

## Article

# English-Origin Verbs in Welsh: Adjudicating between Two Theoretical Approaches

Margaret Deuchar <sup>1,2,\*</sup> and Jonathan R. Stammers <sup>1,†</sup>

<sup>1</sup> Centre for Research on Bilingualism, Bangor University, College Rd., Bangor, Gwynedd LL57 2DG, UK; jon.stammers@gmail.com

<sup>2</sup> Department of Theoretical and Applied Linguistics, University of Cambridge, 9 West Rd., Cambridge CB3 9DP, UK

\* Correspondence: m.deuchar@gmail.com; Tel.: +44-(0)7947-805380

† These authors contributed equally to this work.

Academic Editors: Usha Lakshmanan, Osmer Balam and Tej K. Bhatia

Received: 16 October 2015; Accepted: 12 April 2016; Published: 25 May 2016

**Abstract:** In this paper we address the question of whether it is possible to compare two theoretical approaches to the same phenomenon or whether these should be considered incommensurable. We focus on two contrasting approaches to the identification of code-switching *vs.* borrowing by Poplack and Meechan [1] and Myers-Scotton [2,3]. For Poplack the distinction is based on linguistic integration and for Myers-Scotton on frequency. We show how what is a definition for one is a hypothesis for the other, and *vice versa*. Overcoming this apparent incommensurability requires a theory-independent approach in which we define the unit of analysis as “donor-language items” rather than switches or borrowings. Using this unit of analysis in the analysis of English-origin verbs in a Welsh corpus, we examine the assumptions behind the contrasting definitions of CS *vs.* borrowing. First we consider whether it is possible to identify linguistic integration in an unequivocal, categorical way and secondly whether linguistic integration is related to frequency of usage. We show that the identification of linguistic integration depends on the test used and that both frequency of usage and listedness play roles in the integration of English donor-language items in Welsh. In this way we argue that we achieve a theory-independent approach and go some way towards overcoming incommensurability.

**Keywords:** code-switching; borrowing; integration

## 1. Introduction

The term *incommensurability* is used by philosophers of science such as Kuhn [4] and Feyerabend [5] to argue that two competing theories may be ‘incommensurable’ in that their proponents may use different concepts and propose different research questions or hypotheses. This argument may be seen as a challenge to the work of the philosopher Karl Popper [6] whose work established falsificationism as a method for testing scientific theories. According to this approach, a theory could be compared with another in terms of which theory made the most correct predictions. Although this approach to comparing competing theories is assumed to be correct in much current scientific work, Kuhn and Feyerabend set out to challenge and question it.

According to Kuhn [4], for example, incommensurability is due to different research questions being posed by different theories, differing definitions of terms and what he calls “different worlds”. In relation to the latter Kuhn says “In a sense that I am unable to explicate further, the proponents of competing paradigms practice their trades in different worlds” [4] (p. 150). In what follows we shall see how differing definitions of terms mean that what is a hypothesis for one scholar is a definition for the other, and *vice versa*. This situation can easily lead to a perception of “different worlds”. We shall

show here how two theories of code-switching might have been identified by the likes of Kuhn and Feyereabend as incommensurable but how this problem may be overcome. The theories in question are Myers-Scotton's [2,3], Matrix Language Frame (MLF) approach and Poplack and Meechan's [1] variationist approach. The apparent incommensurability lies in the different approaches the proponents take to the distinction between borrowing and code-switching and in particular in the fact that what for one camp is a definition of the distinction forms a hypothesis for the other, and *vice versa*. So while for Myers-Scotton the definition of the two categories (borrowing and code-switching) is based on frequency of word tokens, for Poplack and Meechan it is based on whether or not the other-language word is integrated into the recipient language. However, for Myers-Scotton the extent of integration of the two categories (defined in terms of contrasting frequency) is a research question, while for Poplack the frequency of the two categories (defined in terms of contrasting integration) is the object of investigation. At first sight at least, the contrasting definitions of borrowings *vs.* code-switches alone make the two approaches seem to be incommensurable.

