

Pashto-English Bilingual Data: Testing the Diagnostic Features for the Patterns of Codemixing

Arshad Ali Khan
Nadeem Haider Bukhari
Amina Khalid

Abstract

This study examines the patterns of code-mixing in Pashto-English bilingual data. The framework for the present research is based on the diagnostics features for code-mixing proposed by Muysken (2000). In order to investigate the research question, 25 clips of Pashto speech community have been transcribed and analyzed. The diagnostic features: the single constituent, the nested a b a, and the selected position of the bilingual elements determine that insertion is the dominant pattern of CM. The study empirically support that insertion is predicted in colonial settings where there is asymmetry in the bilingual's proficiency, and where one of the languages plays a dominant role. The study reveals that the subject-verb agreement, morpheme order, and the late system morpheme (bridge, outsider) play a crucial role to maintain the morphosyntactic frame of Pashto language. The switched elements in the bare DP and the bilingual VP are content words following the MO and SM principles of the Pashto language.

Keywords: Patterns of code-mixing, Insertion, alternation, congruent lexicalization morphosyntactic constraints

1. Introduction

The language contact has given birth to various phenomena, which include interference, borrowing, convergence, pidginization and code-switching. The focus of the present research is on the alternative use of two languages (Pashto and English) in term of code-mixing which is a prevalent style of the Pashto speech community of Khyber Pakhtunkhwa (KP), where large numbers of morphosyntactic and/or lexical material are found in the Pashto-English bilingual data. The term code-mixing refers to, "all cases where lexical items and grammatical features from two languages appear in one sentence" (Muysken, 2000, p. 1).

A number of researchers have proposed different types of constraints on the basis of empirical data from different varieties of code-mixing in different languages, e.g. Poplack (1980) on Spanish-English, Sridhar and Sridhar (1980) on Kannada-English, etc. These constraints have been proposed by a number of different researchers, some of which are claimed to be language-universal and applicable to typologically diverse varieties of code-mixing, e.g., Matrix Language Frame Model (Myer-Scotton, 1993), Equivalence and Free Morpheme constraints (Poplack, 1980) and the Typological approach to code-mixing (Muysken, 2000). Within the Chomskyans' perspective, Belazi *et al.*, (1994) have come up with the Functional Head Constraint Model of code-mixing. MacSwan (1999, 2000), working within the boundary of syntactic theory, has developed Model of Minimalist constraints on code-mixing.

In order to explore the embedded nature of the Pashto-English bilingual data, the present study is based on Muysken's (2000) Typological Approach to code-mixing (CM). In the typological approach to bilingual data, Muysken (2000) proposed the three patterns of CM: insertion, alternation, and congruent lexicalization. The main objective of the study is to answer the research question: what are the different patterns of code-mixing in the Pashto-English bilingual data? Before going to probe into Pashto-English bilingual data from the Typological Approach (Muysken, 2000), it is requisite to discuss Muysken (2000) Typological Approach to code-mixing.

2. Muysken's (2000) Typological Approach to Code-mixing

Muysken's (2000) work on code-mixing is about intra-sentential level. His book, "The Typology of Code-mixing" (2000), has provided a comprehensive insight into the bilinguals' speech, and their linguistic capacity in language interactions of different languages. Muysken (2000) uses the term code-mixing to all cases where lexical items and grammatical features from two languages appear in one sentence. In his study, he has separated the cases of code-mixing from that of lexical borrowing. He has also opposed the model based on the single lexicon and single grammar in the study of code-mixing. He states that in most models, the function of the speaker and listener is portrayed on the basis of a single grammar and a single lexicon.

Weinreich (1953, P. 73) thought of code-mixing at the intra-sentential level as incompetence, lack of proficiency, and interference of bilinguals. An ideal bilingual 'switches from one language according to appropriate changes in the speech situation (interlocutor, topics, etc.) but not in an unchanged speech situation and certainly not within a single sentence'. In response to this statement, Muysken (2000) replies that bilinguals often produce mixed sentences in ordinary conversation. He says that it has made him curious as to how bilinguals produce such mixed sentences with great ease and complete fluidity. For some speakers, it is an unmarked code in certain circumstances, and neither does it reflect limited proficiency in either of the languages involved (Myers-Scotton, 1993a). Poplack (1980) states that speakers who code-mix fluently and easily have been considered quite efficient bilinguals.

Muysken (2000) has suggested that we cannot assume that either code-mixing is a product of word-finding difficulties or specific cultural pressures (even if language contact itself is culturally conditioned).

In the last two decades, a large number of studies have been carried out at the intra-sentential level in code-mixing phenomenon. In specific cases, code-mixing has been analyzed from a grammatical perspective. It has been found that intra-sentential code switching is not distributed randomly, but it is a rule governed phenomenon that occurs at specific points.

The work of Muysken (2000) is grounded both in structural linguistics and in sociolinguistics. He has analyzed this complex phenomenon of code-mixing in detail with the help of integrating the results of many previous studies. This work can be considered to represent a taxonomic phase in the discipline. The main focus of this work is on grammar and all those factors which influence code-mixing, such as degree of bilingual proficiency, mode of bilingual processing and political balance between the languages, language attitudes, and types of interactive settings.

At the very outset, Muysken (2000) has opposed the single model phenomenon in code-mixing at intra-sentential levels. He says:

I do not propose a single model of code-mixing, since I do not think there is such a model, apart from the general models provided by grammatical theory and language processing. The challenge is to account for the patterns found in terms of general properties of grammar. Notices that only in this way can the phenomena of code-mixing help refine our perspective on general grammatical theory.

(Muysken, 2000, p. 3)

2.1 Different Processes of Code-mixing at Intra-sentential level

In the light of previous research, it has been found that code-mixing has different patterns at intra-sentential levels from one another. Code-mixing is a very complex phenomenon because several distinct processes are at work. The different patterns of code-mixing are:

- **Insertion** of material (lexical items or entire constituents) from one language into a structure from other languages.
- **Alternation** between structure from languages
- **Congruent lexicalization** of material from different lexical inventories into a shared grammatical structure.

