation or choice of lemmas must be based on the information sent to the mental lexicon. Regarding lemma activation, especially lemma activation in speech production involving ICS, this paper draws heavily on the theoretical assumptions presented by Myers-Scotton and Jake (1995, 2000) that all lemmas include three distinctive but related levels of abstract lexical structure. The first is the level of lexical-conceptual structure. Lexical access takes place on the basis of the information contained in the speaker’s preverbal message, that is, the speaker’s intention before speech production. It is the speaker’s preverbal message in the conceptualizer which activates language-specific semantic/pragmatic feature bundles at the interface between the conceptualizer and the mental lexicon. These activated semantic/pragmatic features are then mapped onto abstract entries in the mental lexicon (i.e. lemmas), as lexical-conceptual structure. Green (1993) also holds that a

lemma is activated if it matches part of the conceptual structure created by the conceptualizer. The second is the level of predicate-argument structure, which maps the thematic structure of a particular verb onto grammatical relations (i.e. thematic role assignment). The third is the level of morphological realization patterns, which deals with surface grammatical relations (e.g. word order, agreement, inflectional morphology for tense/aspect marking, etc.).

This paper claims that lemmas contained in the bilingual mental lexicon are language-specific and such language-specific lemmas are in contact during a discourse involving ICS at these levels of abstract lexical structure. Accordingly, ICS is described and explained in terms of activation of language-specific lemmas at any of these levels of abstract lexical structure. This paper specifically deals with the issues of unequal activation of bilingual lemmas from the language pairs involved in ICS. In addition, it discusses issues of lemma congruence checking between the languages involved in ICS as an organizing principle governing such a bilingual activity. This type of the study of ICS leads to several specific hypotheses about structural principles governing the bilingual speech production process.

This paper presents three arguments: (1) Lemmas are not equally activated in either monolingual or bilingual speech production. Unequal activation of lemmas from the bilingual mental lexicon motivates and constraints the speaker's choice of morphemes in ICS. (2) Morphosyntactic procedures activated by lemmas from the bilingual mental lexicon must involve parallel processing, but with one procedure of one language blind to the workings of another. It is the structural procedures dictated by the ML (more commonly called "base/host language") which set the sentential frame for sentences containing items switched from the EL (more commonly called "guest language") (Myers-Scotton 1993a [1997]). (3) Bilinguals can access whatever language known to them as the EL in ICS, but the activated lemmas from the EL must be sufficiently congruent with the ML counterparts at each of the three levels of abstract lexical structure (i.e. lexical-conceptual structure, predicate-argument structure, and morphological realization patterns). Insufficient congruence or incongruence between languages at any of these levels restricts choices and affects possible ICS configurations. The major assumption underlying this paper is that ICS itself is a linguistic system, and, like any linguistic system, it is governed and constrained by a set of morphosyntactic principles and rules. The so-called "mixed" speech phenomenon cannot be accounted for without exploring the nature of the bilingual mental lexicon and the activity of bilingual lemmas in contact during the speech production process.

The representative ICS instances for the study were from a research project on language pairs involved in ICS, such as Chinese/English, Japanese/English,

English/Chinese and English/Japanese. The naturally occurring instances of ICS were collected from the daily conversations of the native speakers of Chinese and Japanese who were students studying at the University of South Carolina or other universities and colleges in South Carolina, USA. All the bilinguals observed in their codeswitching behavior were adult learners of English as a second language. Their English proficiency levels were relatively high, either at the high intermediate or advanced levels, because they must have achieved the acceptable scores of TOFEL and/or GRE for their admissions to American universities or colleges. The data collection was conducted over a period of three years. The ICS instances were either note- or tape-recorded, transcribed, coded and analyzed for their morphosyntactic features by those who knew and spoke the languages. Each of the ICS instances discussed in this paper is selected from this data pool. The findings provide empirical evidence for the claims and arguments presented in this paper.

2. Roles of participating languages in ICS

The Matrix Language Frame (MLF) model proposed by Myers-Scotton (1993a [1997]) is the model used in this paper for describing and explaining ICS. As an over-arching principle of the MLF model, the Uniform Structure Principle specifies (Myers-Scotton 2002:8-9):

A given constituent type in any language has a uniform abstract structure and the requirements for well-formedness for this constituent type must be observed whenever the constituent appears. In bilingual speech, the structures of the Matrix Language are always preferred, but some Embedded Language structures are allowed if certain conditions are met.

This principle makes two asymmetries in language contact (“bilingual speech” being studied in this paper) obvious: matrix languages (MLs) vs. embedded languages (ELs), and content vs. system morphemes. The first asymmetry determines how switched items (i.e. morphemes) can occur in a sentence, and the second asymmetry determines what types of morphemes can be switched. It is these asymmetries or distinctions which govern and constrain ICS.

2.1 Matrix vs. embedded languages

The MLF model identifies one of the languages participating in ICS as the ML and the other(s) as the EL(s) (Myers-Scotton 1993a [1997]). As noted in

Myers-Scotton (1993a:66-67), a consideration of “psycholinguistic and sociolinguistic criteria point to a definition” of the ML but provides “a frequency based criterion” to objectively identify the ML: “the ML is the language of more morphemes in interaction types including intrasentential codeswitching” (1993a:68). The ML vs. EL asymmetry is much more than a heuristic device for labeling constituents. Rather, it underlies the differential participation of the ML and the EL in shaping utterances involving ICS, with the ML having the dominant role. It is the ML which is the sole or main source of the morphosyntactic frame structuring the bilingual utterance. This dominance is dramatic in ML + EL constituents: only the ML can project the sentential frame (i.e. the word order, the inflectional morphology and other functional elements). Example (1) illustrates the Uniform Structure Principle and the MLF model.

(Chinese/English)

- (1) Wo you liang-fen **paper** mingtian bixu jiaoshangqu,
I have two-CL tomorrow must turn in
ke wo xianzai yi-fen hai mei **finish** ne.
but I now one-CL yet not PART/AFFIRM
'I have two papers [which] I must turn in tomorrow, but at the moment I haven't finished one yet.'¹

In the ML + EL constituents of this example, there are two elements from English (*paper* and *finish*), but the elements that structure the sentence all come from Chinese (the ML), including the word order, the way that plurality of *paper* is indicated, and the way that negation of *finish* is realized. As assumed in the MLF model, the languages participating in ICS have unequal roles. One language is more central than the other in sentential frame-building. The ML is more activated than the EL in a discourse involving ICS and the occurrence of its morphemes is more frequent and freer than that of the EL. The ML is the language which the speakers engaged in ICS as the “main language” being used.

2.2 Content vs. system morphemes

The asymmetry of content vs. system morphemes is another opposition structuring ICS utterances (Myers-Scotton 1993a [1997]). Content and system morphemes are

¹ Abbreviations: AFFIRM = affirmative; AUX = auxiliary; CL = classifier of noun; COP = copula; INTERROG = interrogative; INTJ = interjection; LOC = location; MANN = manner; NOM = nominative; OBJ = object; PART = particle; PAST = past; PERF = perfect; POSS = possessive; PREP = preposition; PROG = progressive; QUE = question; TOP = topic.

accessed differently in either monolingual or bilingual speech production (Jake 1994, Myers-Scotton and Jake 1995, Wei 2000a). Prototypical content morphemes are nouns, most verbs, descriptive adjectives, most prepositions, and free-standing pronouns. More formally, content morphemes have a plus setting for the lexical feature [\pm thematic role assigner/receiver]. That is, they can be characterized as categories that are potential thematic role assigners or receivers. In contrast, system morphemes have a minus setting for this feature. That is, system morphemes do not assign or receive a thematic role of any kind. Prototypical system morphemes are quantifiers, specifiers, and inflectional affixes. Example (2) illustrates the distinction between content and system morphemes.

(Chinese/English)

- (2) Ni nei-pian **article** hai mei **finish** a?
 you that-CL yet not PART/AFFIRM-QUE
 ‘You haven’t finished that article yet?’

In this example, *article* and *finish* are content morphemes from English, but *nei* (equivalent to ‘that’) and *pian* (a Chinese classifier) go together as a determiner. A Chinese classifier is a word or morpheme that indicates a semantic class of nouns and regularly accompanies any noun of that class in certain syntactic constructions. A Chinese classifier must always be used together with a singular demonstrative like *zhe* (‘this’) or *nei* (‘that’) or a specific number like *yi* (‘one’), *liang* (‘two’), and so on. When this combination occurs with a content morpheme head, it syntactically functions as a system morpheme. However, when the same combination receives a thematic role, it must be a content morpheme. For example, when we say *gei wo nei-ge* (‘give me that’). Here, *neige* (‘that’ or ‘that one’) is assigned a thematic role of THEME. It should also be noticed that there is no perfect aspect making on the English verb *finish*, since Chinese does not have verb morphology of any sort for this and other grammatical purposes. Chinese does have a morpheme which realizes the concept of aspect, but the verb itself is not inflectional. Also, as shown in Example (1), the negation of the verb is indicated by *mei* (‘not’), one of the Chinese system morphemes for negation of verbs. It seems that there is a clear distinction between content and system morphemes, and content and system morphemes are not equally accessed in producing utterances involving ICS.

