

## Insertion, Alternation, Congruent Lexicalization Corpus-based approaches to bilingual speech

Roeland van Hout (University of Tilburg)  
Pieter Muysken (University of Amsterdam)

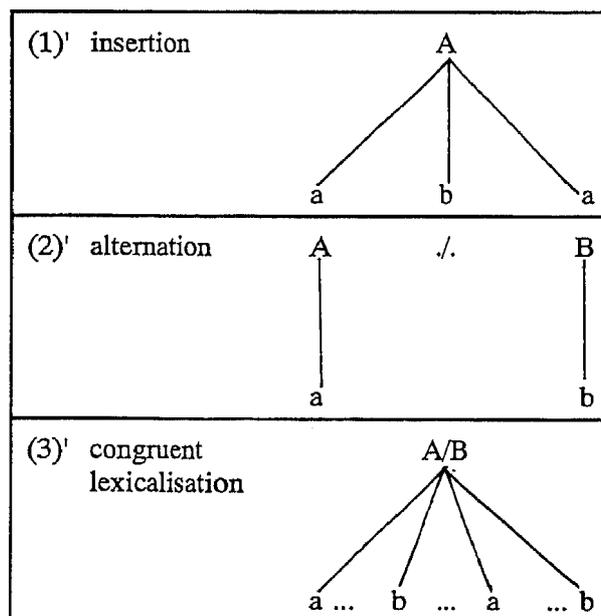
In this brief note we aim (a) to make plausible that there are three separate patterns of code-switching, and (b) to propose ways to study these patterns through the systematic exploration of bilingual corpora, rather than by looking at individual examples. The three patterns are insertion, alternation, and congruent lexicalization. Examples are given in (1)–(3):

- (1) INSERTION  
Yo anduve in a state of shock pa dos días. (Pfaff, 1979)  
I walked in a state of shock for two days.
- (2) ALTERNATION  
Andale pues and do come again. (Peñalosa, 1972)  
That's alright then ...
- (3) CONGRUENT LEXICALIZATION  
Bueno, in other words, el flight que sale de Chicago around three o'clock. (Pfaff, 1976)  
Good, in other words, the flight that leaves Chicago around three o'clock.

The interpretation of these patterns can be

- structural: in terms of labels in tree configurations;
- psycholinguistic: different degrees of activation of components of both languages;
- sociolinguistic: bilingual strategies (Poplack & Sankoff's equivalence versus insertion).

The structural interpretation is what concerns us most here. Consider the trees (1)'–(3)', where A, B are language labels for non-terminal nodes, and a, b are labels for terminal, i.e. lexical, nodes:



Now the issue of constraints, so dominant in the literature on code-switching, can be rephrased in terms of structural profiles. These are roughly as in (4):

- |                           |   |
|---------------------------|---|
| (4) insertion:            | matrix order dominates                  |
|                           | matrix function words dominate          |
| alternation:              | basic linear equivalence at switch site |
| congruent lexicalization: | none?                                   |

The sociolinguistic profiles corresponding to the three patterns are, very tentatively, as in (5):

- |                           |   |
|---------------------------|---|
| (5) insertion:            | (neo-/ex-)colonial settings, recent migrant communities                                 |
| alternation:              | stable communities with tradition of language separation                                |
| congruent lexicalization: | second generation migrant groups, dialect/standard diglossia, closely related languages |

Our main concern here is to outline diagnostic features for the three patterns. How do we know which pattern is dominant in a given bilingual corpus? We will suggest a number of criteria which may be applied to settle this.

### Constituency

- When several constituents in a row are switched, which together do not form one unique constituent, insertion is not plausible. A number of elements form a unique constituent if that constituent contains no other elements.
  - With several constituents, we would have to assume multiple contiguous insertions.
  - When the switched element is a single, well-defined constituent, e.g. a noun phrase or a prepositional phrase, insertion is a plausible option; this holds a fortiori for single words.
  - When several words are switched which do not form one or more constituents together, congruent lexicalization is plausible.
- (6) *single constituent*  
 žib li-ya een glas water of zo (Moroccan Arabic-Dutch; Nortier, 1990: 131)  
 get for me a glass of water or something
- (7) *several constituents*  
 He was sitting down en la cama, mirandonos peleando, y really, I don't remember si el nos separó  
 ... on the bed, watching us fight, and ... if he took us apart.
- (8) *not a unique constituent*  
 Yeah, but I buy 'em mostly pa' el hamburger meat. (Lance, 1978)  
 ... for the ...
- (9) E puntanan aki no [tin niks] [mee te maken]. (Kook, 1993)  
 The points here [have nothing] [to do with it].

### Peripherality

- When the switched element is at the periphery of an utterance, alternation is a clear possibility.
- (10) les étrangers ze hebben geen geld hè  
the foreigners they don't have any money, do they?  
(French-Dutch in Brussels; Treffers, 1990:271)

### Nestedness

- When the switched string is preceded *and* followed by elements from the other language, which are structurally related, it is a case of insertion.
  - However, when the switched string is preceded *and* followed by elements from the other language, which are not structurally related, it is a case of alternation or congruent lexicalization.
- (11) Es una little box asina y ya viene ... (Lance, 1978:145)  
It is a little box like this and it comes already ...
- (12) (A) Se me hace que (B) I have to respect her (C) porque 'ta ... older  
It appears to me that ... because [she] is ... (Lance, 1978:143)

### Selectivity

- If a switched fragment forms a constituent selected by an element in the fragment in the other language, insertion is a good possibility. If not, alternation is a plausible option.
- (13) Duș kal elke keer kal matje bilang iets a sing reageren of aku bikin eh heel dom.  
Thus when each time when she say something I NEG react or I act eh very dumb.  
(Moluccan Malay/Dutch; Huwaë, 1992)  
'Thus each time she says something I don't react or I act very dumb.'
- (14) We've got all these kids here right now, los que están ya criados aquí.  
..., those that already grew up here. (Gumperz & Hernandez Chavez, 1975:157)

### Equivalence

- Is there linear equivalence between the two languages? If so, alternation and congruent lexicalization are a clear possibility.