These three basic processes are constrained by different structural conditions and are operant to a different extent and in different ways in specific bilingual settings. This produces much of the variation in mixing patterns encountered. The three processes correspond to dominant models for code-mixing that have been proposed.

The three processes are the dominant model for code-mixing patterns. Muysken (2000) has derived these three processes from previous research. He has given a sequence to these three patterns according to its grammatical perspective and social setting.

2.1.1 Features of the Three Code-mixing Patterns

2.1.1.1 The Pattern of Insertion

The pattern of insertion, which is akin to borrowing, is based on the framework of the MLF model (Myers-Scotton, 1993b). In this pattern, the process of code-mixing takes place at lexical, or phrase level, e.g., noun or noun phrases.

Muysken (2000) has analyzed different features of insertions. He has discussed in detail the constituent structure, which is a central notion in the analysis of a sentence. In code-mixing, what is inserted is constituent. It has been found that in a single sentence (base) structure, it is not only the single element which is inserted, but sometimes several constituent elements can be in numbers, in a matrix language. Muysken (2000) has called it multiple contiguous insertions.

The insertion of single constituent into a matrix language frame model:

1. Kalau dong bukan dong bukan bikinWhen they always make they always
makeVoor acht personen dek orang cuma nganga dong makanFor eight persons

and then people only look they eat ‘When they [cook], it is always for eight people, and then they only look at it, they eat ...’

(Huma 1992; cited in Muysken, 2000, p. 4)

In (1) the above example, the inserted constituent is from Dutch (prepositional phrase) language into a Moluccan Malay sentence. It means the base language is Malay and the embedding language is Dutch.

(2) na ‘iish-crash la
 Isg:pass out-crash EMPH
 ‘I am about to pass out.’

(Canfield, 1980, P. 219)

In the above example, it is the English verb stem that is inserted in a complex Navajo verbal structure.

(3) Yo anduve *in a state of shock* por dos dias.
 ‘I walked in a state of shock for two days.’

(Pfaff, 1979, P. 296)

In instance (3), it is the Spanish language which has provided base to English inserted ‘*in a state of shock*’.

In the above three different instances from Muysken (2000), it has been proved that at insertion level we have two types of languages, Matrix (base) language and embedding language. Such a relationship of embedding and matrix language is also known as hybridization. The phenomenon of loanwords and borrowing is also akin to insertion. It is this pattern of code-mixing which paved the way to borrowing, or loanwords. Any foreign word in a language is first introduced through code-mixing, and with the passage of time it becomes a borrowed word and an integral part of that language.

2.2 The pattern of alternation

Muysken (2000) has taken the pattern of alternation from the work of Poplack (1980). In this level he has focused on the Free Morpheme and Equivalence Morpheme constraints. Muysken (2000) has differentiated the level of alternation in the light of Auer’s distinction between code-switching and transfer (1995: 126).

The pattern of alternation in code-mixing is taking place at intra-sentential level. In alternation, one language is replaced in a sentence half way by the other. Often the two languages function at clause level.

(4) maar ‘t hoeft niet li-‘ *anna ida seft ana* . . .
 But it need not for when I-see I
 ‘but it need not be for when I see, I . . .’

(Notier, 1990, P. 126; cited in Muysken, 2000, p. 5)

- (5) Les femmes et le vin, *ne ponimayu..*
 ‘Women and wine, I don’t understand.’
 (Timm, 1978, P. 312; cited in Muysken, 2000, p. 5)
- (6) Andale pues *and do come again.*
 ‘That’s all right then, and do come again.’
 (Gumperz & Hernandez-Chavez 1971: 118; cited in Muysken, 2000, p. 5)

In this type of code-mixing the switch is taking place between two languages in which both grammar and lexicon are involved. In examples (4), (5) and (6), it is evident that different languages with different structure and lexicon work at intra-sentential level. Example (6) shows that the first segment is embedded in Spanish language and the second segment in English language.

2.3 The Pattern of Congruent Lexicalization

This pattern of code-mixing can be found in the studies of Labov (1972) and Trudgill (1986). The focus of this study lies on style shifting and dialect/ standard variation. In typology of code-mixing, Muysken (2000) has analyzed the data from previous research and established a link in the three different approaches to different extra-linguistic phenomena. His main finding is based on intra-sentential level. In intra-sentential level, the pattern of alternation is between languages, insertion into a base or matrix language, and congruent lexicalization in the code switching data.

In this pattern of code-mixing, there is largely a shared structure between the two languages which is lexicalized by elements from either language (Muysken, 2000: 5). Consider the following examples:

- (7) Weet jij [*waar*] jenny is?
 ‘Do you know where jenny is? (Dutch: waar jenny is)
 (Crama and van Gelderen, 1984; cited in Muysken, 2000, p. 5)

Example (7) shows congruency in the English word *where* and Dutch *waar*, jenny is a name in both languages, and *is* is *homophonous*.

- (8) (A) why make Carol *sentarse atras* (B) *pa que* everybody
 sit at the back so that has to move (C) *pa que sesalga*
 So that [she] may get out.
 (Poplack, 1980, p. 589; cited in Muysken, 2000, p. 6)

In (8), the sentence fragments (B) is complement to (A), and fragment (C) is complement to (B). The first sentence fragment is that of Spanish which contains a verb phrase, *sentarse atras* and a purposive complementizer, *pa que*. There is no particular grammatical relation between the two fragments English and Spanish as in (8). It does not mean that they obey the rule specific to the matrix constituent, but rather rules are common to both languages.