## 2. Code-Switching *vs.* Borrowing

While code-switching is understood in its clearest manifestation as the use of material from both of a bilingual's languages A and B in the same conversation, problems sometimes arise in determining whether a given item should be counted as a switch from language B or a borrowing into language A. For example, few would doubt that the word *restaurant* in English is a borrowing from French rather than a switch from French, if only on the grounds that it is used by English monolinguals. If it were used by English-French bilinguals, however, there might be some doubt as to whether the speakers were switching from English into their other language, French, when using it, or whether it should be considered a *bona fide* English word. This kind of problem is particularly acute in dealing with Welsh-English data, because all Welsh speakers also speak English. Only about 20% of the inhabitants of Wales are fluent Welsh speakers, but all of these are also exposed to and learn to speak English, whether at an early age at home or in the community. The issue of how to distinguish between code-switching and borrowing is both theoretical and practical. It is a practical issue for all code-switching theorists because its resolution determines what is and what is not included in the theory. It is also a theoretical issue insofar as various researchers have attempted to provide linguistically or psycholinguistically motivated rationales for distinguishing between the two concepts. There are psycholinguistic implications, for example, for assumptions about what is in a speaker's mental lexicon. It is no longer considered obvious (see, e.g., Pavlenko [7]) that bilinguals have two separate mental lexicons for their two languages, but most theorists assume that there are at least language-specific tags for lexical items from different languages, otherwise it is unclear how bilinguals could opt (as they clearly do) to speak only one language on specific occasions. So whereas code-switching is often viewed as the insertion by a speaker of an item from the mental lexicon of language A among other items which are from the mental lexicon of language B, a borrowed item would be one which used to belong to the lexicon of language B, but which over time has been added to the lexicon of language A, like 'restaurant' in English. The issue of the dividing line between switches and borrowings applies particularly to lone 'other-language' words, or single words from language B being inserted in language A. The larger the stretch of 'other-language' material, the less controversial is the identification of this material as a switch. Least controversial is thus intersentential switching, where the 'other-language' material is an entire sentence or clause. This is illustrated in example (1) from our Spanish-English data<sup>1</sup>:

---

<sup>1</sup> See [8] and Deuchar, Davies, Herring, Parafita Couto and Carter [9].

1. NO THEY DON'T KILL PEOPLE THEY DON'T KILL ellos no  
PRON.3PL.M NEG  
mataban a su propia gente<sup>2</sup>  
kill.IMP.3PL to POSS.3SG own.F people  
'no, they don't kill people, they don't kill ... they didn't kill their own people' (herring 7)<sup>3</sup>

In example (1) the first two clauses are in English ((no they don't kill people) (they don't kill)) while there is a switch to Spanish for the third clause ((ellos no mataban a su propia gente)). No analyst would presumably wish to argue that the third clause is a borrowing into English rather than a switch into Spanish. In example (2), however, from our Welsh-English data, we see an example of a switched phrase, WIDE-ANGLE LENSES, which is still relatively unproblematic to classify as a switch, but in the same utterance there are two examples of single 'other-language' words, '*emphasize*' and '*foreground*', which are more difficult to classify.

2. *pan* *dach* *chi* 'n *defnyddio* WIDE-ANGLE LENSES  
when be.2PL.PRES PRON.2PL PRT use.NONFIN wide-angle lenses  
*dach* *chi* 'n EMPHASIZE-*io* 'r FOREGROUND  
be.2PL.PRES PRON.2PL PRT emphasize.NONFIN DET Foreground  
'when you use wide-angle lenses, you emphasize the foreground.' (fusser 17)

How do we decide whether the words *emphasize* and *foreground* in example (2) are switches into English or English borrowings that have been incorporated into the Welsh language? And can this decision be made according to linguistic criteria? Poplack and Meechan [1] consider that it can, although they recognise that their approach is controversial. Summarising their view of the issue, they say the following:

In virtually all bilingual corpora empirically studied, mixed discourse is overwhelmingly constituted of lone elements...of one language embedded in...another. The status of these items is notoriously ambiguous. They may be codeswitches or borrowings... They are at the heart of a fundamental disagreement among researchers about data... At one end of the spectrum, where lone items are defined as codeswitches, researchers tend to consider the relationship between languages...(as) asymmetrical... Where lone items are classified as borrowings...both languages are postulated to play a role in constraining codeswitching.