In the current context of study, the content vs. system morpheme distinction is slightly recast in light of the sources of morphemes. The sources of morphemes refer to differences at the levels of abstract lexical structure. Some morphemes are directly elected (i.e. activated) at the conceptual level to realize the lexical-conceptual

structure of lexemes. This is because at the conceptual level, speaker intentions are mapped onto semantic/pragmatic feature bundles. This is the mapping of speaker intentions to lemma entries in the mental lexicon. Such directly-elected morphemes (Bock and Levelt 1994, Wei 1996, 2000a, 2000b, Myers-Scotton 2002) supported by those activated lemmas in the mental lexicon are content morphemes. In contrast, certain morphemes are the result of the spelling out of aspects of the morphological realization of a lexeme. Such morphemes do not reflect the semantic/pragmatic features conflated in the lexical-conceptual structure but are only activated in the projection of morphophonological encoding at the positional level. For example, case assignment reflects predicate-argument structure, some phi-features license phonetic realizations, and inflectional morphology realizes agreement or tense/aspect marking. Morphemes for such grammatical purposes are not elected or activated by semantic/pragmatic feature bundles as required by speaker intentions. Such morphemes are “structurally-assigned” morphemes (Wei 1996) or system morphemes (Myers-Scotton 1993a [1997]).

2.3 Central structural principles governing ICS

In the MLF model (Myers-Scotton 1993a [1997]), several central structural principles are proposed in explaining grammatical constraints on ICS and predicting permissible ML + EL constituents. In addition to the Uniform Structure Principle, two principles under the Matrix Language Hypothesis are proposed: The System Morpheme Principle and the Morpheme Order Principle. The System Morpheme Principle specifies that in ML + EL constituents, all syntactically relevant system morphemes come only from the ML, and the EL only supplies content morphemes to the sentential or morphosyntactic frame projected by the ML. The Morpheme Order Principle specifies that in ML + EL constituents, the ML surface morpheme order must not be violated. In addition, the Lemma Congruence Checking Principle (Myers-Scotton & Jake 1995, Wei 2001, 2002) specifies that lemmas activated from the EL must be congruent with the ML counterparts; otherwise, compromise strategies must be taken for possible ICS configurations. With the ML vs. EL distinction and content vs. system morpheme distinction, these central structural principles govern ICS. These principles are further tested in this study of the nature and activity of lemmas in the bilingual mental lexicon during bilingual speech.

3. A bilingual lemma activation model

Adopting Levelt's (1989) speech production model and Myers-Scotton and Jake's (1995) bilingual language competence and production model, Wei (2006) proposes a bilingual lemma activation model to explain bilingual speech production process involving ICS. This model is reintroduced here to explain why bilingual lemmas are in contact during the process of ICS. The model (see Figure 1 below) consists of four sequentially connected and ordered levels of speech production: conceptual level → lemma level → functional level → positional level. At the conceptual level the CONCEPTUALIZER generates messages by attending to the bilingual speaker's communicative intention about the discourse mode (cf. Grosjean 1997) and preverbal message (cf. Myers-Scotton and Jake 1995, 2000) to be desired. If the speaker chooses the monolingual mode, no codeswitching will occur; if the speaker chooses the bilingual mode, then he/she must decide whether intersentential or intrasentential codeswitching should be performed. If the speaker decides to perform ICS, the CONCEPTUALIZER then generates preverbal message about his/her choice of the language as the ML to be used and semantic/pragmatic feature bundles to be desired for his/her communicative intention. The output of the CONCEPTUALIZER is the speaker's preverbal message which gives input to the BILINGUAL MENTAL LEXICON at the lemma level. It is at this level that language-specific lemmas are in contact for activation. The activation of language-specific lemmas alone is not sufficient enough for ICS to occur. Lemma congruence checking between the languages involved must come into play. Myers-Scotton and Jake (1995:985) define lemma congruence as "a match between the ML and the EL at the lemma level with respect to linguistically relevant features." Lemma congruence between the languages involved must be checked at the level of lexical-conceptual structure, at the level of predicate-argument structure, and at the level of morphological realization patterns. Lemma congruence checking is regarded as an organizing principle for ICS (Wei 2001, 2002, 2006). If the lemmas of the EL are congruent or sufficiently congruent with the counterparts of the ML at each of these levels, they can be activated for the speaker to proceed with the bilingual mode for ICS; otherwise, the speaker must take some compromise strategies for possible ICS realization (or give up the bilingual mode and go back to the monolingual mode. Only when there is a match between the EL and ML lemmas or compromise strategies are taken, directions will be sent to the FORMULATOR at the functional level for morphosyntactic encoding by observing the structural principles governing ICS. The successfully encoded morphosyntactic material will then be sent to the ARTICULATOR at the positional level for morphophonological encoding. The successfully encoded morphophonological

material will then produce surface forms of word order and phonetic string, that is, speech output, for speech comprehension.

The bilingual activation model adopted here to describe and explain bilingual speech production processes also draws on some contemporary proposals and assumptions about the nature and activity of bilingual language modes in speech production. For example, regarding the activation and deactivation of language modes, Faerch and Kasper (1986) and Grosjean and Soares (1986) assume that the bilingual's language systems can be kept separate because they may be activated at different levels, depending on whether they are used at the moment. Green (1986) and Paradis (1987, 1997) propose that the language system of a bilingual is organized in two subsets, one for each language, that can be selectively activated, activated simultaneously to various degrees, or deactivated independently of one another. When the bilingual is in a monolingual mode, it is assumed that the other language must be deactivated. Green (1986) even proposes that the other language is inhibited. Green (1986) postulates that bilingual speakers who wish to select a particular language for the current communication must ensure that its activation exceeds that of the competing language(s). So, when bilinguals speak L2, L2 must be selected and highly activated, and L1, which is still active, must be deactivated. According to Soares and Grosjean (1984), Green (1986) and Grosjean (1997), deactivation, in fact, is rarely total, as clearly evidenced in the interference effects in bilinguals' production of the speaker-specific deviations from the target language being spoken. This is clearly due to the influence of the deactivated language. Paradis (1997) also assumes that when bilingual speakers select one language rather than another, the activation threshold of the nonselected language must be raised; however, the language not being selected for use at the moment is never totally deactivated. Paradis (1989) and Berg and Schade (1992) even claim that deactivation of one language may be so difficult that abundance involuntary mixing or hybridization may occur. Therefore, inhibition or deactivation of one of a bilingual's languages is a matter of degree (Paradis 1997). Green's (1986) model of activation vs. deactivation of language modes offers two advantages. One advantage is that it postulates a specifier, which specifies how the system is to be controlled especially in the case of codeswitching, that sets the activation going. The other advantage is that it pays much attention to the resources that are needed to control the activation levels. Green describes such sources as energy or fuel, which must exist in order for the system to function. Green's postulation of such resources makes it possible to explain why beginning learners of L2 experience more L1 interference effects than advanced learners. This is because beginning learners need to invest much more energy or fuel in speaking L2

before their L2 production becomes automatized. In other words, they need more L2 resources to suppress the activation level of L1.

However, Grosjean and Soares point out that “this deactivation has led to much theorizing and much controversy around the notion of a language switch or a monitor system,” but “what is certain, however, is that bilinguals rarely deactivate the other language totally. This is seen in various types of production interference – the involuntary influence of one language on the other – ...” (1986:146). However, when the bilingual is in the bilingual mode, both languages are activated, with the base language more strongly activated than the other. In explaining codeswitching, Grosjean (1997) proposes that if the bilingual is in the monolingual mode the guest language becomes deactivated as best as possible, and, as a consequence, the deactivated guest language does not act upon the base language often. In contrast, if the bilingual is in the bilingual mode, the guest language becomes activated to a relatively high level but less so than the base language, and thus the activated guest language intrudes upon the base language very often. According to Grosjean (1985, 1989, 1994, 1997), bilinguals find themselves in their everyday lives at various points along the language mode continuum, which corresponds to different levels of activation of the two languages but, particularly, of the guest language, because the base language is always more strongly activated than the guest language, that is, it probably never descends much below full activation. Thus, it is assumed that it is the language mode chosen and the relative degree of activation of the base and guest languages that cause the amount of language mixing that takes place in the bilingual’s speech.

The model (Wei 2006) was adapted from the models of Levelt (1989), Myers-Scotton and Jake (1995) and Wei (2002). This model adds to those proposals and assumptions the notion of sequentially ordered levels of bilingual speech production and the notion of bilingual lemma activation and congruence checking in bilingual speech production involving ICS.

