

Minority Language Survival: Code-mixing in Welsh

Margaret Deuchar
University of Wales, Bangor

1. Introduction

The aim of this paper is to report on the results of some preliminary analysis of Welsh-English code-mixing data, in order to determine which structural pattern of code-mixing is predominant. This is the first step of a more wide-ranging analysis which will deal with a larger amount of data and will contribute to the issue of which linguistic and extralinguistic factors may influence the occurrence of one type of code-mixing rather than another. A secondary aim of this study is to test the viability of a typological approach by developing a methodological tool to identify the dominant pattern of code-mixing based on diagnostic features. The term *code-mixing* is used following Muysken (2000:1) to refer “to all cases where lexical items and grammatical features from two languages appear in one sentence” and the focus of interest will be on “intrasentential mixing” or mixing where elements from both languages appear in the same sentence. Many authors use the term *code-switching* to refer to the same phenomenon.

2. Previous research

2.1 On code-mixing in general

Previous research on code-mixing has followed three main trends: (a) the search for universal constraints, typified by the seminal work of Poplack (1980); (b) the assumption of asymmetry, initiated by Joshi (1985) and developed by Myers-Scotton (1993) and (c) the typological approach, advocated by Muysken (2000). Muysken suggests that instead of one code-mixing model serving for all language pairs, there are three main types of code-mixing: insertional, alternational and congruent lexicalization. Insertional mixing is the type proposed by Myers-Scotton, alternational mixing the type proposed by Poplack, and congruent lexicalization is a type which Muysken identifies as occurring in a pair of languages or varieties where the grammar is largely shared but the vocabulary is different. Muysken suggests that the dominant code-mixing pattern in a particular speech community can be predicted on the basis of both linguistic and extralinguistic factors. For example, typological distance may predict either insertional or alternational code-mixing, but not congruent lexicalization, and a colonial setting may predict insertion.

2.2 On code-mixing in Welsh and Irish

There has been virtually no previous work on structural aspects of code-mixing in Welsh, though some pioneering work has been done by Jones (1993, 1995, 2000) within a discourse analysis perspective. Jones (1993) discusses reasons for code-mixing by Welsh speakers, Jones (1995) offers an analysis of both spoken and written data emphasizing the heterogeneity of language use and Jones (2000) examines code-mixing as a part of bureaucratic discourse.

Structural aspects of code-mixing in Irish have nevertheless been examined by Stenson (1990). This work is relevant to us since Irish, like Welsh, is a Celtic VSO language which is spoken by bilinguals living in a society where English is the dominant language. Stenson found evidence in her data for the violation of both Poplack’s equivalence and her free morpheme constraint. Poplack (1980) had proposed that code-switching was subject to two main constraints, the equivalence and the free-morpheme constraint. According to the equivalence constraint, a switch point between two languages can only occur where the order of items on both sides of the switch point is compatible with the grammar of the two languages. This is illustrated by Poplack (1980:586) in (1), where the slashes indicate the only possible switch points between Spanish and English parts of the utterance:

- (1) (Yo) / le dije / eso / pa' que / (el) / la trajera / ligero
 I / told him / that /so that / he / would bring it / fast
 I him told that so that he it would-bring fast'

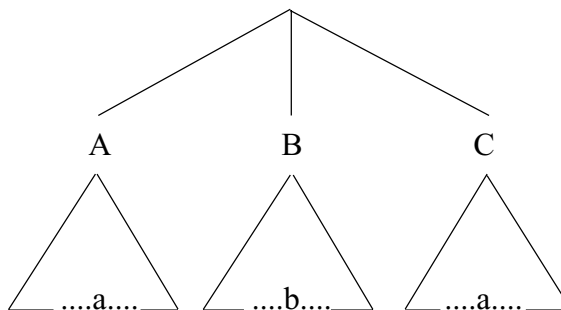
Stenson found that in Irish-English code-mixing data, the equivalence constraint was violated by switches between an Irish noun and a following English adjectival phrase (as in *carr light green*, 'light green car'). Poplack's other constraint, the free-morpheme constraint, is captured succinctly by Appel and Muysken (1987) who define it as ensuring that "no switch may take place between two morphemes which are morphologically bound to one another" (Appel & Muysken 1987). Thus an example like **eatiendo*, with a stem from English and an inflected present participial ending from Spanish, cannot occur. However, Stenson found that this free morpheme constraint was violated by the appearance in her data of English nouns with Irish plural suffixes (as in *jobanna*, 'jobs').

Poplack's constraints were subsequently tested on many sets of data, and Stenson's was one of those which appeared to provide counter-examples: see also e.g. Bentahila & Davies 1983, Berk-Seligson 1986. If a typological approach were taken to the Irish-English data as suggested by Muysken (2000) and outlined above, it seems likely that the data would be identified as predominantly insertional pattern (see below for a description of this). For reasons to be outlined, we would also expect this to be the dominant pattern in the Welsh-English data.