—Poplack and Meechan [1] (pp. 127–128)

As we shall see, Poplack and Meechan associate themselves more with the second view, while Myers-Scotton is associated with the first.

### 3. Verbs in Welsh

As Stammers [11] (p. 81) reports, Welsh has two main types of verbal constructions in finite clauses, synthetic and periphrastic. The synthetic type is illustrated in (3) below, where the verb *ddigwyddodd* is a finite form of the nonfinite form *digwydd* 'to happen'. The suffix *-odd* identifies it as third person singular in the past tense.

3. *dyna* *be* *ddigwyddodd* *wrth gwrs*.  
there what happen.3S.PAST of course  
'that's what happened, of course' (fusser 4)

<sup>2</sup> In illustrative examples, as in tables, ENGLISH WORDS are given in capitals, *Welsh words* in italics, WORDS THAT COULD BE EITHER WELSH OR ENGLISH are given in capital italics, and words in Spanish are underlined.

<sup>3</sup> Material in parentheses following examples indicates the name of the recording: see [10].

In contrast, example (4) illustrates the periphrastic type of construction where the non-finite form *ddigwydd* ‘happen’ is in construction with a past tense, finite verb of the verb *gwneud*, which functions here as a dummy auxiliary.

4.    *a*            *be*            *wnaeth*            *ddigwydd*  
       and        what      do.3S.PAST      happen.NONFIN  
       ‘and what happened?’            (fusser 19)

In the synthetic construction illustrated in (3) the main verb is finite whereas in the periphrastic construction illustrated in (4) the main verb is nonfinite, appearing in construction with an auxiliary or ‘light’ verb. Periphrastic constructions are more common than synthetic types in informal speech, although both appear in our corpus. Whereas the verb *digwydd* ‘happen’ illustrated above has no verbal suffix in its nonfinite form, other Welsh verbs which are derived from nouns and adjectives appear with suffixes including *-u*, *-o*, *-io*, *-i*, *-a* and *-au* in their nonfinite forms. King [12] (p. 131) provides some examples including *pleidleisio* with *-io* ‘to vote’ from the noun *pleidlais* ‘vote’, *talw* with *-u* ‘to pay’ from the noun *tâl* ‘pay’, and *rhyddhau* with *-au* ‘to free’ from the adjective *rhydd* ‘free’. King [12] (p. 132) furthermore points out that it is the *-io* suffix (or *-o* in South Wales) which is commonly used to derive nonfinite verb forms from English words, as in, e.g., *parcio* (from park) and *stopio* (from stop). A use of *stopio* is illustrated in example (5) below from our corpus. In this example *stopio* is used in a periphrastic construction with the finite auxiliary *dw*<sup>4</sup> (‘be’, first person, present tense):

5.    *dw*                    *'m*            *yn*            *stopio*                    *nhw*  
       be.1S.PRES      NEG      PRT      stop.NONFIN          PRON.3PL  
       ‘I’m not stopping them.’                    (fusser 19)

While *parcio* and *stopio* are established verbs in Welsh which can be found in Welsh dictionaries, the *-io* suffix is very productive in ‘coining’ new ‘mixed’ nonfinite forms of verbs with English stems and Welsh suffixes, like *EMPHASISE-io* in example (2) above and *RECOGNISE-io* in example (6) below. Neither *EMPHASISE-io* nor *RECOGNISE-io* occur in Welsh dictionaries.