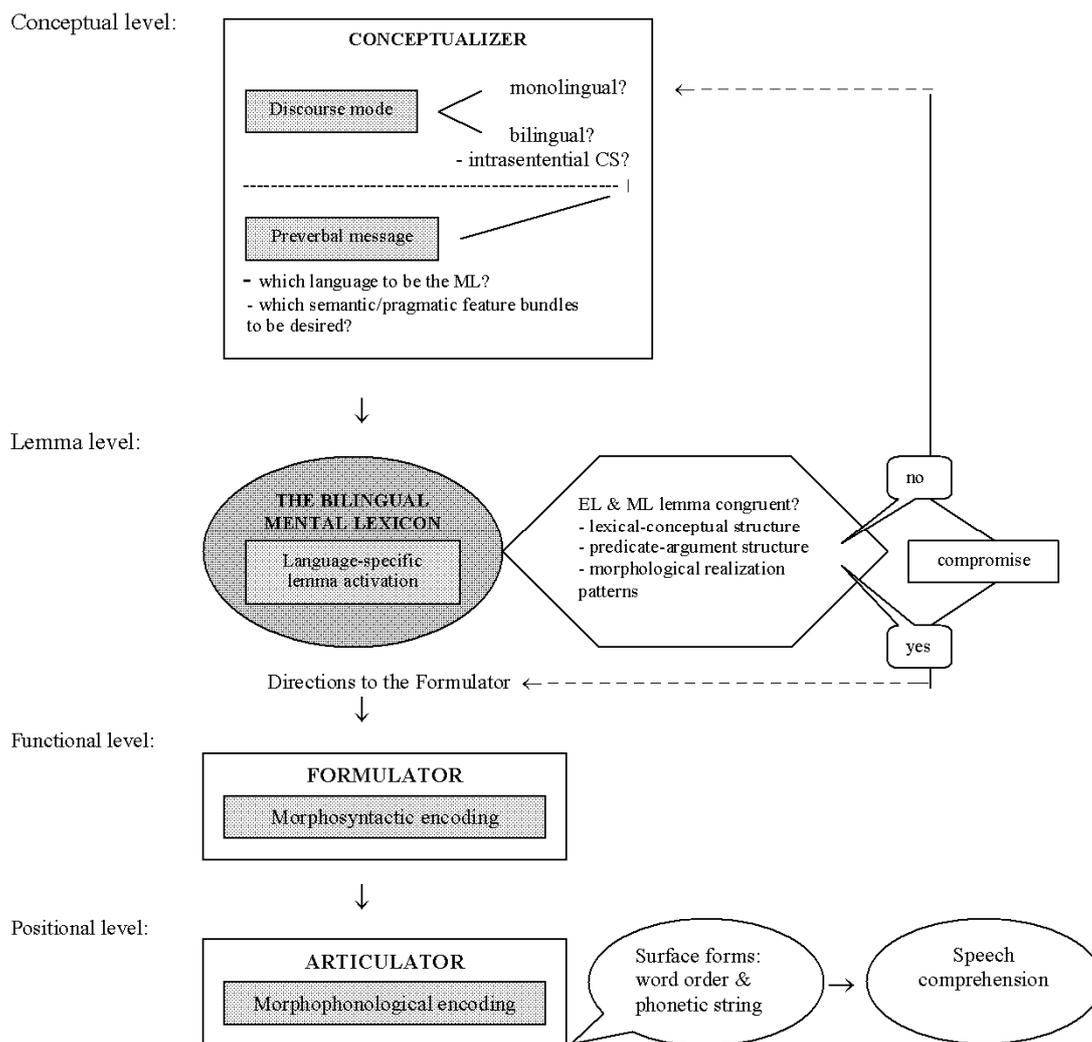


Figure 1. A bilingual lemma activation model

(Wei 2006, adapted from Levelt 1989, Myers-Scotton and Jake 1995, Wei 2002)

4. Bilingual lemmas in contact in ICS

This paper assumes that ICS juxtapositions which may surface do not have much to do with surface linear of typological correspondences between the participating languages. Instead, it proposes that ICS juxtapositions originate with directions contained in lemmas (cf. Myers-Scotton and Jake 1995, Wei 2001). As introduced earlier, lemmas are abstract entries in the speaker’s mental lexicon which support the surface realization of actual lexemes. This is because lemmas contain semantic, syntactic, pragmatic, and morphological information about lexemes stored in the mental lexicon. Thus, lemmas in the mental lexicon are defined as the speaker’s internal representation of knowledge about surface forms. This paper further argues

that lemmas in the bilingual mental lexicon are language specific, and such lemmas are in contact in bilingual speech, especially in ICS.

According to Green (1986, 1993) and de Bot and Schreuder (1993), the lexical items that belong to different languages must be organized in subsets which are either activated or deactivated in their entirety. According to Myers-Scotton (1993a [1997]), there is a universal set of semantic and pragmatic features available for the lexical-conceptual structure of lemmas. This paper assumes that lemmas in the bilingual mental lexicon are language-specific and lexicalization patterns across languages reflect the fact that there are different configurations of semantic and pragmatic features across related lemmas in different languages. Hypotheses about cross-linguistic differences in how information is organized at the level of lexical-conceptual structure and at the level of predicate-argument structure, whether semantic or pragmatic or semantic with morphological consequences, affect the code choices and structures which will appear in ICS.

The type of ICS discussed in this paper is called “classic codeswitching” in Myers-Scotton (2002:8), which refers to speech containing switched items produced by the bilingual speakers who are proficient enough to produce well-formed utterances in each of the languages participating in ICS, and who are proficient enough in using either participating language as the ML (i.e. the abstract morphosyntactic frame of the utterances containing switched items). In other words, classic codeswitching requires speakers to have full access to the morphosyntactic frame of one of the participating languages and also have enough proficiency in the other language either to switch EL content morphemes into mixed constituents framed by the ML or produce well formed EL islands (EL islands are constituents that show structural dependency relations and are well formed in the EL grammar. Thus, EL islands include not just EL content morphemes, but all relevant EL system morphemes), or both. It becomes obvious that speakers engaging in classic codeswitching must have some proficiency in the EL(s). However, the extent of their proficiency may be less than what is needed for them to produce well-formed ML morphosyntactic frames and syntactically relevant ML grammatical elements. What is needed in classic codeswitching is that speakers’ proficiency must be sufficient enough for them to assess whether EL content morphemes can appear in the ML frame. Also, speakers’ EL grammatical proficiency level must be high enough for them to produce EL islands in ICS.

4.1 Unequal lemma activation for morphemes

As explained earlier, lemmas contain bundles of semantic and pragmatic features which encode the lexical-conceptual structure representing the speaker's communicative intention or preverbal message generated by the CONCEPTUALIZER (cf. Levelt 1989). This is because each individual conceptual chunk is specified for a particular language in the speaker's preverbal message (Poulishse and Bongaerts 1994). What needs to be emphasized here is that it is at the level of lexical-conceptual structure that the speaker seeks appropriate linguistic material for his/her communicative intention. It is at this level that the speaker conceptualizes a message and activates the appropriate concepts accordingly. The activated concepts will then spread activation to the corresponding lemmas in the mental lexicon. Sufficiently activated lemmas will then spread activation to the associated lexemes (cf. Roelofs 1992, Levelt 1995, Wei 2001). It is assumed that there is a universal set of semantic and pragmatic features available for the lexical-conceptual structuring of lemmas (Myers-Scotton 1993a [1997]); however, it is also assumed that the presence and conflation of universally available semantic and pragmatic features may vary cross-linguistically (Wei 2001). In this paper, the structures appearing in ICS are seen as evidence of the relative importance of cross-linguistic lexical-conceptual differences in lemmas in the bilingual mental lexicon, and the consequences of ICS are considered as variation in semantic/pragmatic feature bundles.

One of the major assumptions underlying this study is that lemmas in the bilingual mental lexicon are language-specific, that is, they are tagged for a specific language and support the realization of actual lexemes. This paper suggests that it is individual content morphemes which encode the speaker's specific communicative intention. However, at the conceptual level, the speaker does not produce surface level morphemes but rather makes choices about the semantic and pragmatic information which he/she wishes to convey. This paper further suggests that one of the major reasons why content and system morphemes are accessed differently is that it is content morphemes, rather than system morphemes, which contain semantic/pragmatic feature bundles. The speaker may choose certain EL content morphemes at a certain point during a discourse to convey his/her intended or desired meaning as generated in his/her preverbal message. It is in this sense that content and system morphemes are not equally activated, and it is also in this sense that certain language-specific lemmas are conceptually projected in ICS.