2.3 Types of code-mixing

Muysken (2000) suggests that there are three main code-mixing patterns which may be found in bilingual speech communities: insertion, alternation and congruent lexicalization. One pattern will usually dominate, though not necessarily to the exclusion of other patterns. In the **insertion** pattern, one language determines the overall structure into which constituents from the other language are inserted: this is illustrated in Figure 1, based on Muysken (2000:7). This pattern is assumed by the

Figure 1: the insertion pattern



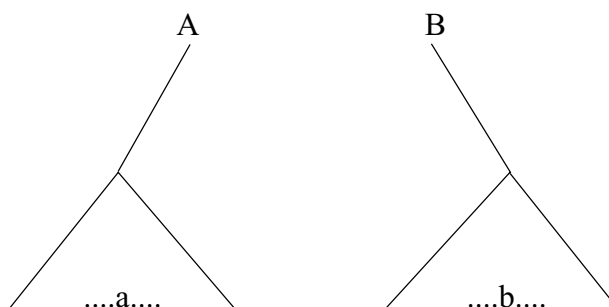
Matrix Language Frame (MLF) theory proposed by Myers-Scotton (e.g. 1993). It can be illustrated in (2) by a Swahili-English example from Myers-Scotton (1993:86):

- (2) *A-na-ku-l-a* ***plate m-bili*** *z-a murram*
 3s-PRES-INFIN-eat-INDIC plate CL 10-two CL 10-of maize
 ‘He eats two plates of maize’

In this example the word order is as in Swahili, including the phrase ***plate m-bili*** ‘two plates’, and all the inflectional morphology is from Swahili. The asymmetry between the two languages involved in the insertional pattern is captured in the MLF by labelling the main language the “matrix” language and the other the “embedded” language.

In the **alternation** pattern, both languages occur alternately, each with their own structure, as illustrated in Figure 2 based on Muysken (2000:7).

Figure 2: the alternation pattern



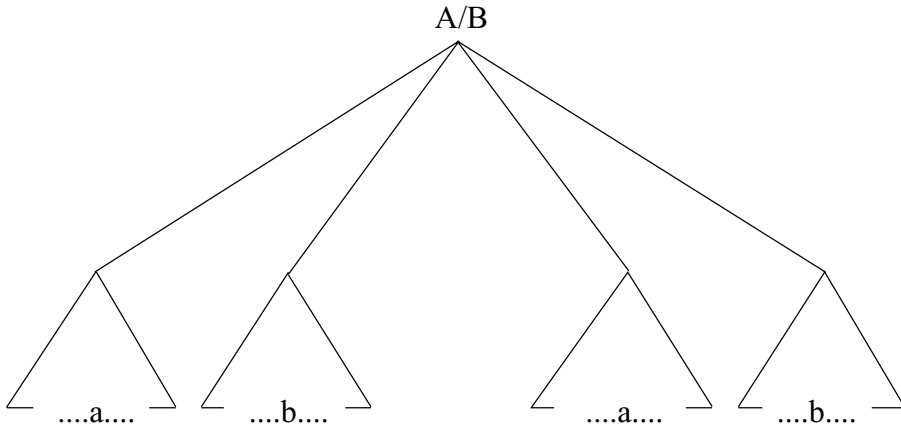
This type of code-mixing is assumed in Poplack’s work (e.g. 1980) and is well represented in her data in examples like (3) (Poplack:1980:589):

- (3) *Why make Carol* ***sentarse atrás pa’ que*** *everybody has to move* ***pa’ que se salga?***
 sit-REFL at-back so-that for that REFL get-out-SUBJUNC
 ‘Why make Carol sit in the back so that everybody has to move for her to get out?’

In this example each language stretch, whether English or Spanish, has its own language-specific syntax and morphology, with neither language providing an overall structural frame for the utterance.

In the third type of code-mixing, **congruent lexicalization**, “the grammatical structure is shared by languages *A* and *B*, and words from both languages *a* and *b* are inserted more or less randomly” (Muysken 2000:8). This is illustrated in Figure 3 based on Muysken (2000:8). Muysken proposes this type with reference mainly to standard/dialect mixing.

Figure 3: the congruent lexicalization pattern



This type can be illustrated by example (4) from the Ottersum dialect (in bold) and standard Dutch:

(4) **ja maar bij** *ouwe mensen komt dat* *gauwer tot stilstand als bij jonge mense wa*
 yes but with / older people/ comes that /more quickly to a halt than / with younger people eh

(Muysken 2000:130, citing Giesbers 1989:147)

In this example the fragments from each variety apparently have no internal grammatical cohesion as is expected in this type of mixing.