6.    *dw*                    *i*                    *'n*            *RECOGNISE-io*            *fan*                    *hyn*  
       be.1S.PRES      PRON.1S      PRT      recognise.NONFIN      place                  this  
       ‘I recognise this place!’                    (fusser 27)

These mixed verbs almost always appear in periphrastic rather than synthetic constructions in our data, although Stammers [11] (p. 88) reports the very occasional example of a mixed verb in a synthetic construction, such as that in (7) below:

7.    *pan*                    *CRASH-odd*            *fy*                    *cyfrifiadur*  
       when              crash.3S.PAST          POSS.1S          computer  
       ‘when my computer crashed’                    (fusser 14)

Note that the periphrastic constructions with mixed nonfinite verbs and finite Welsh auxiliaries are quite different from the type of construction occurring in other language pairs (see e.g., [13]), where an inflected light verb occurs with a bare other-language item in a novel construction compared to that of unmixed verbs. Instead, the mixed English-Welsh verb appears in exactly the same periphrastic constructions as unmixed verbs, *i.e.*, where a finite auxiliary is combined with a non-finite form of the main verb. If the main verb has been derived from another form, e.g., a Welsh noun, Welsh adjective or English verb as described above, it has a verbalising suffix as outlined.

<sup>4</sup> In periphrastic constructions the verb ‘to be’ is used in the present tense as in (4) while ‘to do’ is used in the past tense as in (3).

#### 4. Review of the Relevant Literature by the Protagonists

In this section we will summarize the work of Poplack and Meechan [1], representing one end of the spectrum mentioned above, and the work of Myers-Scotton [2], representing the other. We will use these two studies in this paper to illustrate the problem of incommensurability and its solution.

Poplack and Meechan focus in their paper on how to determine the status (code-switch or borrowing) of “lone other-language items” in conversations involving the use of more than one language [1] (p. 128). As they point out, the status of such items when viewed in isolation is ambiguous, but they use a variationist approach to achieve disambiguation. The variationist approach is derived from variation theory as pioneered by Labov (e.g., [14]) and developed by others (e.g., Sankoff [15]). A key feature is that it recognizes that all languages (including those in contact in bilingual situations) are inherently variable and exploits this fact to determine whether a lone other-language item is a switch or borrowing. Poplack and Meechan’s method involves a quantitative analysis of relevant morphosyntactic patterns in the two contact languages when they are used monolingually or without mixing, and then a comparison of the results with the morphosyntactic patterns in mixed discourse in which a lone item from another ‘donor’ language is being used in a recipient language. If the patterns in which the donor-language item is being used are similar to those used in the recipient language when it is unmixed, they consider the item to be linguistically integrated and therefore to be a borrowing rather than a switch. If on the other hand the morphosyntactic patterns are more similar to those of the unmixed donor language, then the lone other-language item is classified as a switch. This classification can be further checked by comparing the patterning of the lone other-language item with that found in what they consider to be unambiguous (multiword) switches and with that of well-established loanwords. If the item patterns with well-established loanwords it is likely to be a borrowing, whereas if it patterns with its occurrence in multiword switches then it is likely to be a switch. As an example, Poplack and Meechan cite work by Buzhak-Jones [16] showing that “English-origin nouns in otherwise Ukrainian discourse are inflected with Ukrainian case markers following the same system speakers use to inflect Ukrainian nouns in Ukrainian discourse” [1] (p. 133). Poplack and Meechan [1] (p. 136) report that their research shows that in fact, lone other-language items almost always pattern with items in the recipient language, even if they are infrequent or ‘nonce’<sup>5</sup>, and they thus find that borrowings are overall more frequent than switches.