Below are some typical examples of ICS which show that EL content morphemes can be switched because they are projected from the EL lemmas which are sufficiently congruent with those of the ML.

(Chinese/English)

(3) Wo xianzai **course** xiu de chabuduo le,
 I now take PART/MANN almost PART/PERF
 dan hai dei xiu liang-men **foreign language**.
 but still must study/take two-CL
 ‘I’ve now taken almost all the courses, but I still must take two foreign language courses.’

(4) Naxie **visiting scholar** bu shi hen youqian ma,
 those not/EMPH COP very rich PART/AFFIRM
 bi women **student** you qian duo le.
 PREP/than us have money more PART/AFFIRM
 ‘Aren’t those visiting scholars very rich? They have a lot more money than us students.’

(5) Wo **daughter** mingtian guo **birthday**.
 my tomorrow have
 ‘My daughter will have her birthday tomorrow.’

(6) Shi-bu-shi qu nei-ge **new library**?
 yes-not-yes go that-CL
 ‘Are we going to that new library?’

(7) Wo **summer** bu **take course** le.
 I not PART/AFFIRM
 ‘I won’t take any course in summer.’

- (8) You xuduo **homework** yao zuo; hai you
 have a lot of must do in addition have
 hao ji-pian **article** xiang qu **library** -de **computer**
 quite a few-CL want go -POSS
 shang **check** yixia.
 on once
 ‘I’ve a lot of homework to do. In addition, I’ve quite a few articles (that) I want to go to check on the library’s computer.’

- (9) Mei you zijide jiqi feichang bu fangbian
 not have own machine very not convenient
 feichang **inconvenient**.
 very
 ‘It’s very inconvenient if I don’t have my own machine, very inconvenient.’

- (10) Tingshuo ni **in May graduate**, shi ma?
 hear you right PART/INTERROG
 ‘I heard that you will graduate in May, won’t you?’

The nouns and noun phrases *course* and *foreign language* in (3) and *visiting scholar* and *student* in (4) are all English, the EL, content morphemes, but the EL system morpheme for plural marking do not appear, and the demonstrative determiner *naxie* (‘those’) is an ML system morpheme. The nouns *daughter* and *birthday* in (5) are English content morphemes without the English possessive determiners. The noun phrase *new library* in (6) appears with the Chinese, the ML, demonstrative determiner ‘that’. Again, the noun *summer* and the verb phrase *take course* in (7) are English content morphemes. It should be noted that the equivalent Chinese preposition *zai* (‘in’) is optional in realizing an adverbial of time or place. Also, while English requires an auxiliary verb, a system morpheme, to negate a predicate verb, Chinese does not have one for this grammatical purpose. *bu* (‘not’), a system morpheme, comes from the ML. The nouns *homework*, *article*, *library*, *computer* and the verb *check* in (8) are all English content morphemes, but the system morphemes *xuduo* (‘much/a lot of’), *ji-pian* (‘a few-CL’) and *-de* (adjective marker) come from Chinese, and the English infinitive marker *to*, a system morpheme, does not appear to introduce the verb *check*. The adjective *inconvenient* in (9) is an English content morpheme, but *feichang* (‘very’), a system morpheme indicating degree, is from Chinese. The prepositional phrase *in May* and the verb *graduate* in (10) are English content

morphemes, but the English auxiliary verb, a system morpheme, to indicate a future activity does not appear.

These examples show that EL content morphemes, rather than EL system morphemes, can be activated in ICS. It is the ML which supplies all the system morphemes (if they are available in the ML) for the grammatical purposes. The similar phenomena are also observed in the Japanese/English ICS instances.

(Japanese/English)

(11) Muzukashikat-ta to iu ka, aa sore **essay**
 difficult-PAST PART say PART PART that
 datt-ta kara wakara-nai, um.
 PART-PAST because understand-not PART
 ‘It could be said to be difficult, mm because I didn’t understand that essay, mm.’

(12) Ii desu keredomo **tuition** ga totemo **expensive**
 good COP/be but PART/NOM very
 desu.
 COP/be
 ‘It’s good, but the tuition is very expensive.’

(13) Nihon demo saikin kekko **drug** o ya-tte iru
 Japan also recently rather PART/OBJ do-PROG AUX
 hito ga ooi yo.
 people PART/NOM many INTJ
 ‘Recently in Japan people who are doing drugs are also many.’

(14) Ima wa **summer course** o tot-te
 Now PART/TOP PART/OBJ take-PROG
 iru n.
 AUX/be PART
 ‘I’m taking summer courses now.’

(15) Dore gurai koko ni **stay** suru no?
 how long about here LOC do QUE
 ‘About how long will you stay here?’

The noun *essay* in (11) is an English, the EL, content morpheme, but the demonstrative determiner *sore* ('that'), a system morpheme, is from Japanese, the ML.

The noun *tuition* and the adjective *expensive* in (12) are EL content morphemes, but the EL definite article, a system morpheme, does not appear before *tuition*, and the system morpheme *totemo* ('very') modifying *expensive* is from the ML. The noun *drug* in (13) is an EL content morpheme, but the EL plural *-s*, a system morpheme, does not appear. The noun phrase *summer course* in (14) is from the EL, but *o* marking the accusative case, a system morpheme, is from the ML. The noun *stay* in (15) is an EL content morpheme used in conjunction with the ML *suru* ('do') to meet this particular lexical structure in Japanese.

The above examples provide the evidence that in ICS, bilingual lemmas are in contact and are not equally activated. EL content morphemes can be freely activated to be switched for the speaker's communicative intention at a certain point during a discourse, but EL system morphemes cannot. It should also be noticed that although Chinese and Japanese have a different basic word order and Japanese possesses morphology for case and tense/aspect marking but Chinese does not, the System Morpheme Principle applies to both Chinese/English and Japanese/English ICS. That is, all the system morphemes must come from the ML.

4.2 Unequal lemma activation for morphosyntactic procedures

Central to the discussion of ICS is the claim that the language pairs involved do not participate equally. In a bilingual mode, although both language networks are activated, one is more activated than the other (Green 1986, Grosjean 1997). According to the MLF model (Myers-Scotton 1993a [1997]), ICS occurs within the constraints of a sentential frame which is set by structural procedures dictated by the ML for ML + EL constituents. That is, this sentential frame is determined by the ML morpheme order and syntactically relevant relational or functional elements (i.e. system morphemes). If the speaker chooses to engage in ICS at a certain point during a discourse, he/she automatically selects one of the participating languages as the ML to provide the sentential frame. Figure 1 shows that the activated language-specific lemmas send directions to the FORMULATOR at the functional level for morphosyntactic encoding. As assumed, at the conceptual level the discourse mode is chosen with one language as the ML and then corresponding language-specific lemmas are activated at the lemma level to realize the speaker's preverbal message, resulting in language-specific morphosyntax.

It is up to the speaker to choose any of the languages he/she knows as the ML based on various factors, such as the languages known to the participants in the same discourse (i.e. the interlocutors), the particular speech settings (i.e. the contexts), the particular conversation topics, the speaker's social motivations, the speaker's pragmatic intentions, the speaker's semantic implications, etc. (see e.g. Gumperz 1977, Poplack 1980, Scotton 1983, Appel and Muysken 1987, Myers-Scotton 1993b, Li and Milroy 1995, Milroy and Muysken 1995, Romaine 1995, Li 1996, Nishimura 1997). The study of various factors involved in choosing one language rather than the other as the ML is beyond the scope of this paper. What is emphasized here is that whichever language chosen by the speaker as the ML always plays a central role in building the sentential frame for ML + EL constituents throughout ICS.

The above Chinese/English and Japanese/English ICS examples provide evidence for the morphosyntactic procedures controlled by the ML. That is, it is the ML which provides the sentential frame into which EL items are switched. Although Chinese and English share the same basic SVO order, Chinese is flexible in the arrangement of words in a sentence. Once Chinese is chosen as the ML, it sets up the sentential frame into which the EL content morphemes are switched. In (3) the object *course* goes before the verb *xiu* ('take'); in (4) the prepositional phrase for comparison *bi women students* ('than us students') goes before the predicate; in (5) *mingtian* ('tomorrow') stands between the subject and the predicate, rather than either in the sentence initial or final position as in English; in (7) the adverbial of time (*in*) *summer* follows the subject rather than being placed in the sentence initial or final position as in English; in (8) the adverbial of place *library-de computer shang* ('on the library's computer') goes before the predicate verb *check*, and the prepositional phrase itself follows the Chinese: noun phrase + preposition; in (10) the adverbial of time *in May* stands between the subject and the verb, another common Chinese word order. These are just some of the common Chinese arrangements of sentential elements, but they are not permitted in English. The Japanese/English ICS examples provide further evidence. Unlike English, Japanese has the SOV order. Once Japanese is chosen as the ML, the Japanese verb final order is always maintained. In (11) the object *sore essay* ('that essay') goes before the predicate verb *wakara-nai* ('not understand'); in (12) the adjective *expensive* goes before copula *desu*; in (13) the object *drug* goes before the verb *ya-tte* ('be doing'); in (14) the object *summer course* goes before the verb *tot-te* ('be taking'); in (15) the adverbial of place *koko ni* ('in here') goes before the verb *stay suru* ('stay').