2.3.1 Diagnostic features

Muysken (2000: 230-231) suggests that specific diagnostic features can be applied to a corpus, and that the set of values for each feature will match one pattern more closely than the others. The features are grouped under the headings of *constituency*, *element switched*, *switch site* and *properties*, and are listed in the first column of Table 1. Some features only apply to some switches, when certain conditions are met (see below), while others apply to all, and these features are underlined in Table 1. The three next columns of Table 1 show Muysken's expected values of each of the features for the patterns insertion, alternation and congruent lexicalization.

The first group of features relates to constituency, and as is shown in the Table, the insertion pattern is distinguished from the others by positive values for the features *single constituent* and *nested aba*, the alternation pattern by positive values for *several constituents* and *non-nested aba*, and the congruent lexicalization pattern by positive values for *non-constituent* and *non-nested aba*. The contrast between 'nested' and 'non-nested' aba applies to switches which have other-language material both before and afterwards, but the material before and after the switch is grammatically related in the 'nested' case, but not in the 'non-nested' case. I assume these two features do not apply to non 'aba' switches, i.e. those which are either at the beginning or the end of a turn, and thus do not have other-language material both before and afterwards. Where the value of a feature in columns 2, 3 and 4 is given as 0, this indicates that its value is not relevant for the determination of the overall mixing pattern given in the column heading.

Table 1: Analysis of data using quantification of Muysken's features

MUYSKEN'S FEATURES	INSERTION	ALTERNATION	CONGRUENT LEXICALIZATION	example: <i>camouflaged</i>	Score on INS	Score on ALT	Score on CON
<u>single constituent</u>	+	0	0	+	1	0	0
<u>several constituents</u>	-	+	0	-	1	-1	0
<u>non-constituent</u>	-	-	+	-	1	1	-1
nested aba	+	-	0	+	1	-1	0
non-nested aba	-	+	+	-	1	-1	-1
diverse switches	-	0	+	Type of switch: Past participle			
<u>long constituent</u>	-	+	-	-	1	-1	1
<u>complex constituent</u>	-	+	-	-	1	-1	1
<u>content word</u>	+	-	-	+	1	-1	-1
<u>function word</u>	-	-	+	-	1	1	-1
<u>adverb, conjunction</u>	-	+	-	-	1	-1	1
<u>selected element</u>	+	-	+	+	1	-1	1
emblematic or tag	-	+	0	N/A	0	0	0
<u>major clause boundary</u>	0	+	0	-	0	-1	0
<u>peripheral</u>	0	+	0	+	0	-1	0
embedding in discourse	0	+	0	N/A	0	0	0
<u>flagging</u>	-	+	-	-	1	-1	1
<u>dummy word insertion</u>	+	0	-	-	-1	0	1
<u>bi-directional switching</u>	-	+	+	Switch into English			
<u>linear equivalence</u>	0	+	+	+	0	1	1

<u>telegraphic mixing</u>	+	-	-	-	-1	1	1
<u>morphological integration</u>	+	-	+	-	-1	1	-1
<u>doubling</u>	-	+	-	-	1	-1	1
<u>homophonous diamorphs</u>	0	-	+	-	0	1	-1
<u>triggering</u>	0	0	+	N/A	0	0	0
<u>mixed collocations</u>	0	-	+	-	0	1	-1
<u>self-corrections</u>	-	+	-	-	1	-1	1

The next group of features relates to the nature of the element switched. The feature *diverse switches* actually applies to the whole corpus rather than individual instances: in my analysis I have merely recorded the grammatical category of individual switches in order to determine the overall diversity of switches afterwards. A high diversity of switches would favour the congruent lexicalization pattern, whereas a low diversity would be indicative of the insertion pattern. The insertion pattern is also characterised by positive values for the features *content word* and *selected element* (applied to an object or complement) in this group, and the alternation pattern by positive values for the features *long constituent*, *complex constituent*, *adverb or conjunction*, and *emblematic or tag*. This last feature involves the mixing in of tags or interjections (Muysken 2000:99). I have assumed that it does not apply to a switch which has a positive value for any of the features *long constituent*, *selected element* or *morphological integration* (see below). The congruent lexicalization pattern is characterized by positive values for the features *function word* and *selected element* in this group.