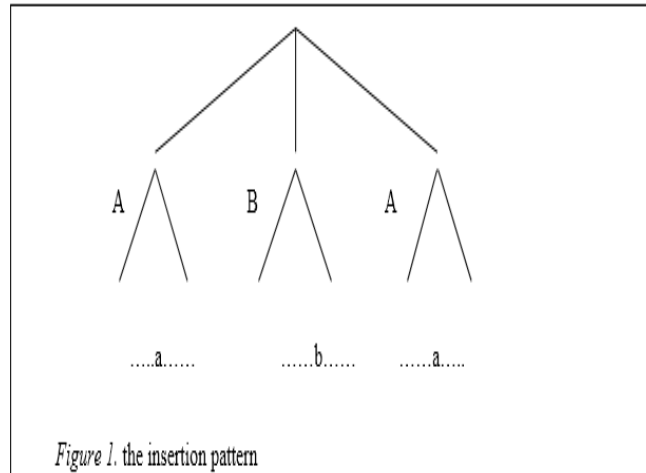
So far, the two notions of insertion and alternation of CM patterns are considered, both of them take place in a sentence at different intervals and often do not share the same structure. In congruent

lexicalization, the two languages often share the structure, and the elements from the two languages are inserted as constituents, or as words into a shared structure.

2.4 The Structural Interpretation of the Three Patterns

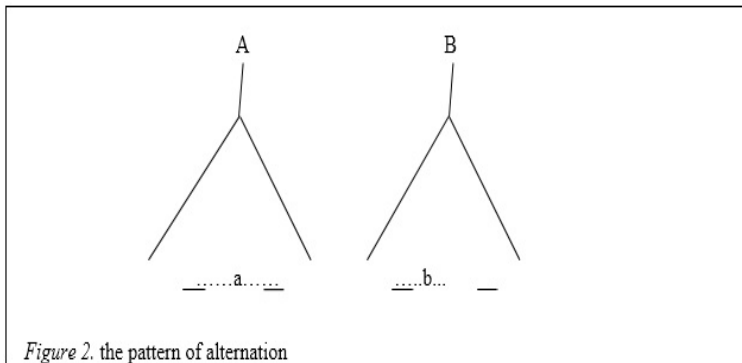
The structural interpretation of the three patterns of code-mixing by Muysken (2000) is as follows. In the following trees, *A* and *B* are languages, *a* and *b* are labels for the terminal, i.e., lexical, nodes, indicating that the words chosen are from a particular language.

The pattern of insertion (Muysken, 2000, p. 7) is also assumed by the Matrix Language Frame Model (MLF) theory proposed by Myers-Scotton (1993)



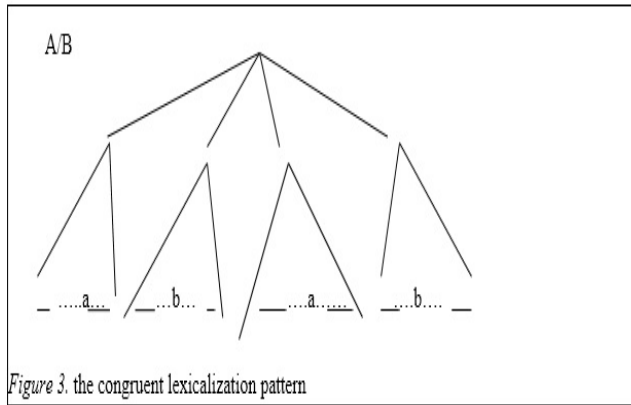
In *figure 1* the insertion pattern, one language (*A*) determines the overall structure into which constituent *B* (with words *b* from the same language), and words from that language *A* are inserted.

The pattern of alternation (Muysken, 2000, p. 7) is also assumed by Poplack (1980).



In *Figure 2*, the two languages *A* and *B* are working at alternation, each having its own structure. In this situation, a constituent from language *A* (with words from the same language) is followed

by a constituent from language *B* (with words from that language). Neither of the language provides over all structure for the utterance. The congruent lexicalization pattern (Muysken, 2000, p. 8) can also be found in the study of Labov (1972) and Trudgill (1986).



In Figure 3, the two different languages *A* and *B* share the same grammatical structure, and words *a* and *b* from both languages are inserted randomly.

Table 2.1: **The relationship between code-mixing patterns and extra linguistic factors (Muysken, 2000)**

code-mixing	Linguistic factors favoring this patterns	Extra linguistic factors favoring this patterns
Insertion	Typological distance	Colonial settings, Recent migrant communities Asymmetry in speaker's proficiency in two languages
Alternation	Typological distance	Stable bilingual communities tradition of language separation
Congruent lexicalization	Typologically similar languages	Two languages have roughly equal prestige, No tradition of overt language separation

The above three patterns of code-mixing can give sociolinguistic meanings as bilingual communicative strategies. Insertion type of code-mixing can be bilingual strategies in colonial settings and recent migrant communities. Muysken (2000) argues that there will be a considerable asymmetry in the speaker's proficiency in the two languages. In migrant communities, language shift towards the dominant language can be seen in the third generation. Insertion is taking place

in the host language of the country. The process of alternation is taking place in stable bilingual communities with a tradition of language separation. In such setting, often the languages claim the same prestige. The third process, congruent lexicalization, may be associated with the second generation. The two languages in contact frequently have the same structure and claim the same prestige.

3. Predicting the CM Patterns

The pattern of CM is based on linguistic and extra linguistic factors (Muysken, 2000). He proposed that the difference between the three patterns, insertion, alternation and congruent lexicalization, is favored by the typological distance and likeness between the languages in contact. The insertion and alternation is favored between typological distance and extra linguistic factors. The insertion is predicted in colonial settings where there is asymmetry in the bilingual's proficiency, and one of the languages plays a dominant role. Alternation is predicted in more stable bilingual communities. Both the languages involved in the present study have a distinct word order, with the colonial setting predicting insertion as the dominant pattern of CM.

Before going on to diagnose the Pashto-English CM data for the dominant pattern, it is important to establish the definition of *constituent structure* in the present data. Defining the constituent structure, Muysken (2000: 61) writes: "a constituent can be a syntactic unit, either a lexical item (e.g. a noun) or a Phrase (e.g. a prepositional phrase)". In a different work, the embedded islands are defined as: the embedded islands are fixed expressions, idioms, formulaic expressions, chunks, and now also known as composite expressions (Muysken, 2000; Myers-Scotton, 2002; Backus, 1999b). In the Pashto-English CM data, the embedded islands, or single elements, are working as single units in the morphosyntactic frame of Pashto (Khan, 2014).