According to de Bot and Schreuder (1993), bilingual speakers are able to separate the systems of the languages they know and to mix them in a bilingual mode. According to Grosjean (1989, 1997), the amount of language mixing depends on the

language mode the bilingual speaker is currently in, whether monolingual, bilingual, or anywhere else on the language-mode continuum between these two modes. This paper assumes that it is particular levels of activation of the ML and the EL which determine the bilingual speaker's adoption of a particular position on the language-mode continuum. This is because bilingual speakers know that the ML and the EL play unequal roles in ICS. That is, bilingual speakers can clearly "separate" the two language systems and switch items from one language into the other. Bilingual speech is not so-called "mixed" speech but is governed by a set of underlying structural principles, such as those proposed in the MLF model (Myers-Scotton 1993a [1997], 2002, Myers-Scotton and Jake 1995, Wei 2001). Below are more examples of such a language separation.

(Chinese/English)

(16) Ta meitian qu **office** huoshi qu **library**, hen shao
 he everyday go or go very seldom
 zai jia gongzuo.
 PREP/at home work
 'Everyday he either goes to the office or goes to the library, but very seldom works at home.'

(17) Mingtian wo xiang dao **Columbia Mall** qu. Ni qu-bu-qu?
 tomorrow I want PREP/to go you go-not-go
 'I want to go to Columbia Mall tomorrow. Are you going?'

(18) Ni keyi ba zhe-zhang dade **sleeping sofa** fang
 you can PREP/OBJ this-CL big put
 zai xiaode **bedroom** li.
 PREP/LOC small PART/in
 'You can put this big sleeping sofa in the small bedroom.'

(19) Na wo yi dian **come to pick you up**.
 so I one o'clock
 'So, I'll come to pick you up at one o'clock.'

In (16) the verb phrases *qu office* ('go office') and *qu library* ('go library') show that unlike the English maximal category projection of the prepositional phrase where *to* as the head designates GOAL, in Chinese, the ML, GOAL is realized in the verb itself (e.g. *lai* ('come') as in *lai office* ('come office'), and *qu* ('go') as in *qu school*

content morpheme *jiangxuejin* ('scholarship') is switched into the ML syntactic structure. In (21) both the main clause and the embedded clause are from the ML. In (22) the predicate *fabiao wengzhang* ('present papers') goes before the adverbial of place (*at conferences*), which is the ML word order.

(English/Japanese)

(23) It's **totemo muzukashi** to find a convenient and **yasui** apartment here.
very difficult cheap
'It's very difficult to find a convenient and cheap apartment here.'

(24) **Supa** is close from here, but I have to walk **juugo fun gurai**
supermarket fifteen minutes about
to the bus stop.
'The supermarket is close from here, but I have to walk about fifteen minutes to the bust stop.'

In (23) the infinitive clause with the formal subject pronoun *it* is a typical English construction, into which a Japanese adjective phrase is switched. In (24) *walk juugo fun gurai to the bus stop* is the English verb initial construction, but in Japanese the verb final construction is always maintained.

As Grosjean (1989, 1997) claims, the bilingual speaker's language system is organized in two subsets that can be activated and deactivated independently or simultaneously, each to a particular degree. This paper suggests that such an activation or deactivation occurs at the conceptual level (see Figure 1) of the speech production process. If at the conceptual level the speaker decides to choose the bilingual mode, he/she will then choose intersentential or intrasentential codeswitching, and if the latter is chosen, he/she must choose one of the languages as the ML. If the bilingual mode and ICS are chosen, language-specific semantic/pragmatic feature bundles will activate language-specific lemmas in the bilingual mental lexicon at the lemma level. Such activated lemmas will send directions to the FORMULATOR at the functional level for the projection of language-specific morphosyntactic procedures to be realized at the positional level. In other words, once one of the languages is chosen as the ML, it sets up the sentential frame for ICS configurations.

5. Lemma congruence checking as an organizing principle

Lemma congruence between the languages involved in ICS is defined as “a match between the ML and the EL at the lemma level with respect to linguistically relevant features” (Myers-Scotton and Jake 1995:985). As introduced earlier, all lemmas include three levels of abstract lexical structure, which figures in explaining and predicting possible ICS configurations. Green (1986, 1993) assumes that a lemma is activated if it matches part of lexical/conceptual structure created by the CONCEPTUALIZER (see Figure 1) at the level of lexical-conceptual structure. At the level of predicate-argument structure thematic structure is mapped onto grammatical relations. At the level of morphological realization patterns, word order, agreement morphology, case marking, tense/aspect marking, and phonological forms are realized. This paper further claims that in order for possible ICS configurations, lemma congruence checking between the languages involved must come into play, and this checking must take place at each of the levels of abstract lexical structure: lexical-conceptual structure, predicate-argument structure, and morphological realization patterns. Relevant to the current study is lemma congruence checking at the first two levels.

5.1 Lemma congruence for lexical-conceptual structure

As shown in Figure 1, at the conceptual level the speaker makes appropriate choices about the semantic/pragmatic feature bundles (i.e. pieces of information) that he/she intends to convey. However, although, as generally assumed, there is a universal set of semantic/pragmatic features available for the lexical-conceptual structuring of lemmas, there is cross-linguistic variation in the presence and conflation of these features. Thus, the EL lemmas activated by the semantic/pragmatic feature bundles chosen at the conceptual level must be sufficiently congruent with those of the counterparts of the ML in order for ICS to occur. Most ICS instances studied in this paper show sufficient congruence between the EL and ML lemmas which support the existing lexemes in both languages.

As commonly observed, bilingual speakers may switch to another language either intersententially or intrasententially in a discourse. In other words, codeswitching may become part of bilingual speakers’ daily practice or natural speech patterns (Li 1996, Nishimura 1997, Wei 2001, 2002). Grosjean (1982) reports that some codeswitches are motivated by the lack of a particular word in one of the languages or by the greater availability of a word in the other language. Thus, such switches are deliberate or intentional. In a similar way, this paper claims that because the language pairs

involved may differ in their lemmas at the level of lexical-conceptual structure, and it is this type of difference which motivates ICS. In terms of lemma congruence between the languages involved, if such a difference is only partial, there is still sufficient cross-linguistic congruence. This paper suggests that a partial difference at the level of lexical-conceptual structure is one of the major reasons why certain morphemes are switched from the EL as chosen.

(Chinese/English)

(25) Wo xiawu qu jian wode **advisor**. Wo bu neng he ni
 I afternoon go see my I not can with you
 yiqi qu **mall** le.
 together go PART/AFFIRM
 ‘I’m going to see my advisor this afternoon. I can’t go to the mall with you.’

(26) Zhu zai zheili hen fangbian, meitian you **school**
 live PREP/LOC here very convenient everyday have
bus.
 ‘It’s very convenient to live here since there is a school bus every day.’

(27) Wo you liang-fen **paper** mingtian bixu jiaoshangqu, ke
 I have two-CL tomorrow must turn in but
 wo xianzai yi-fen hai mei **finish** ne.
 I now one-CL yet not PART/AFFIRM
 ‘I have two papers [which] I must turn in tomorrow, but at the moment I haven’t finished one yet.’ (= (1))

In (25) an English *advisor* assumes more responsibilities than a Chinese advisor. In the academic setting, an English advisor is a professor or instructor who gives advice or counsel to students regarding their academic progress, improvement, course requirements and sequential arrangements, thesis or dissertation writing, and so on. In addition, most English advisors are those who recommend their students to the job market or professional agencies. Though Chinese has the equivalent word *daoshi* (‘advisor’), it does not necessarily mean that a *daoshi* assumes the same responsibilities as an English advisor. In China only a graduate student may have a *daoshi*, whose only or main responsibility is to guide the student in writing his/her thesis or dissertation. In (26) a *school bus* in English means a bus mainly for transporting students to and from a school. In China the equivalent word *xiaoche* (‘school bus’) usually only transports a school’s sports or performance team or

equipment. In (27) *paper* in English may mean any written piece of work, such as an article, a report, an essay or a composition, but the Chinese equivalent word *zhi* ('paper') itself only means a piece of paper to wrap things up in or to write something on.

(Japanese/English)

(28) Anata wa **registration** o shimashi-ta
 you PART/TOP PART/OBJ do-PAST
 ka?

PART/INTERROG

'Have you done your registration?'