The third group of features relates to the switch site involved. In this group the feature *bidirectional switching*, like *diverse switches* discussed above, will apply to the corpus as a whole rather than to individual switches: in my analysis I have noted the direction of each switch so that the proportion of Welsh-English vs. English-Welsh switches can be determined afterwards. A roughly equal proportion of switches in each direction will be indicative of the *alternation* or *congruent lexicalization* pattern whereas a predominance of switches in one direction rather than another will favour the *insertion* pattern. Insertion is also characterised by the individual feature *dummy word insertion*. This refers to the insertion of semantically empty elements (cf Muysken 2000: 105). However, if a switch is at a *major clause boundary* or *peripheral* to the clause, this will indicate the alternation pattern, as will *embedding in discourse*. This applies when the same language is maintained across a turn boundary (Muysken 2000:104): I have assumed that this feature only applies to switches which occur at the beginning or end of a turn. *Flagging*, another feature characteristic of alternation, refers to mixing which is highlighted by the insertion of a discourse marker (see Muysken 2000:101); the absence of flagging, however, is indicative of the congruent lexicalization pattern.

The features in the fourth and final group relate to properties of the switch itself other than constituency, which are dealt with in the first group. Insertion is favoured by positive values of the features *telegraphic mixing* and *morphological integration*: Telegraphic mixing occurs where elements have been omitted which should have been present in one or both of the languages involved. Morphological integration, which is also indicative of congruent lexicalization, occurs where one of the languages determines the overall grammatical framework, and where items switched from the other language are morphologically integrated into the main or matrix language. I have assumed this feature does not apply where the possibility of morphological integration does not arise, e.g. where there is no bound morpheme which could be replaced by a bound morpheme from the matrix language. Other features of the congruent lexicalization pattern are *linear equivalence*, *homophonous diamorphs*, *triggering*, and *mixed collocations*. As is clear from the Ottersum dialect/standard Dutch example

(Muysken 2000:13) above, apparently unconstrained mixing back and forth is only possible because of the linear equivalence or close word order parallels in the two varieties involved. Closely related varieties may also have some homophonous diamorphs or pairs of words which are phonetically similar in both varieties. Triggering is a phenomenon identified by Clyne (e.g. 1967) and is an interpretation of a multi-word switch where the choice of one of the words in the switch may have led to other words to be switched as part of a longer string. This feature appears to involve more interpretation than objective description, but has been applied in my analysis as carefully as possible. Mixed collocations occur where two elements of an idiomatic collocation from one of the languages are from different languages, e.g. *mynd off* ‘go off’ in Welsh-English conversations. The feature *linear equivalence* is also indicative of the alternation pattern: this feature receives a positive value where there is word order equivalence between the two languages on either side of the switch. *Doubling* and *self-corrections* are additional features of the alternation pattern. Doubling applies where the semantic value of the switch is the same as that of another morpheme in the original language, as e.g. when plurality is marked twice, as in the doubly marked Welsh-English plural in *llwynogods*¹ ‘foxes’ (Welsh *llwynog* ‘fox’, *-od* Welsh plural, *-s* English plural). Self-corrections are switches that involve repetition of similar material in the other language, often after a hesitation. A good example can be found in Poplack (1998: 53) where a speaker is relating in French a conversation he had in English (note that I have indicated the self-correction by underlining) : *J’ai dit “you don’t swim.” Il dit, “sure” il dit, “I can swim”. Il dit “sure” “Well, I says –j’ai dit “show it to me”.* (‘I said, “you don’t swim.” He says, “sure” he says, “I can swim”. He says “sure”. “Well, I says, I said, “show it to me”.’)

2.3.2 Predicting the code-mixing pattern

Whereas previous approaches to code-mixing predicted that a specific model would account for all patterns to be found in any speech community, Muysken (2000:8-9) proposes that the pattern will vary according to both linguistic and extralinguistic factors. My interpretation of one way in which he suggests this might work is summarised in Table 2. In this table we see that, in terms of linguistic factors, insertion and alternation are favoured by typological distance between the languages involved, whereas congruent lexicalization is seen to be more likely when the two languages are typologically similar. Although both insertion and alternation are predicted by typological distance, the two patterns are associated with different extralinguistic factors. For example, insertion is predicted to be more likely in colonial settings and where there is asymmetry in the speaker’s proficiency in the two languages, whereas alternation is predicted in stable bilingual communities where there is a tradition of language separation. Where there is no such tradition, however, and the languages are typologically similar, congruent lexicalization is more likely. If we apply the predictions summarized in Table 2 to Welsh-English code-mixing, we might expect to find a pattern of either insertion or alteration on structural grounds. While English is an SVO language, Welsh is VSO: this word order difference makes congruent lexicalization seem an unlikely pattern. Of those extralinguistic factors proposed by

Muysken as favouring insertion that of “colonial settings” is probably most applicable. It can be argued that Wales has been a colony of England since it was conquered in 1284, the Welsh language being excluded from legal use after the Acts of Union in 1536 and 1542. (This situation did not change substantially until 1942: see Thomas 1982: 87). It is interesting to note that the pattern found in Stenson’s (1990) Irish data, collected in a community with similar linguistic and extralinguistic characteristics, appears to have been predominantly one of insertion.