3.1 Research design

This study is descriptive in nature, including qualitative methods to investigate morphosyntactic constraints in Pashto-English code switching.

3.1.1 Sample description

The collected data covers television shows of a Pashto language TV channel AVT Khyber, i.e. Khyber News, Khyber Watch and Khyber club. The data covers a wide range of topics i.e. political, social, cultural, pedagogical, and women's rights. Two different formats have been used in these clips; some are group discussions about a certain topic and some of the clips are based on face to face interviews with participants.

Most of the participants are highly educated and they shared the same cultural norms of the Pashto speech community. The data collected from YouTube are categorized as in Table 1.

Table 3.1: **Topic, gender, and time of TV fragments**

Clip #	Khyber Khyber	Watch/ Topic	Participants (M = male; F = female)	Duration in minutes
1	Khyber club EP# 1	Computer	M: 3; F: 1	14
2	Khyber club EP# 11	Electronic/ print media and its influence on our life	M: 3; F: 1	14
3	Khyber club EP# 16	The use of tuition in education	M: 3; F: 1	14
4	Khyber club EP# 16	The use of tuition in education	M: 3; F: 1	12
5	Khyber club EP# 14	Peace, prosperity and development	M: 4	14
6	Khyber club EP# 11	Electronic media	M: 3; F: 1	14
7	Khyber club EP# 13	Electronic media	M: 3; F: 1	14
8	Khyber News	My right (women issues)	F: 2	8
9	Khyber News	My right (women issues)	F: 3	8
10	Khyber News	My right (women issues)	F: 3	14
11	Khyber News	Ismail (Singing)	M: 2; F: 1	19
12	Khyber News	Personal political views	M: 2; F: 1	14
13	Khyber News	Personal political views	M: 2; F: 1	14
14	Khyber News	Personal political views	M: 2; F: 1	14
15	Nun Sahar	(Education and progress) VC Sarahad University	M: 2	52
16	Nun Sahar	Self and Reality	M: 1; F: 1	14
17	Nun Sahar	Self and Reality	M: 1; F: 1	14
18	Khyber Club	Politics: Care taker Prime minister	M: 4	14
19	Nun Sahar	Unemployment	M: 4	9
20	Nun Sahar	Unemployment	M: 4	14
21	Nun Sahar Classics	Motorway Issues	M: 1; F: 1	39
22	Nun Sahar Classics	Islambad police	M: 1; F: 1	27
23	Khpal Etwar	Doctor opinion	M: 1; F: 2	27
24	Khpal Etwar	Doctor opinion	M: 1; F: 2	21
25	Khpal Etwar	Female education	F: 3	14

3.2 Data collection Procedure

Approximately 12 hrs of 35 clips were downloaded from the online source (YouTube) for the analysis of the CM data. Out of 35 clips only 25 clips, a total of approximately eight/8 hrs were selected. In order to select reliable and appropriate data the clips were selected by six native speakers of Pashto language.

The first criteria of the selection of the different shows were their language and participants. Most of the participants were educated and they were using English words/phrases with ease and spontaneously. The second criteria of selection of the online data were their topics. Their topics were about the culture, politics, unemployment, education and women rights in Pashto speech community of KP. Keeping the notion of “observer’s paradox” (Labov, 1972) in mind the data in the present research has been downloaded in a very natural setting. There was no such risk that the

participants in shows felt observed for the research purposes, and their speech style remains unchanged throughout the debate held in the shows.

3.3 Data Coding

In order to transcribe English-Pashto CS data, the recordings were transcribed into Roman English. In order to give it proper glossing and morphemic identification, the data were transcribed with the help of *Tool Box*. All data have been transcribed in three layers. The first layer represents the data at the morphemic level. The second layer represents the data at the gloss level, and translation at the third level. It is found that the data were quite homogenous.

3.4. The Proposed Diagnostic Features for the Pattern of Insertion

The roadmap is set for the insertion as the dominant pattern of CM, as linguistically, English (SVO) and Pashto (SOV) are typologically distance languages; the setting of the speakers are colonial, with asymmetry in the speaker's proficiency, and Pashto is a dominant language of the KP speech community. In such settings the pattern of insertion is plausible.

The following table (3) shows the diagnostic features proposed by Muysken (2000) for the three patterns of CM. The features are distributed according to its type of constituency, nature of the element switched, the switch site, and the properties of the switch (Muysken, 2000). The study discusses and analyzes the transcribed data with the help of the diagnostic features according to the constituency type, nature, and its switch site for the pattern of CM.

Table 3.2: Muysken's diagnostic features for the patterns of CM

Patterns	Constituency	Nature of the element switched	The switch site	properties of the switch
Insertion	1. Single constituent 2. Nested aba	3. Content words 4. Selected element	5. Dummy word insertion	Telegraphic Mixing Morphological integration
Alternation	1. Several constituents 2. Non-nested aba	3. long constituent 4. complex constituent 5. adverb, conjunction	6. emblematic or tag 7. major clause boundary 8. peripheral 9. embedding in discourse 10. flagging 11. bidirectional switch 12. linear equivalence	13. doubling 14. self-correction
Congruent Lexicalization	1. non-constituent 2. non-nested aba	3. function word 4. selected element	5. bi-directional switching 6. linear equivalence	7. morphological integration 8. homophonous dimorphs 9. triggering 10. mixed collocations

The diagnostic features in table (2) have been developed after studying the works of Poplack (1980), Muysken (1994), and Nortier (1990), for the pattern of insertion. The pattern of alternation is based on Treffers-Daller's work (1994). The pattern of congruent lexicalization is based on Giesbers (1989). In order to determine the pattern of CM, the following bilingual Pashto-English data (9 to 29) is analyzed with the help of the proposed set of diagnostic features.