(29) Futatsu no **bedroom** ga ate, hitori, Maria
 two POSS PART/NOM COP one person
 to iu ko wa hitori de **one bedroom**
 and call person PART/TOP one person PREP/by

o mot-te imasu yo.
 PART/OBJ have-PROG AUX PART/AFFIRM

'We have two bedrooms. One person, called Maria, has one bedroom.'

In (28) the speaker switches to *registration* for the possible reason that in Japanese universities/colleges, though students must register for the courses to take, they are not free to select the courses which they are interested in taking. The speaker may choose the English word to express his/her intended meaning more accurately. In (29) the speaker switches to *bedroom* for the possible reason that the concept of *bedroom* is relatively new to Japanese. A traditional Japanese room is often used not only for sleeping but also for eating, studying, entertaining guests, or for other daily family activities.

As observed in the above examples, bilingual speakers may switch to particular lexical items of another language at a certain point during a discourse most probably because of cross-linguistic differences in language-specific lemmas underlying the lexical-conceptual structure. As assumed in this paper, although any language is capable of expressing its speakers' semantic and pragmatic intentions, lexicalization patterns are language-specific (cf. Talmy 1985). As observed by Li (1996), Nishimura (1997), and Wei (2001, 2002), in many cases exact translation across languages is incomplete or inaccurate. This is because language cues may have different values. In the case of ICS, the speaker may ignore the ML cue and switch to an EL lexical item when no ML lexical item is available or sufficient enough to express the speaker's

intended meaning. In other words, when the language cue specifies a particular language at a certain point of bilingual speech production, the lexical item in that language receives activation. de Bot and Schreuder (1993) assume that bilingual speakers are able to separate different language systems and to mix them for their communicative intentions. This paper explains (see Figure 1) that bilingual speakers are able to do so by making language choices in the preverbal message. This means that conceptual information and the language cue must work together in activating appropriate language-specific lemmas in the bilingual mental lexicon to serve bilingual speakers' communicative intentions

5.2 Lemma congruence for predicate-argument structure

The MLF model (Myers-Scotton 1993a [1997], 2002) specifies that one of the languages involved in ICS must be the ML. This is because it is the ML which provides the sentential frame into which the EL content morphemes are switched. Thus, only the ML controls the predicate-arguments structure, and only the ML supplies system morphemes, a subcategorization frame for the verb, and morpheme order. The bilingual lemma activation model adopted in this paper (Wei 2006, see Figure 1) assumes that before morphosyntactic directions are sent to the FORMULATOR, lemmas from both languages can be activated at a certain point during bilingual speech. Thus, lemma congruence checking at the level of lexical-conceptual structure alone is not sufficient enough for ICS to occur. Lemma congruence checking at the level of predicate-argument structure must also come into play.

As observed, the speakers in the Chinese/English ICS tend to use the EL verbs and verb phrases as well as the EL nouns and noun phrases. One of the reasons for this is the fact that Chinese and English share the same basic V-O order. The speaker can easily switch the EL verb/verb phrases or nouns/noun phrases into the syntactic slots prepared by the ML. Most examples discussed in this paper show that there is sufficient congruence between the EL and ML lemmas underlying the realization of an EL content morpheme, either a single verb or a verb with its complement noun.

(Chinese/English)

- (30) Ta gang dao, ta dei xue **drive**.
he just arrive he must learn
'He just arrived and must learn how to drive.'

5.3 Consequences of lemma incongruence

Languages may differ at any of the three levels of abstract lexical structure. The MLF model (Myers-Scotton 1993a [1997], 2002) claims that since not all languages lexicalize concepts in the same way and lexical representations are independent across languages (i.e. lexical representations are language-specific), whenever an EL lemma is selected but does not have a match with that of the ML, some compromise strategy must be taken for possible ICS. In other words, the existence of lack of sufficient congruence between the languages involved in ICS requires some compromise strategies for ICS to occur. One of such strategies is the production of EL islands (Jake and Myers-Scotton 1997, Wei 2001, 2002). An EL island is a constituent in which an EL content morpheme occurs entirely with only other EL morphemes, including EL system morphemes. Such compromise strategies can be taken at the level of lexical-conceptual structure or at the level of predicate-argument structure. This is possible because in producing an EL island, the EL directs the FORMULATOR to activate only the EL morphosyntactic procedures. Lemma incongruence between the language pairs involved in ICS in regard to lexical-conceptual structure and predicate-argument structure is one of the major reasons for EL islands to be produced if the EL lemmas are selected for the speaker's communicative intention.

In Levelt's (1989) model of speech production, the second component, the FORMULATOR, gives language-specific directions. This is because different procedures must be applied to the grammatical and phonological encoding of typologically different languages. This also can be true if the language pairs involved in ICS cannot possibly be encoded by means of the same morphological procedures. Because bilingual speakers have two speech plans available for bilingual speech, they may stop the encoding of one of them and continue with the other so as to solve the problem occurring in the production process. This means that at a certain point of bilingual speech production, for some reason or other, speakers may consider it more appropriate to continue in one of the languages. Different from Levelt's model of speech production is that in ICS the choice of one EL procedure versus another is determined by the larger ML frame.

(Chinese/English)

(35) Name ni mingtian **call me.**

Then you tomorrow

'Then you call me tomorrow.'

(36) Ni neng-bu-neng **give me a ride?**

you can-not-can

‘Can you give me a ride?’

(37) Na wo yi dian **come to pick you up.**

so I one o’clock

‘So, I’ll come to pick you up at one o’clock.’

In (35) the semantic features of “communicate with by telephone” are conflated in the English verb *call*, but the Chinese equivalent of *to call me* is *da dianhua gei wo* (literally translated as ‘make phone to me’), that is, the same meaning is realized by both the verb and its direct object as well as its indirect object. Since the speaker chooses the EL lemma underlying *call me*, the EL lexical-conceptual structure is activated and the verb phrase is accessed as an EL island. In (36) *give me a ride* is incongruent with the ML counterpart *song wo yixia* (literally translated as ‘send me one time’). In the EL the lexical-conceptual structure of the means of transportation is conflated in the noun *ride* as the direct object of the verb, but in the ML it may be conflated in the verb *song* (‘send’) because the verb itself may not contain the meaning of transportation at all. The speaker chooses the EL expression probably because he/she wants to be more specific than he/she can be with the Chinese verb. Thus, when the EL lemma is activated, the whole EL verb phrase is accessed as an EL island. In (37) the speaker chooses the EL lemmas underlying the lexical-conceptual structure of *come to pick you up*, and thus the whole verb phrase with a pronominal object before the particle *up* is accessed. The infinitive *to*, an EL system morpheme, also appears in the island. The speaker prefers *pick up* for the possible reason that this phrasal verb contains the meaning of ‘to take on as a passenger’, but the Chinese equivalent verb *jie* (‘meet’/‘welcome’) may not. It should also be noticed that *come* is accessed together with the infinitive clause *to pick you up* as an EL island. The possible explanation is that the English infinitive marker *to* becomes obligatory if two successive EL verbs are activated and accessed simultaneously. Such cross-linguistic differences in lexical-conceptual structure are also shown in Japanese/English ICS instances.

These instances illustrate why lemma incongruence between the ML and the EL in lexical-conceptual structure of certain lexical items may induce the production of EL islands. However, it needs to be noted that two distinct codes may exist within the same conversation or even within the same sentence (i.e. utterance). If the former occurs, intersentential codeswitching will come into play. This means that the speaker may produce certain sentences in one language but others in another language within

a single conversation. The practice of intersentential codeswitching is commonly observed in fluent bilinguals in particular discourse contexts or speech situation, such as register, field, tenor, and mode (cf. Halliday 1978, Brown and Yule 1983, Coulthard 1985). The study of intersentential codeswitching is beyond the scope of the current paper. As observed in the previous sections, in ICS, it is rather common for the speaker to switch to certain EL items within the same sentence. For example, in (35) it is very possible for the speaker to say “call wo (‘me’)” rather than *call me*. If this happens, no EL island is produced. Similarly, in (36) the speaker may say “give wo (‘me’) a ride”, and in (37) the speaker may say “come to pick ni (‘you’) up”. If this happens, the EL island still exists in terms of both lexical-conceptual structure and predicate-argument structure (see the discussion on incongruence in predicate-argument structure below).

(Japanese/English)

(38) Nan to iu no desu ka, Amerika de
 what PART say NOM COP/BE QUE America PREP/LOC
Christian toka ga ooi desh-oo. Dakara
 so on PART/NOM many COP/be-AFFIRM so
 nanka **before marriage sex** ga dame mitai
 something PART/NOM prohibited like
 da.
 COP/be
 ‘Whatever you say, in America there’re many Christians and other such people,
 so things like before marriage sex is prohibited, or the like.