The three main patterns of code-mixing are not expected to be either watertight or static, however. More than one pattern of code-mixing may be found in a corpus of data, though it is likely that one will be predominant. It is also possible that the predominant pattern of code-mixing may change over time. Muysken (2000:249) suggests that prolonged language contact may lead to a change of pattern, and that in particular, an insertional pattern may change to one of either alternation or congruent lexicalization, again depending on both linguistic and extralinguistic factors. He suggests that alternation will be favoured by strong norms, language competition and typological distance, whereas congruent lexicalization is more likely in the case of loose norms, balance between languages and structural

1 I am grateful to Winifred Davies for this example.

parallels. In Wales the tradition of language separation seems best established in the more formal or written contexts, but not in informal contexts (see Thomas 1992: 267). We might thus predict different patterns according to the formality of the context, with signs of alternation appearing in more formal or semi-formal contexts, and congruent lexicalization in less formal contexts. In this paper preliminary results from the analysis of data collected in an informal context are presented.

Table 2 : Muysken’s view of the relation between code-mixing patterns and extralinguistic factors

Code-mixing pattern	Linguistic factors favouring this pattern	Extralinguistic factors favouring this pattern
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

3. Data analysis

The data to be analysed were collected by a native Welsh speaker from an informal social gathering to which she invited some colleagues. All of the seven participants in the conversation are native Welsh speakers who are bilingual in Welsh and English. (Aside from infants, monolingual Welsh speakers are virtually non-existent, given the extent of exposure to English at home (at least in the media), school and in the community, where monolingual speakers are to be found in all areas, to a greater or lesser extent.) All are women, and six are in their thirties while one is in her fifties. In order to determine the main code-mixing pattern in the Welsh-English data, a method was devised to quantify the values assigned to Muysken’s diagnostic features in order to measure the extent to which these values matched each of the three possible patterns. Each switch in the data was coded for the diagnostic features listed in Table 1. (Note that items considered to be English loans because of their listing in the authoritative *Geiriadur Prifysgol Cymru* ‘Dictionary of the University of Wales’, <http://www.cymru.ac.uk/geiriadur>) were not considered to be switches. For each diagnostic feature listed in Table 1, one of three values was assigned: ‘+’, ‘-’ or ‘N/A’ (non-applicable). (The only exceptions to this procedure arose from the two features identified above as applying to the whole corpus rather than to individual switches, *diverse switches* and *bidirectional switching*: for each switch qualitative information was noted as to the syntactic category of the switch and its direction, either Welsh-English or English-Welsh.) These values were then given a score for the extent to which they matched each of the three possible patterns, insertion, alternation and congruent lexicalization, for which the expected values for each feature are given in Table 1. A score of ‘1’ was assigned where a feature value matched the positive or negative value expected for a particular pattern, ‘-1’ where it did not match, and ‘0’ where the feature did not apply to the switch in question. A score of ‘0’ was also assigned where Muysken’s expected value was shown as ‘0’ in Table 1, indicating that the value of the feature was neutral in relation to the pattern in question. The example switch in the fifth column of Table 1, *camouflaged*, occurred in the utterance *Ia, mae o’n reit camouflaged yn dydi*, and can be used to demonstrate how the scoring works. For example, the first feature, *single constituent*, receives a positive value for *camouflaged* as a switch. This matches the expected value of the feature for the

insertion pattern, as shown in the second column, second row of Table 1. Thus the score on insertion for *camouflaged* is 1. The value of the feature is neutral for the alternation and congruent lexicalization patterns, however, as indicated by '0' in the third and fourth columns of Table 1, second row. To take an example of a feature in the fourth group, morphological integration, we can see from the second to fourth columns, sixth row from the bottom, that the expected value of this feature is positive for the insertion and congruent lexicalization patterns, and negative for the alternation pattern. The value for the switch *camouflaged* is negative, since this past participle has not been morphologically integrated into Welsh, but retains the English *-ed* inflection. Thus on this feature the switch receives a score of 1 for the alternation pattern (see penultimate column of Table 1, sixth row from the bottom) which it matches in this respect, but a score of -1 for the insertion and congruent lexicalization pattern, which it mismatches (see sixth and eighth columns of Table 1). If we add up the scores for *camouflaged* for each pattern we obtain a total score of 11 for insertion, -6 for alternation and 3 for congruent lexicalization. I interpret this as indicating that, for this switch, the insertion pattern is the most strongly indicated one; alternation is counter-indicated, while congruent lexicalization is a possible secondary pattern. If the scores for all the switches on each pattern in a particular corpus are added together we obtain an overall score which can be interpreted as indicating the overall dominant pattern. In addition, the qualitative information derived from coding a large number of switches in terms of the syntactic category of the switch and the direction of the switch can be used as further evidence in determining the dominant pattern. In Table 3 the results of scoring five further switches as well as *camouflaged* are recorded. The numbers in the top row refer to the numbers of utterances listed in the Appendix. In Table 4 the total scores for each of the items on each pattern are given, which makes it possible to determine the dominant pattern in each case. Table 5 shows the total score of the first 50 switches in the corpus. The score for insertion is 408, for alternation is -168, and for congruent lexicalization is 82. These figures suggest that insertion is the dominant pattern overall, that alternation is counter-indicated (the minus figure shows that the number of features mis-matched were greater than those matched) and that there is some evidence for congruent lexicalization as a secondary pattern.