3.5 Data Analysis and Discussion

3.5.1 Patterns of CM for Nouns

The study shows that 57% of nouns are the dominant pattern of insertion in the Pashto-English bilingual data.

3.5.2 Nouns in Bare DP Constructions

The EL bare noun 'youth' as single constituent has been inserted in the abstract DP construction; it does not pose any threat to the ML and the sentence is well-formed, as in example (9). The second embedded lexical verb 'target' in the bilingual VP is making a single constituent inflected by the Pashto auxiliary *kaw -u* for tense and agreement. The switched elements are in the nested *aba* structure following the insertion patterns. The embedded noun 'youth' is a selected element making a core argument with its verbal predicate. The switched elements in the bare DP and the bilingual VP are content words following the MO and SM principle of the Pashto language. The diagnostic features show that insertion is the dominant pattern.

9.

pa de ke monga [IP[NP *youth*] [V *target kaw -u*]]
 at MED.PROX in.OBL POSS.1PL.NOM youth target do.PRS.IPFV -1PL
 'In this we target the youth.'

IP-internal [[NP][V]] a. NP-internal [N] b. V-internal [V light verb]

In (10), the EL bare noun 'cassette' as single constituent has been inserted in the morphosyntactic frame of Pashto. The second embedded lexical verb 'release', in the bilingual VP, is making a single constituent inflected by the Pashto passive auxiliary *shw -a* for the tense and agreement. The switched elements are in the nested *aba* structure, as the preceding *che*, and the following bilingual VP, are clearly parts of the same clause. The embedded noun 'cassette' is the selected element in the passive construction, making a core argument with its verbal predicate in the object position. The switched elements in the bare DP and the bilingual VP are content words following the MO and SM principles of the Pashto language. The diagnostic features show that insertion is the dominant pattern.

10.

Halaq way -i Che [IP [DP [NP *cassette*][release shw -a]]
 People say.PRS.PFV COMP cassette release COP.PRS.PFV -3SG
 'The people say that the cassette has been released.'

IP-internal [[NP][I]] a. NP-internal [N] b. V-internal [V light verb]

3.5.3 Nouns in Determiner Phrase Constructions

The EL noun 'issue' is making a single constituent in the determiner construction of the morphosyntactic frame of Pashto as shown in (11). The second embedded adjective 'sensitive' in the predicate position has been marked with the Pashto intensifier *daira* 'very' which is modifying the bilingual NP. The switched elements are in the nested aba structure as the preceding *che*, and the following COP *da* are clearly parts of the same Pashto clause. The bilingual NP is a selected element functioning in a core argument with its verbal predicate in the subject position. The switched elements in the bilingual NP and the modifier in the predicate positions are content words following the MO and SM principles of the Pashto language. The diagnostic features: the single constituent, the nested a b a, and the selected position of the bilingual elements determine that insertion is the dominant pattern of CM.

11.

Che [IP [NP daa *issue*] [AP *daira sensitive*]] da]
 COMP this.MED.PROX issue very sensitive COP.PRS.IPFV.F.3SG
 'That this issue is very sensitive.'

IP-internal [[NP] I] a. NP-internal [DETN] b. VP-internal [AP]

3.5.4 Nouns in Prepositional Phrase Constructions

In (12), the two EL nouns 'show' in the PP phrase and 'call' in the bare DP are making single constituent in the morphosyntactic frame of Pashto. The switched elements are in the nested aba structure as the preceding Ergative subject and the following transitive auxiliary are clearly parts of the same Pashto clause. The embedded nouns 'show' in the bilingual PP and 'call' in the direct object position are the selected elements functioning in a core argument with its verbal predicate. The switched elements in the bilingual PP and the embedded noun "call" in the object positions are content words. The diagnostic features: the single constituent, the nested a b a, the selected elements and their property as content words are ample evidence to determine that insertion is the dominant pattern of CM. The embedded element 'show' is a typical selected element in the oblique position. The oblique position is always marked as indirect object or complement.

12.

taso mong ta [IP[PP pa *show* ke] [NP *call*] [V wa -k -u]]
 2PL.ERG POSS.1PL to at show in.OBL call PFV- do.PST -3SG
 'You called us in the show.'

IP-internal [[NP] I] a.NP-internal [N] IP-internal [[PP] V] a.PP-internal [P N P]

3.5.5 Nouns in the Possessive Constructions (də)

The two EL nouns 'tuition' and 'concept' in the possessive construction are marked by the possessive marker *də* as single constituent in the morphosyntactic frame of Pashto, as shown in (13). The switched elements are in the nested aba structure as the preceding subject *zə* 'I' (1SG) and Pashto fragment *helaf zaka* are clearly parts of the same Pashto clause. The embedded nouns in the possessive construction in the complement position are the selected elements functioning in a core argument with its verbal predicate. The switched elements in the bilingual PP in the complement position are content words. The diagnostic features: the single constituent, the nested

a b a, the selected elements and their property as content words are ample confirmations to determine that insertion is the dominant pattern of CM.

13.

zə [IP [PP da *tuition* da *concept*]]helaf zeka [V yu -m]
 1SG.NOM of tuition of concept against because COP.PRS.IPFV -1SG
 ‘That is why I am against the concept of tuition.’

IP-internal [[PP] V]

a.PP-internal [P N P N]

3.6 EL Adjective in ML Constructions

The adjective ‘proper’ in attributive possession in the bilingual NP is preceded by the Pashto indefinite article *yaw* and is followed by its head noun *jang*, ‘fight’, that make a single constituent in the morphosyntactic frame of Pashto, as shown in (14). The switched elements are in the nested aba structure as they are preceded and followed by the fragments of the same grammar. The embedded adjective ‘proper’ in the attributive position modifying the noun in the object position is a selected element. The switched element in the bilingual NP in the attributive position is a content word. The diagnostic features: the single constituent, the nested a b a, the selected element and their property as content word is a clear indication that insertion is the dominant pattern of CM.