### Length

- The more words a switched fragment contains, the more likely that it is alternation rather than insertion.

### Complexity

- The more complex structure a switched fragment contains, the more likely that it is a case of alternation rather than insertion.
- (15) (A) Right to 104th Street (B) donde tenía una casa (C) which were furnished rooms.  
(... where I had a house ...). (Poplack & Sankoff, 1979)

### Embedding in discourse

- Consider a mixed clause starting in language A and ending in language B. If the preceding utterance is in A, and the following clause is in B, alternation is a plausible analysis.

### Structural position

- If the switch takes place at a major clause boundary, alternation is a plausible option. An example is (2) above.

(16) ... maar 't hoeft niet li-'anna ida šeft ana ...  
... but it need not (be) for when I-see I ...

(Moroccan Arabic-Dutch; Nortier, 1990:126)

### Flagging

- If the switch is flagged, alternation between codes is a plausible option.

(17) Tu sais, l'affuteur de scies hein, daan gink bij die beenhouwers, de zager.  
You know, the knife-grinder eh, he went to the butchers, the sawyer.

(French-Dutch in Brussels; Treffers, 1990:266)

The diagnostic values of the whole set of features can be listed in the following way:

|                         | insertion | alternation | congruent<br>lexicalization |
|-------------------------|-----------|-------------|-----------------------------|
| ○ single constituents   | +         |             |                             |
| ○ several constituents  |           | +           |                             |
| ○ non-constituency      |           |             | +                           |
| ○ peripherality         |           | +           |                             |
| ○ nested a b a          | +         |             |                             |
| ○ non-nested a b a      |           | +           | +                           |
| ○ selected element      | +         |             |                             |
| ○ linear equivalence    |           | +           | +                           |
| ○ length                |           | +           |                             |
| ○ complexity            |           | +           |                             |
| ○ discourse embedding   |           | +           |                             |
| ○ major clause boundary |           | +           |                             |
| ○ flagging              |           | +           |                             |

This list of features is neither exhaustive nor mutually exclusive. Such a list can be used of course for determining the underlying pattern of individual examples of code-switching. The real impact of these features, however, can be evaluated only by studying their properties on the level of bilingual corpora. In using and analysing bilingual corpora it is necessary to distinguish two interdependent entrances:

- (1) a stepwise, incremental procedure in handling and coding bilingual corpora;
- (2) the (simultaneous) development of measures/indices for diagnosing code-switching patterns.

An incremental coding procedure for code-switching analysis will contain minimally the following successive steps:

- mark turns;
- mark utterances;
- assign global language label on the word level (etymology);
- assign specific language label (phonetic and morphological incorporation);
- assign categorial labels;
- perform morphemic analysis (including language labels), where appropriate;
- mark constituent boundaries;
- give structural analyses.

There is not much experience in applying quantitative measures or indices to bilingual corpora. In developing measures/indices concepts like distribution, density, and dependency will have to be taken into account. Some important aspects of corpus-related measures are:

- the distribution of the languages across discourse;
- the distribution of the languages across turns, speakers, utterances and words;
- transitional patterns across turns, speaker, sentences and words;
- patterns on the speech management level (turn taking, feedback, etc).

Key concept of analysis seems to be the concept of transitional patterns in distributions of adjacent elements or sets of elements with different language labels.

## The effect of code-switch density and semantic context on the recognition of code-switches by bilingual listeners

Markus Leuenberger

University of Neuchâtel and University of Basel, Switzerland

The aim of this study was to explore how code-switch density and semantic context influence the recognition of code-switches in bilingual speech. To do this, we asked the following questions:

- 1) Does code-switch density affect the recognition of code-switched words (i.e. does an increase in code-switching facilitate the access of code-switches)?
- 2) Does semantic context also influence the recognition of code-switches (i.e. does the reference to a domain which normally involves the other language speed up the access of code-switches)?

The rating task was used to present sentences in Swiss-German with French code-switches to Swiss-German – French bilingual subjects and the dependent variables measured were the isolation point, the confidence ratings and the candidates proposed.

### 1. Method

#### Subjects

48 bilingual students (Swiss-German – French) from the University of Basel took part in the experiment.

#### Material

Fifteen French nouns which are often used in a Swiss-German – French bilingual context were selected for the experiment. They were embedded in the second clause of a two-clause Swiss-German sentence. The first clause was modified so as to vary the two independent variables (code-switch density and semantic context). This produced four types of sentences:

Type a. The first clause of the sentence contained no code-switch and did not constrain the code-switch in the second clause.

e.g. Sit em Gspröch am Gänfersee weiss är, dass är PRAGMATIQUE nimmt.

Type b. The first clause contained no code-switch but it did constrain the code-switch in the second clause.

e.g. Är isch vo de Sprächakt e so fasziniert, dass är PRAGMATIQUE nimmt.

Type c. The first clause contained a code-switch but did not constrain the code-switch in the second clause.

e.g. Sit em Gspröch am LAC LEMAN weiss är, dass är PRAGMATIQUE nimmt.