4. Discussion

As predicted, we find insertion to be the dominant pattern in the Welsh-English code-mixing data. If we examine the two diagnostic features which apply to the corpus as a whole, diverse switches and bidirectional switching, we also find that the overall results are compatible with the insertion pattern. 64% of the switches are either single nouns or noun phrases, which makes for an overall low diversity of switch types, and matches the negative value of this feature for the insertion pattern shown in Table 1. As for directionality of switching, 96% of the switches are from Welsh to English rather than English to Welsh, which indicates a lack of bidirectionality, also matching the negative value of this feature for insertion shown in Table 1.

However, the score of 82 in Table 4 indicates congruent lexicalization as a possible secondary pattern, as predicted above for data produced in informal contexts. This pattern is doubtless the result of prolonged language contact between English and Welsh, and can be compared to other aspects of language contact which appear to result in closer approximation of the structures of the two languages. One of these, also noted by Thomas (1992), is the calquing or loan translation of English phrasal verbs, as in the following examples which I have observed: *troi allan* 'turn out', *sortio allan* 'sort out', *ffindio allan* 'find out', *adïo i fyny* 'add up', *pigo i fyny* 'pick up', *rhedeg i lawr* 'run down', *torri i lawr* 'break down'. This type of loan translation also tends to be one of the focuses of prescriptive attempts to avoid anglicisms in Welsh. Another kind of English influence on the use of Welsh verbs, according to Thomas (1992) is the use of the modal auxiliaries in constructions with uninflected main verbs. Thomas goes so far as to suggest that the establishment of this auxiliary + main verb pattern could have led to the breakdown of synthetic tense-marking. This is particularly apparent in the spoken language, where periphrastic constructions involving an inflected form of *bod* 'be' and an uninflected main verb are much more common than in the written language.

Table 3: Analysis of first six examples in Appendix

MUYSKEN'S FEATURES	<i>(1) camou- flaged</i>	<i>(2) exercisio</i>	<i>(3) Just as well</i>	<i>(4) off</i>	<i>(5) squeekio, toys</i>	<i>(6) ngwasi</i>
single constituent	+	+	+	+	-	+
several constituents	-	-	-	-	-	-
non-constituent	-	-	-	-	+	-
nested aba	+	+	N/A	+	-	N/A
non-nested aba	-	-	+	-	+	N/A
<i>type of switch</i>	Past participle	V	Adv	Particle	V+N	tag
long constituent	-	-	-	-	-	-
complex constituent	-	-	-	-	-	-
content word	+	+	-	-	+	-
function word	-	-	-	-	-	-
adverb, conjunction	-	-	+	+	-	+
selected element	+	+	-	+	-	-
emblematic or tag	N/A	N/A	-	N/A	-	+
major clause boundary	-	-	+	-	-	+
peripheral	+	-	+	-	-	+
embedding in discourse	N/A	N/A	N/A	N/A	N/A	+
flagging	-	-	-	-	-	-
dummy word insertion	-	-	-	-	-	-
<i>direction of switch</i>	English	English	English	English	English	Welsh
linear equivalence	+	+	+	+	+	+
telegraphic mixing	-	-	-	-	-	-
morphological integration	-	+	N/A	N/A	N/A	N/A
doubling	-	-	-	-	-	-

homophonous diamorphs	-	-	-	-	-	-
triggering	N/A	-	-	-	-	N/A
mixed collocations	-	-	N/A	+	N/A	-
self-corrections	-	-	-	-	-	-

Table 4: Scores on mixing patterns for examples and pooled data

SCORES ON MIXING PATTERNS	(1) <i>camou-flaged</i>	(2) <i>exercisio</i>	(3) <i>Just as well</i>	(4) <i>off</i>	(5) <i>squeekio, toys</i>	(6) <i>ngwasi</i>
Insertion	11	13	4	8	-1	3
Alternation	-6	-8	4	-5	-9	7
Congruent lexicalization	3	5	4	6	9	3
Dominant pattern	Insertion	Insertion	Alter-nation	Insertion	Congruent Lexicalization	Alter-nation