14.

us ra-ze [IP [NP yaw *proper* jang] [V kaw -u]]
 now 1PL.come one.M.SG proper fight do.PRS.IPFV.1PL
 ‘Now let’s come to start a proper fight.’

IP-internal [[NP] V]

a. NP-internal [DET A N]

3.7 EL Adverb in ML Constructions

In example (15), the EL adverb ‘either’ is functioning as tag-switching (Poplack, 1980) to the Pashto clause. Such a tag-switch functions as adjoining, not as insertion, with the Pashto language frame. The structural position of such switching is very important to determine its pattern: If the switch takes place at a major clause boundary, alternation is a plausible option (Muysken, 2000). Muysken (2000) has explained the pattern of alternation in terms of Peripherality: the distinction between clause-central and clause-peripheral code-mixing. According to its clause-peripheral position and function as tag-switching, the adverb ‘either’ clearly qualifies the pattern of alternation.

15.

[AP *Either*]k pa kali ke da aghwi [PP pa kor ke] [V na wi]
 Either if at village in.OBL of them.DIST at home in.OBL not- do.PRS.IPFV.3SG
 ‘Either if in the village it is not in their home then.’

AP-internal [A]

In example (16), the EL adverb ‘somehow’ is functioning as tag-switching (Poplack, 1980) to the Pashto clause. Such tags are clause-peripheral and often take switch at the major clause boundary. The structural position of such switching is very important to determine its pattern. The adverb

'somehow' is modifying the whole sentence and is known as sentence adverbial. According to its clause-peripheral position and function as tag-switching, the adverb 'somehow' clearly qualifies the pattern of alternation.

16.

[AP*Somehow*] de sa [VP [NP *serious* [V na day]
Somehow DEM.PROX.M.3SG some serious not COP.PRS.IPFV.M.3SG
'Somehow he is not serious.'

a. AP-internal [A]

The two examples, (16) and (17), where the adverbs are at clause-peripheral position, clearly qualify the pattern of alternation. In (17), the clause-central embedded adverb 'repeatedly' in the predicate position is modifying the verb *kaw -o*. It is a single constituent and is following the nested a b a structure of the Pashto language. The adverbial modifier is not only inserted syntactically but also pragmatically. Thus follows the reduplication strategy of the Pashto language.

17.

Monga [IP [AP *repeatedly repeatedly*] da habara [kaw
-o]]
1PL.NOM repeatedly repeatedly DEM.PROX talk do.PRS.IPFV-1PL
'We are talking repeatedly.'

IP-internal [[AP] V] a. AP-internal [A]

3.8 Embedded Element + Light verb Constructions

In example (18), the EL root 'conclude' in the bilingual VP is inflected by the Pashto auxiliary *kaw* for tense and aspect. The embedded root inflected with the transitive auxiliary makes a single constituent VP which is a plausible example of insertion. The switched element is in the nested aba structure. In the bilingual VP the EL root is suffixed by the auxiliary verb *kaw -am* and is preceded by the Pashto object *habara* which are the fragments of the same clause. It is a selected element functioning as VP in the predicate position. The switched element in the bilingual VP is an example of content word. The diagnostic features: the single constituent, the nested a b a, the selected elements and their property as content word demonstrate that insertion is the dominant pattern of CM.

18.

za ba da habare [V *conclude* kaw -am]
1SG.NOM CL.FUT DEM.PROX discussions.NOM conclude do.PRS.PFV -1SG
'I will conclude this discussion.'

V-internal [V light verb]

In example (19), the EL root 'start' in the bilingual VP is inflected by the Pashto auxiliary *kaw -o* for tense, aspect and subject-verb agreement. The embedded root inflected with the transitive auxiliary, making a single constituent VP which is a plausible example of insertion. The EL word

is a selected element functioning as VP in the predicate position. The switched element in the bilingual VP is a clear example of a content word. The diagnostic features: the single constituent, the nested a b a, the selected element and the EL property as content word support that insertion is the dominant pattern of CM.

19.

...che [IP[PP da *deformat*] ba anga [V *start* kaw -o]]
 ...COMP of DEM.PROX format CL.FUT how start do.PRE.IMP-1PL
 'You should say to the dear audience that how we will start this format.'

IP-internal [[PP] [V]] a. PP-internal [P DET N] b.V-internal [V light verb]

In the above examples (9 to 19), it can be seen that for the single embedded elements, the pattern of insertion is the dominant pattern in the Pashto-English bilingual data. The diagnostic features proposed in Muysken (2000) are compatible with the present study. In the entire single element embedding data, the structural position plays a vital role to determine the nature of the pattern. In case of adverbs switched at a major boundary (outside) of the clause determine alternation as plausible pattern.

3.9 Embedded Language Islands in Pashto-English Code-mixing

In the following examples (20 - 29), the EL Islands in the different constructions of the Pashto morphosyntactic are discussed. The examples are tested for the pattern of CM with the help of diagnostic features proposed in Muysken (2000).

The EL NP 'check and balance' is a fixed expression (collocation) and inserted as single constituent in the ML frame as shown in (20). The switched NP is in the nested aba structure following the insertion patterns. The embedded NP in the Pashto frame is a selected element functioning in the complement position to its verbal predicate. The switched elements are content words. The diagnostic features show that insertion is the dominant pattern.

20.

Che pa de bande [IP[NP *check and balance*][V wasat -u]]
 COMP at CL on check and balance keep.PRS.IPFV -2PL
 'That we should keep check and balance on it.'

IP-internal [[NP] I] a. NP-internal [N]

In (21), the EL island 'live show' as fixed expression is inserted in the morphosyntactic frame of the Pashto language. The adjective 'live' collocates with the different nouns depending upon special occasions, context, and time. The very frequent uses of 'live' as attributive adjective in the Pashto frame are 'live call', 'live match', 'live show', and 'live program, etc. All the basic diagnostic features, single constituent, nested a b a, selected element and content word are applicable to the collocation 'live show'. This determines that the dominant pattern of CM is insertion.