In (38) the speaker switches to the English lexical item *sex* and also the prepositional phrase *before marriage*, an EL island, where the EL order is observed. The speaker chooses the EL expression probably because most Japanese keep their traditional concept of sex before marriage. In America sex before marriage is not necessarily bounded with marriage, but in Japan it is usually, if not always, so.

In addition to potential lemma incongruence in lexical-conceptual structure across the languages involved in ICS, lemma incongruence in predicate-argument structure also occurs. That is, in some cases, predicate-argument structures across the two languages may differ. If such incongruence occurs, but the semantic/pragmatic feature bundles desired by the speaker activate the EL lemma for his/her communicative intention, a radical compromise strategy must be taken in order for the EL material to be accessed. In other words, even if the lexical-conceptual structures between the two languages are sufficiently congruent, the ML predicate-argument structure will reject

the mapping if a particular EL predicate-argument structure does not match that of the ML.

(Chinese/English)

(39) Tingshuo nei-ge **professor** hen **crazy**. Ta jingchang **fails**
 hear that-CL very she often
students in exams.

‘I heard that professor is very crazy. She often fails students in exams.’

(40) Ni biye hou keyi **teach English to nonnative speakers**.
 you graduate after can

‘After you graduate, you can teach English to nonnative speakers.’

(41) Ta jingchang bangzhu wo **with my computer work**.
 he often help me

‘He often helps me with my computer work.’

In (39) the verb phrase headed by *fail* is an EL island (i.e. with all the morphemes, including the system morphemes, from the EL). In English *fail* can be used as a causative verb and thus takes the grammatical subject as the AGENT who makes the failure happen, but the Chinese equivalent verb *shibai* means ‘be defeated in ...’ and is used only as a noncausative verb and thus takes the grammatical subject as the EXPERIENCER. The speaker prefers the EL concept, but there is incongruence between the EL and the ML in predicate-argument structure. The result is the production of an EL island. (It is possible for the speaker of a particular Chinese variety to say “tingshuo nei-ge **professor hen crazy**. Ta jingchang dang **students**.” In this case, the Chinese verb *dang* is equivalent to the English verb *fail* in terms of the predicate-argument structure. Like *fail*, *dang* is used as a causative verb with the grammatical subject as the AGENT. If the speaker chooses *dang* rather than *fail*, of course, no EL island will be produced. In other words, *shibai* and *dang* are two separate lexical entries.) In (40) the RECIPIENT is introduced by the prepositional phrase headed by *to*, the English indirect object dative construction. By contrast, the equivalent Chinese verb phrase headed by *jiao* (‘teach’) only permits the double object construction (e.g. “jiao ta English” (‘teach him English’)). Again, since the speaker prefers the EL material, but the ML rejects the mapping which the EL preposition phrase would project at the level of predicate-argument structure, the result is the production of an EL island. In (41) the THEME *my computer work* is introduced by the preposition *with* in English, but in Chinese any THEME must be

introduced by the verb itself. Again, the speaker's choice of the EL material is realized as an EL island because of the incongruence between the EL and ML in predicate-argument structure.

6. Bilingual competence and outcomes of ICS

The so-called "classic codeswitching" is defined as codeswitching by bilinguals who have full proficiency in the participating language that becomes the ML in ICS, and who have limited to full proficiency in the other language (Myers-Scotton 2002). Muysken (1995:177) points out that "... code-switching is a quite normal and widespread form of bilingual interaction, requiring a great deal of bilingual competence". The reality is that few bilinguals have equal proficiency in their languages for a variety of reasons. Bilinguals' different linguistic competence, that is, unequal proficiency in the languages they know, may give rise to different types of codeswitching (Appel and Muysken 1987, Muysken 1995). For example, in order to practice intersentential codeswitching, speakers must have full command of the morphosyntax of both languages or a high level of proficiency in both languages. Also, it seems obvious that ICS requires some proficiency in the speaker's other language in terms of both frequency of ICS and various EL items to be switched. As often observed, proficient bilinguals may produce more EL islands than less fluent bilinguals. This is because to produce EL islands formulated in the language which is not the speaker's native language requires the speaker's full knowledge of both the lexical-conceptual structure and the predicate-argument structure (also, morphological realization patterns) of particular lexical items. However, speakers need not have equal proficiency in their both languages to engage in ICS. This is because under some codeswitching models, notably the MLF model, the premise is that languages do not participate equally in ICS. In other words, speakers must be able to produce and recognize well-formed utterances in one of the languages as the ML (i.e. whichever language as the ML), they need not have full command of the morphosyntax of the other language.

In addition to bilinguals' linguistic competence (i.e. bilingual proficiency) in relation to types of codeswitching (which is not the main topic of the current paper), it needs to be mentioned that there must be other factors that may influence speakers' code choices (e.g. bilingual mode vs. monolingual mode, intersentential vs. intrasentential codeswitching, and EL lexical items vs. EL islands (see Figure 1)). Other factors may include topic, setting, and genre (Gumperz 1982), speakers' social motivations (Myers-Scotton 1993b), different contact situations (Muysken 1995), multilingual variables and socio-cultural dynamics (Fishman 2000), marked vs.

unmarked choices (Scotton 1983, Myers-Scotton 2000), among others. Research in such factors can provide insights into grammatical (i.e. morphosyntactic) outcomes of ICS. However, the exploration of such factors is beyond the purpose and scope of this paper.

7. Conclusion

This paper has explored the bilingual speech production process and constraints. The general assumption underlying the study is that the mental lexicon contains not only lexemes but lemmas, which are abstract entries in the mental lexicon containing syntactic, semantic, pragmatic, morphological, and phonological information about the surface realization of actual lexemes. The specific claim is that lemmas in the bilingual mental lexicon are language-specific, language-specific bilingual lemmas are in contact during a discourse involving ICS, and such a contact occurs at three distinct but related levels of abstract lexical structure: lexical-conceptual structure, predicate-argument structure, and morphological realization patterns. Thus, this paper regards ICS as bilingual lemmas in contact at any level of abstract lexical structure. The bilingual lemma activation model presented in this paper describes and explains ICS, including its processes, mechanisms and constraints.

The representative ICS examples from several language pairs discussed in this paper provide sufficient evidence for the arguments and claims about the structural principles governing ICS in relation to the nature and activity of the bilingual mental lexicon. The major research findings indicate that ICS can be better accounted for at a deep or abstract level. The study reaches several conclusions.

First, the bilingual mental lexicon contains lemmas from the languages known, these lemmas are tagged for their specific language, and language-specific lemmas are in contact in ICS. If the speaker chooses the bilingual mode and decides to engage in ICS, he/she may activate the language-specific lemmas as desired from his/her bilingual mental lexicon. However, lemmas are never equally activated in either monolingual or bilingual speech. ICS is no exception. Only EL content morphemes can be activated and switched to express the speaker's intended or desired meaning.

Second, although bilingual speakers' languages are turned "on" during a discourse involving ICS, they are never equally activated at the same time. The ML is more activated than the EL. It is the speaker who chooses whichever language as the ML. It is only the ML which provides the sentential frame, controls morphosyntactic procedures, and provides all system morphemes as well as content morphemes at a much higher frequency. The EL only supplies certain content morphemes as desired by the speaker to be switched into the ML sentential frame. This is because only

conceptually activated EL lemmas in the bilingual mental lexicon can appear in ML + EL constituents.

Third, bilingual speakers can activate lemmas from whichever language as the EL during a discourse involving ICS, but the activated EL lemmas must be sufficiently congruent with the counterparts of the ML at the three levels of abstract lexical structure or some combination of these levels: lexical-conceptual structure, predicate-argument structure, and morphological realization patterns. If lemma incongruence of insufficient congruence occurs between the language pairs at any of these levels, radical compromise strategies, such as production of EL islands, must be taken in order for possible ICS to occur; otherwise, such a bilingual speech will become impossible.

The main purpose of this paper has been to explain the phenomenon of ICS as bilingual lemmas in contact. As assumed, different aspects of abstract lexical structure affect ICS, and ICS, like any natural language speech behavior, is a rule governed bilingual behavior.

References

- Appel, René, and Pieter Muysken. 1987. *Language Contact and Bilingualism*. London: Edward Arnold.
- Berg, Thomas, and Ulrich Schade. 1992. The role of inhibition in a spreading-activation model of language production. *Journal of Psycholinguistic Research* 21:405-462.
- Bock, Kathryn, and Willem J. M. Levelt. 1994. Language production: Grammatical encoding. *Handbook of Psycholinguistics*, ed. by Morton A. Gernsbacher, 945-984. New York: Academic Press.
- Brown, Gillian, and George Yule. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Coulthard, Malcolm. 1985. *An Introduction to Discourse Analysis*. New York: Longman.
- de Bot, Kees, and Robert Schreuder. 1993. Word production and the bilingual lexicon. *The Bilingual Lexicon*, ed. by Robert Schreuder and Bert Weltens, 191-214. Amsterdam: John Benjamins.
- Faerch, Clause, and Gabriele Kasper. 1986. Cognitive dimensions of language transfer. *Crosslinguistic Influence in Second Language Acquisition*, ed. by Eric Kellerman and Michael Sharwood Smith, 49-65. New York: Pergamon Press.