Table 5: Scores on first 50 switches

SCORES ON MIXING PATTERNS	Total based on 50 switches
Insertion	408
Alternation	-168
Congruent lexicalization	82
Dominant pattern	Insertion

Code-mixing, then, is one aspect of language contact, and it may contribute, along with other processes, to the achievement of greater congruence between the two languages over time. Congruence is a notion which has received explicit attention relatively recently (cf Myers-Scotton 1993: 120, Myers-Scotton & Jake 1995, Sebba 1998). One means of achieving congruence at the morphological level is by means of morphological integration, which is a diagnostic feature listed in Table 2. Here we see that the morphological integration feature has a positive value in the insertion and congruent lexicalization patterns. However, this feature seems to apply to a differing extent to nouns versus verbs in our data, and to a differing extent to certain parts of verbs rather than others. In the case of switches to English nouns with plural marking, the English plural *-s* seems to be retained rather than a Welsh plural being used. This is illustrated by the use of the form *toys* in (5) in the Appendix and *toothpicks* in (7). There are various possible Welsh plural suffixes available (see e.g. Thorne 1993: 100-115) but none of these are used here. Thomas (1982:210) points out that Welsh plural suffixes do occur on words originating from English, like *roced* ‘rocket’ and *bom* ‘bomb’: he says that the Welsh plural forms as in *rocedi* and *bomiau* are likely to occur in more formal contexts, while the “imported” English plural forms as in *rocets* and *boms* would occur in the vernacular. He points out that this imported English plural can even be used with indigenous Welsh forms, as in *pregethwrs* ‘preachers’

as an alternative to *pregethwyr*, the indigenous plural. Thus when English plural nouns are switched into Welsh (and recall that these are nouns which would not be classified as loans as they are not listed in a Welsh dictionary) it seems that their plural marking is considered to be not only congruent with Welsh plural affixes, but even to count as a possible Welsh plural form.

The morphological marking of English verbs switched into Welsh follows a different pattern which depends on the part of the verb switched. If the verb is to be used in a way similar to an English infinitive, the English verb is usually switched with the addition of the suffix *-io*, as in *exercisio* in example (2) and *squeekio* in example (5) in the Appendix. The *-io* suffix in Welsh is a productive way of deriving both Welsh verbs from nouns (e.g. *llywio* ‘steer’ from *llyw* ‘helm’: see King 1993:132) and Welsh verbs from English verbs. (Bolonyai (2003) reports on a similar “denominal verbaliser” in Hungarian which is used both to derive Hungarian verbs from Hungarian nouns and Hungarian verbs from English verbs.) Verb forms like *llywio*, *exercisio* and *squeekio* are nonfinite, and as such are referred to as ‘verbnouns’ by Welsh grammarians because their distribution in some constructions is similar to that of nouns (see Borsley 1993). While Welsh has just this one non-finite form of the verb, English has both infinitives and participles: hence it can be argued that an English verb is made congruent with a Welsh verbnoun by the addition of the verbalizing suffix *-io*. However, no comparable process of morphological integration applies to past participles like *camouflaged* in example (1) in the Appendix: this could be because this category is already congruent with Welsh predicative adjectives, so does not need further adaptation. Present participles, however, like *second guessing* do not have an equivalent category in Welsh, where the nonfinite form of the verb (like those ending in *-io* sketched above) would normally be used in a periphrastic construction with the verb ‘to be’. Switching of present participles is in fact rare in the data, perhaps because of this lack of congruence, and it has been suggested² that *second guessing* in example (8) should be treated as a noun phrase, and the verb ‘to be’ (*fod*) as a copula rather than an auxiliary.

The role of congruence in determining which switches require morphological integration in Welsh-English data clearly requires further research. However, in the meantime we can speculate that where the equivalent syntactic categories in the two languages are congruent, morphological integration will not be necessary, as in the case of noun plurals and past participles with adjectival function. However, where congruence is lacking, morphological integration will serve to adapt the switch to a congruent category in the main or matrix language, as in the use of the *-io* suffix on English verbs.

5. Conclusion

A quantitative adaptation of Muysken’s typological approach to code-mixing has shown that insertion is the dominant pattern in a preliminary analysis of Welsh-English data from informal conversation. This is the pattern predicted on the structural grounds of typological distance between Welsh and English as well as the extralinguistic grounds of a postcolonial situation. Some evidence for congruent lexicalization was found as a secondary pattern, and is compatible with other features reflecting prolonged language contact.

Appendix: Examples of code-mixing from informal Welsh-English conversational data

- (1) Ia, mae o'n reit *camouflaged* yn dydi.
‘Yes, he is quite camouflaged isn't he’
- (2) Dwi'n cofio mynd i *exercisio* fo rhyw amser cinio
‘I remember going to exercise him one lunch time’
- (3) Mae pawb 'di mynd yn ddistaw. *Just as well*, fydd rhaid i ti neud yr ymchwil..
‘Everyone’s gone quiet. Just as well, you’ll have to do the research..’