21.

da [IP [NP *live* *show*] [V da]]
 DM.PROX live show COP.PRS.IPFV.F.3SG
 'This is a live show.'

the string of meaningful words. The information about tea is triggered by the meaningless word ‘mubə’.

In a conversation if someone asks:

24. obə mubə shta
Water mater (do) exist.PRS.IPFV.3SG
‘Do you have water mater.’

aw chy shta
yes, tea exist.PRS.IPFV.3SG
‘Yes, tea is available.’

The following example is a typical example of the pattern of insertion. The word is not only following the Pashto strategy but also fulfills other diagnostic features of the nested a b a, selected position and content word.

25.
[PP pa colleges ke] [IP[NP concept muncept] [V na clear keg
i]]
at colleges in.OBL concept muncept no clear become.PRS.IPFV -3SG
‘In college the concept will not going to clear.’

PP-internal [P N P] b.VP-internal [IP [NP][I]] a.NP-internal [N] b.V-
internal [V light verb]

In (26), the bilingual NP ‘golden rules’ is a fixed expression (collocation) and is making a single constituent. In bilingual NP, the adjective ‘golden’ in the attributive position is modifying the head noun ‘rules’. It is in the nested aba structure as it is preceded and followed by the fragments of the same grammar. In the subject position it is a selected element and is making a core argument with the verbal predicate. The switched element in the bilingual NP in the subject position is a content word. The diagnostic features: the single constituent, the nested a b a, the selected element and its property as content word is a clear indication that insertion is the dominant pattern of CM.

26.
də Islam [IP [NP toul golden rules] ba halta [V apply wi]]
of Islam all.DET golden rules CL.FUT DEM.DIST apply COP.PRS.IPFV -3PL
‘All golden rules of Islam will be applied there.’

IP-internal [[NP] [I]] a. NP-internal [DET AN] b. V-internal [V light verb]

In (27), the embedded island ‘previous caller’, the adjective ‘previous’ in attributive position is modifying the head ‘caller’ as fixed expression and making a single constituent. The embedded elements are triggered in different constructions of the Pashto frame, and insertion is the plausible pattern. The switched constituent in the possessive construction is in the nested aba. It is preceded and followed by the grammar of the same language. The switched NP in the possessive construction is a clear example of the content word. The bilingual PP ‘viewer sara’ (with viewer)

is also following the Pashto morpheme order and is a clear pattern of insertion. Similarly, the English root ‘share’ in the bilingual VP is inflected by Pashto auxiliary. The diagnostic features: the single constituent, the nested a b a, and the selected element, demonstrate that insertion is the dominant pattern of CM.

27.

[IP[PP də *previous caller* habara] za[PP *viewer* sra] [V *share* kaw -am]]
of previous caller talk 1SG.NOM viewer with.OBLsharePRS.IPFV -1SG
‘I will share the point of view of the previous caller with the audience.’

IP-internal [[PP] [PP][I]] a. PP-internal [P AN] b. PP-internal [N P]
c. V-internal [V light verb]

In (28), there are two embedded elements in the morphosyntactic frame of the Pashto language. The gerund ‘blackmailing’ is used as a noun, and the embedded PP ‘in a sense’ is used as an idiom functioning as adjunct in the code-mixing sentence. If we apply the term of Peripherality, the distinction between clause-central and clause-peripheral, then the embedding gerund is a plausible example of alternation but it is an integral part of the Pashto clause and is functioning as core argument with the verbal predicate.

The second EL elements, the embedded PP, are a single constituent and functions as idiom in the Pashto clause. In the nested a b a position, it is preceded and followed by Pashto grammar. It is not a selected element and is functioning as adjunct in the clause. Most of the diagnostic features confirm that the embedded idiom is a plausible example of insertion.

28.

[IP[NP *Blackmailing*] hu [PP *in asense*] [TP war -ta]na -shay [way -əl -e]]
blackmailing indeed inasense 2SG-to not- PRS.PFV say -2PL
‘In one sense indeed it cannot be called blackmailing.’

IP-internal [[NP] I] a.NP-internal [N]
PP-internal [NP]

29.

...che [IP [PP *as a profession*] daa də cha [V yi]]
...COMP as a profession DEM.PROX of who COP.PRS.PFV.3SG
‘If as profession someone has it.’

IP-internal [[PP] V] a.PP-internal [ADETN]

In example (29), the embedded element ‘as a profession’ is modifying the demonstrative *daa* which refers to the topic in the discourse. The embedded PP ‘as a profession’ is a modifier and functions as adjunct in the bilingual clause. The embedded modifier ‘as a profession’ is single constituent functioning as single unit. It is in the nested a b a construction preceded by the Pashto complementizer *che* and followed by the Pashto *da*, which are the fragments of the same clause. The bilingual NP, the noun ‘profession’ determines it as content element. The diagnostic features

demonstrate that the dominant pattern of CM is insertion.

4. Conclusion

The Pashto-English bilingual data show that Pashto is the ML and English is the embedded language. The diagnostic features for the pattern of code-mixing are qualitatively used and discussed with the help of different transcribed data. The result and discussion shows that the code-mixing is working under set rules and patterns. In the three patterns of code-mixing insertion is the dominant pattern. The diagnostic features proposed in Muysken (2000) are compatible with the present study.

In the data, different types of constituents: single elements, fixed expressions, prepositional phrases and idiomatic expressions are diagnosed, and it is found that insertion is the dominant pattern in all types of embedded elements found in the ML frame of the Pashto language. According to its clause-peripheral position and function as tag-switching, few examples of the patterns of alternation have been recorded such as the adverb 'either' clearly qualifies the pattern of alternation. The embedded root inflected with the transitive auxiliary in the bilingual compound verbs, making a single constituent VP which is a plausible example of insertion such as in example (10). The very frequent uses of 'live' as attributive adjective in the Pashto frame are 'live call', 'live match', 'live show', and 'live program, etc. All the basic diagnostic features, single constituent, nested a b a, selected element and content word are applicable to the collocation 'live show'.