- Fishman, Joshua A. 2000. Who speaks what language to whom and when? *The Bilingual Reader*, ed. by Li Wei, 89-106. London: Routledge.
- Green, David W. 1986. Control, activation and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language* 27:210-223.
- Green, David W. 1993. Toward a model of L2 comprehension and production. *The Bilingual Lexicon*, ed. by Robert Schreuder and Bert Weltens, 249-277. Amsterdam: John Benjamins.
- Grosjean, François. 1982. *Life with Two Languages*. Cambridge, MA: Harvard University Press.
- Grosjean, François. 1985. The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development* 6:467-477.
- Grosjean, François. 1989. Neurolinguistics, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36:3-15.
- Grosjean, François. 1994. Individual bilingualism. *The Encyclopedia of Language and Linguistics*, ed. by Ron E. Asher, 1656-1660. Oxford: Pergamon.
- Grosjean, François. 1997. Processing mixed languages: Issues, findings, and models. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, ed. by Annette M. B. de Groot and Judith F. Kroll, 225-254. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Grosjean, François, and Carlos Soares. 1986. Processing mixed language: Some preliminary findings. *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*, ed. by Jyotsna Vaid, 145-179. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gumperz, John J. 1977. The sociolinguistic significance of conversational code-switching. *RELC Journal* 8:1-34.
- Gumperz, John J. 1982. Conversational codeswitching. *Discourse Strategies*, ed. by John J. Gumperz, 55-99. Cambridge: Cambridge University Press.
- Halliday, Michael A. K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Jake, Janice L. 1994. Intrasentential codeswitching and pronouns: On the categorical status of functional elements. *Linguistics* 32:271-298.
- Jake, Janice L., and Carol Myers-Scotton. 1997. Codeswitching and compromise strategies: Implications for lexical structure. *International Journal of Bilingualism* 1:25-39.
- Levelt, Willem J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: The MIT Press.
- Levelt, Willem J. M. 1995. The ability to speak: From intentions to spoken words. *European Review* 3:13-23.

- Li, David C. S. 1996. *Issues in Bilingualism and Biculturalism: A Hong Kong Case Study*. New York: Peter Lang.
- Li, Wei, and Lesley Milroy. 1995. Conversational codeswitching in a Chinese community in Britain: A sequential analysis. *Journal of Pragmatics* 23:281-299.
- Lipski, John. 1977. Codeswitching and the problem of bilingual competence. *Aspects of Bilingualism*, ed. by Michel Paradis, 250-265. Columbia, SC: Hornbeam Press.
- Milroy, Lesley, and Pieter Muysken (eds.). 1995. *One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching*. Cambridge: Cambridge University Press.
- Muysken, Pieter. 1995. Code-switching and grammatical theory. *One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching*, ed. by Lesley Milroy and Pieter Muysken, 177-198. Cambridge: Cambridge University Press.
- Myers-Scotton, Carol. 1993a [1997]. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.
- Myers-Scotton, Carol. 1993b. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Clarendon Press.
- Myers-Scotton, Carol. 2000. Code-switching as indexical of social negotiations. *The Bilingual Reader*, ed. by Li Wei, 137-165. London: Routledge.
- Myers-Scotton, Carol. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. New York: Oxford University Press.
- Myers-Scotton, Carol, and Janice L. Jake. 1995. Matching lemmas in a bilingual language competence and production model: Evidence from intrasentential codeswitching. *Linguistics* 33:981-1024.
- Myers-Scotton, Carol, and Janice L. Jake. 2000. Explaining aspects of codeswitching and their implications. *One Mind, Two languages: Bilingual Language Processing*, ed. by Janet Nicol, 91-125. Oxford, UK: Blackwell.
- Nishimura, Miwa. 1997. *Japanese/English Code-Switching: Syntax and Pragmatics*. New York: Peter Lang.
- Paradis, Michel. 1987. *The Assessment of Bilingual Aphasia*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Paradis, Michel. 1989. Bilingual and polyglot aphasia. *Handbook of Neuropsychology*, Vol. 2, ed. by Francois Boller and Jordan Grafman, 117-140. Amsterdam: Elsevier.
- Paradis, Michel. 1997. The cognitive neuropsychology of bilingualism. *Tutorials in Bilingualism: Psycholinguistic Perspectives*, ed. by Annette M. B. de Groot and Judith F. Kroll, 331-354. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Pfaff, Carol. 1979. Constituents on language mixing: Intrasentential codeswitching and borrowing in Spanish/English. *Language* 55:291-318.

- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish *y termino en español* [and finish in Spanish]: Toward a typology of code-switching. *Linguistics* 18:581-618.
- Poullisse, Nanda, and Theo Bongaerts. 1994. First language use in second language production. *Applied Linguistics* 15:36-57.
- Roelofs, Ardi. 1992. A spreading activation theory of lemma retrieval in speaking. *Cognition* 42:107-142.
- Romaine, Suzanne. 1995. *Bilingualism*. Cambridge, MA: Basil Blackwell.
- Scotton, Carol. 1983. The negotiation of identities in conversation: A theory of markedness and code choice. *International Journal of the Sociology of Language* 44:115-136.
- Soares, Carlos, and François Grosjean. 1984. Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition* 12:380-386.
- Sridhar, S. N., and K. Sridhar. 1980. The syntax and psycholinguistics of bilingual code-switching. *Canadian Journal of Psychology* 34:7-16.
- Talmy, Leonard. 1985. Lexicalization patterns: Semantic structures in lexical form. *Language Typology and Syntactic Description III: Grammatical Categories and the Lexicon*, ed. by Timothy Shopen, 51-149. New York: Cambridge University Press.
- Wei, Longxing. 1996. Organizing principles behind codeswitching and interlanguage development in early adult second language acquisition. *Sociolinguistic Variation: Data, Theory, and Analysis: Selected Papers from NWA 23 at Stanford*, ed. by Jennifer Arnold, Renée Blake, Brad Davidson, Scott Schwenter, and Julie Solomon, 417-431. Stanford, CA: CSLI Publications.
- Wei, Longxing. 2000a. Unequal election of morphemes in adult second language acquisition. *Applied Linguistics* 21:106-140.
- Wei, Longxing. 2000b. Types of morphemes and their implications for second language morpheme acquisition. *International Journal of Bilingualism* 4:29-43.
- Wei, Longxing. 2001. Lemma congruence checking between languages as an organizing principle in intrasentential codeswitching. *International Journal of Bilingualism* 5:153-173.
- Wei, Longxing. 2002. The bilingual mental lexicon and speech production process. *Brain and Language* 81:691-707.
- Wei, Longxing. 2006. Intrasentential codeswitching as conceptual projection of lemmas in the bilingual mental lexicon. *Journal of Cognitive Science* 6:149-179.
- Woodford, Ellen. 1983. Bilingual codeswitching and syntactic theory. *Linguistic Inquiry* 14:520-536.

[Received 28 August 2008; revised 5 March 2009; accepted 17 June 2009]

Department of Linguistics
Montclair State University
New Jersey, U.S.A.
Longxing Wei: weil@mail.montclair.edu

句內代碼轉換：接觸中的雙語詞義素

魏龍興

蒙特克萊爾州立大學

本文針對在進行句內代碼轉換時雙語腦詞匯的活動與性質加以研究。文章著重提出所謂的腦詞匯不僅僅包含詞位而且包含「詞義素」。詞義素是存在於腦詞匯中的抽象條目。正是這種抽象條目支撐著實際詞位的表象實體。其原因在於這種抽象條目包含對詞位有關的語義、句法、語用以及詞法方面的信息。文章並提出在腦詞匯中的詞義素是各種語言特定的，而各語言特定的詞義素在進行含有句內代碼轉換的談話過程中相接觸。這種詞義素的相接觸出現在抽象的詞匯結構的三個層次：詞匯概念結構、謂語組合結構以及詞法實現形式。各語言特定的詞義素在抽象的詞匯結構中任何一個層次上的激活便會激發句內代碼轉換的產生。本文中討論到的句內代碼轉換的例子證明參與句內代碼轉換的語言的激活程度是不同的。它們的不同激活程度表現在語言特定的詞義素對詞素及詞素句法程序的激活。句內代碼轉換的例子並且證明為使句內代碼轉換的實現有可能，從內嵌語（EL）那裡激活的詞義素必須與母體語（ML）中相應的詞義素等合（Myers-Scotton 1993a [1997]）。

關鍵詞：腦詞匯、詞義素、雙語、內容、系統、詞素、激活、等合