2 I am grateful to Faith Owen for this suggestion

- (4) Mae nhw'n mynd *off* i chwilio amdan M
'They go off to look for M'
- (5) Powlen o ddwr, powlen o fwyd, dau neu dri gwahanol *chew, toys* i i *squeekio, toys* i er frathu, toys i gnoi.
'A bowl of water, a bowl of food, two or three different chews, toys to squeak, toys to bite, toys to chew'
- (6) Beauty is in the eye of the beholder, *ngwasi*.
'Beauty is in the eye of the beholder, you see'
- (7) Ti'n gwybod gwestai crand, pam bod na ddim handi andies efo dy *toothpicks* di?
'You know the grand restaurants, why aren't there any handy andies with your toothpicks?'
- (8) Mae'n siwr ar y funud fod on um *second guessing*, dydi, i jeckio fod o'n gweithio.
'Probably at the moment it's second guessing, you know, to check that it works'

References

- Bentahila, Abdelâli and Davies, Eirlys E. (1983). The syntax of Arabic-French code-switching. *Lingua* 59: 301-30.
- Berk-Seligson, Susan (1986). Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society* 15: 313-48.
- Bolonyai, Agnes (2003). English verbs in Hungarian/English codeswitching. Paper presented at the Fourth International Bilingualism Symposium, Arizona State University, Tempe, April 30-May 3 2003.
- Borsley, Robert D. (1993). On so-called verb-nouns in Welsh. *Journal of Celtic Linguistics* 2: 35-64.
- Jones, Kathryn (1993). "Dwisio hanner dwshin o wyau a phedwar beefburger": Rhai rhesymau dros drosi o u'n iaith i'r llall". *Barn* 368: 5-7.
- Jones, Kathryn (1995). Code-switching, intertextuality and hegemony: exploring change in bilingual discourse. In *Proceedings of the Summer School, Code-Switching and Language Contact*. Ljouwert/Leeuwarden: Fryske Akademy: 108-118. University of Wales Press.
- Jones, Kathryn (2000). Texts, mediation and social relations in a bureaucratised world. In Martin-Jones, M. and Jones, K. eds) *Multilingual literacies: Reading and writing different worlds*. Amsterdam: John Benjamins, 209-28.
- King, Gareth (1993). *Modern Welsh*. London: Routledge.
- Muysken, Pieter (2000). *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: CUP.
- Myers-Scotton, Carol (1993). Duelling languages: Grammatical structure in code-switching. Oxford: Clarendon Press.
- Myers-Scotton, Carol and Jake, Janice (1995). Matching lemmas in a bilingual language production model: evidence from intrasentential codeswitching. *Linguistics* 33: 981-1024.
- Poplack, Shana (1980). Sometimes I'll start a sentence in Spanish y termino en español. *Linguistics* 18:581-618.
- Poplack, Shana (1998). Contrasting patterns of code-switching in two communities. In Trudgill, P. and Cheshire, J. (eds) *The Sociolinguistics Reader. Volume 1: Multilingualism and Variation*. London: Arnold.
- Stenson, Nancy (1990). Phrase structure congruence, government, and Irish-English code-switching. In Hendrick, R. (ed.) *Syntax and Semantics, Volume 23, The Syntax of the Modern Celtic Languages*. San Diego: Academic Press: 167-197.
- Thomas, Alan (1982). Change and decay in language. In Crystal, D. (ed) *Linguistic controversies: Essays in linguistic theory and practice in honour of F.R. Palmer*. London: Arnold.
- Thomas, Alan (1992). The Welsh language. In Macaulay, D. (ed) *The Celtic Languages*. Cambridge: Cambridge University Press.

ISB4: Proceedings of the 4th International Symposium on Bilingualism

edited by James Cohen, Kara T. McAlister,
Kellie Rolstad, and Jeff MacSwan

Cascadilla Press Somerville, MA 2005

Copyright information

ISB4: Proceedings of the 4th International Symposium on Bilingualism
© 2005 Cascadilla Press, Somerville, MA. All rights reserved

ISBN 978-1-57473-210-8 CD-ROM
ISBN 978-1-57473-107-1 library binding (5-volume set)

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Press.

Ordering information

To order a copy of the proceedings, contact:

Cascadilla Press
P.O. Box 440355
Somerville, MA 02144, USA

phone: 1-617-776-2370
fax: 1-617-776-2271
sales@cascadilla.com
www.cascadilla.com

Web access and citation information

This paper is available from www.cascadilla.com/isb4.html and is identical
to the version published by Cascadilla Press on CD-ROM and in library binding.