List of abbreviations

1, 2, 3	first, etc. person	f	feminine	PFV	perfective
ADV	adverb	FUT	future	PL	plural
BCV	bilingual complex verb	GEN	genitive	POSS	possessive
CL	clitic	IPFV	imperfective	PRN	pronoun
CM	code-mixing	KP	Khyber Pakhtunkhwa	PRS	present
COMP	complementizer	LOC	locative	PRX	proximate
COP	copula	LVC	light verb construction	PST	past
DM	demonstrative	M	masculine	RECP	reciprocal
DST	distal	OBJ	object	REDUP	reduplication
ERG	ergative	OBL	oblique	SG	singular

References

- Akhtar, Raja N. (2003). Aspectual Complex Predicates in Punjabi. *The Yearbook of South Asian Languages and Linguistics*. New Delhi.
- Annamalai, E. (1971). Lexical insertion in a mixed language. *Papers from the Seventh Regional Meeting. Chicago Linguistic Society*, 20-27. Chicago: University of Chicago.
- Babrakzai, Farooq (1999). Topics in Pashto syntax. Manoa, HI: University of Hawai'i at Manoa dissertation.
- Backus, Ad M. (1999) Evidence for lexical chunks in insertional codeswitching. In *Language encounters across time and space*, E.L. Brendemoen & E. Ryen (Eds.), 93–109. Oslo: Novus Press.
- Belazi, Hedi M., Edward J. Rubin and Almeida Jacqueline Toribio. 1994. Code Switching and X--Bar Theory: The Functional Head Constraint. *Linguistic Inquiry* 25:2, 221--237.
- Butt, Miriam. (1995). *The Structure of Complex Predicates in Urdu*. Stanford, California: CSLI Publications.
- Cattell, R., (1984). *Syntax and Semantics: Composite Predicates in English*. Academic Press, London.
- Di Sciullo, Anne-Marie, Pieter Muysken, and Rajendra Singh. 1986. Government and code-mixing. *Journal of Linguistics* 22:1-24.
- Grimshaw, Jane and Armin Mester. (1988). Light verbs and Theta marking. *Linguistic Inquiry* 19, 205–232.
- Khan, A, A. (2011). *Social factors and English code-mixing in Pashto language: code-mixing in Pashto speech community*. LAP LAMBERT Academic Publishing.
- Khan, A.A & Bukhari, H.N. (2014). Descriptive Study of Compound Verbs and Light Verbs Constructions in Pashto-English Code Switching. *Kashmir Journal of Language and Research: ISSN-1028-6640. Volume 17 NO. 2. 2-14*
- Khan, A.A & Bukhari, H.N. (2015). Pashto-English Code switching: A Device of Indigenization and Loanwords. *Kashmir Journal of Language and Research: (Accepted)*
- Labov, William (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lotfabbadi, Leyla Naseh (2002). Disagreement in Agreement-a Study of Grammatical Aspects of Codeswitching in Swedish - Persian Bilingual Speech. Doctoral dissertation, Stockholm University.
- MacSwan, J. (1999). A minimalist approach to intrasentential code switching. New York: Garland Press.
- MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from codeswitching. *Bilingualism: Language and Cognition*, 3(1), 37-54.
- Muysken, Pieter. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.
- Myers-Scotton, C. 2002a. *Contact linguistics, bilingual encounters and grammatical outcomes*. Oxford, UK: Oxford University Press.
- Myers-Scotton, C. 2008. Language contact: Why outsider system morphemes resist transfer. *Journal of Language Contact*, 2, 21–41.
- Myers-Scotton, C., & Jake, J. 2009a, January 8–11). Predicting distributions in creoles by morpheme type. Paper presented at annual meeting, Society for Pidgin and Creole Linguistics, San Francisco, CA.

- Myers-Scotton, C., & Jake, J. 2009b. A universal model of code-switching and bilingual processing and (2002) *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
- Myers-Scotton, C., & Jake, J. 1993b) *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press. (Paperback ed. 1997 with new Afterword.)
- Myers-Scotton, C. (1993 [1997]) *Duelling Languages: Grammatical Structure in Codeswitching* (1997 edition with a new Afterword). Oxford : Clarendon Press.
- Nortier, J. (1990). *Dutch Moroccan Arabic code-switching among young Moroccans in the Netherland*, Dordrecht: Foris.
- Penzel, H. (1961) Western loanwords in Pashto. *Journal of the American Oriental Society*, 81, 1, 43-52.
- Polack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPANOL; *language*, 18, 581-618.
- Poplack, S. (1988) Contrasting patterns of code-switching in two communities. In Heller M.(ed) *Codeswitching, Anthropological and Sociolinguistic Approaches*, 17-44. Berlin: Mouton de Gruyter.
- Poplack, Shana, Sankoff, David & Miller, Christopher (1988) The Social Correlates and Linguistic Processes of Lexical Borrowing and Assimilation. *Linguistics* 26:47-104.
- Roberts, Taylor. (2000). Clitics and agreement. Cambridge, MA: MIT PhD dissertation.
- Robson, Barbara, and Habibullah Tegey.(1996). A reference grammar of Pashto. Washington, D.C:Center for Applied Linguistics. Online: <http://www.eric.ed.gov/PDFS/ED399825.pdf>.
- Romaine, S. (1995) (2nd ed.) *Bilingualism*, Oxford: Blackwell.
- Siegel, Jeffrey (1987) *Language contact in a plantation environment*. Oxford: Oxford University Press..
- Treffers-Daller, J. (1994). *Mixing Two Languages:French-Dutch Contact in a Comparative perspective*. Berlin: Mouton de Gruyter.
- Trudgill, P. (1972). Sex, Covert prestige, and linguistic change in the urban British English of Norwich, *Language in society* 1, 179-195.
- Trudgill, P. (1986). *Dialects in contact*. Oxford: Blackwell.
- Weinreich, U. (1953). *Language in contact*.The Hague: Mouton.
- Woolford, Ellen. 1983. Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14:520-536.