While Poplack and Meechan [1] use linguistic integration of other-language items as a way of identifying borrowings in contrast to switches, Myers-Scotton [2] expects all other-language items, whether switches or borrowings, to be linguistically integrated to at least some extent. This is because of her assumption that the two languages in contact have an asymmetrical relationship, with one, the ‘matrix language’ or ML, providing the morphosyntactic frame and the other, the ‘embedded language’ or EL, having a smaller role. Since providing the morphosyntactic frame involves supplying the grammatical morphemes or ‘closed class’ items, it is not surprising to Myers-Scotton that any content or ‘open class’ morphemes provided by the EL will be linguistically integrated in the frame of the ML by being juxtaposed, for example, by grammatical morphemes. She illustrates this with some data from a Swahili-English utterance (see [2], p. 4) illustrated in example (8) below, where italics are used to indicate an English-origin word:

8.   hata    siku    hizi    ni-me-*decide*    kwanza   kutumia   sabuni   ya   miti  
      even   days   these   1s-PERF-decide   First   to use   soap   of   stick  
      ‘Even these days I have decided first to use bar soap’

In example (8) above, the matrix language is Swahili, an agglutinating language which uses preverbal particles to indicate grammatical functions. In this example the English verb *decide* is

<sup>5</sup> Stammers and Deuchar [17] (pp. 642–643) use their analysis of English-origin verbs in Welsh to argue that the category of ‘nonce borrowings’ is redundant.

used with two preverbal particles which indicate the person of the subject and the tense of the verb. It seems likely that if this utterance were part of a variationist analysis by Poplack and Meechan, they would identify the English verb *decide* as following Swahili rather than English morphosyntactic patterns, leading to its classification as a borrowing rather than a switch. For Myers-Scotton, the distinction between switches and borrowings is not so central to her theory, but since her approach focuses on code-switching [2] (p. 204) she needs to distinguish switches from borrowings for practical reasons. She does this on the grounds of frequency, arguing that switches should not occur more than twice in a relatively large corpus [2] (p. 205). Although Myers-Scotton does not use the degree of linguistic integration to differentiate switches from borrowings, she does expect borrowings (defined in terms of frequency) to receive a greater degree of ‘peripheral’ morphological integration than switches. Her distinction between central and peripheral morphological integration [2] (pp. 183–184) is necessary since she expects both switches and borrowings to receive central morphological integration as predicted by her matrix language frame theory. She exemplifies this with reference to Swahili-English code-switching, where English verbs inserted in Swahili receive preverbal prefixes, as exemplified in (8) above, and these prefixes are considered central morphologically because they function to indicate subject-verb agreement and tense. However, she says that inserted English verbs are rarely integrated by the addition of a final vowel characteristic of Swahili verbs, which she says carries a low functional load and is thus considered part of the peripheral morphology of Swahili.

## 5. The Contrasting Approaches Applied to Mixed Welsh-English Verbs

The contrasting approaches to borrowing *versus* switching outlined above mean that an item classified as a borrowing by one researcher may be a switch for the other, and *vice versa*. This can be illustrated by the case of English-origin verbs used in otherwise Welsh utterances. In such utterances, illustrated in examples (5) and (6) above, Welsh is the matrix language in Myers-Scotton’s terms since it provides the morphosyntactic frame of the utterance. Examples (5) and (6) are repeated as (9) and (10) for convenience, so that we can use them to compare the approaches of Poplack and Meechan on the one hand and Myers-Scotton on the other.

- |     |  |                     |                  |  |                                       |
|-----|--|---------------------|------------------|--|---------------------------------------|
| 9.  | <i>dw</i><br>be.1S.PRES<br>'I'm not stopping them.'  | <i>'m</i><br>NEG    | <i>yn</i><br>PRT | <i>stopio</i><br>stop.NONFIN             | <i>nhw</i><br>PRON.3PL<br>(fusser 19) |
| 10. | <i>dw</i><br>be.1S.PRES<br>'I recognise this place!' | <i>i</i><br>PRON.1S | <i>'n</i><br>PRT | RECOGNISE- <i>io</i><br>recognise.NONFIN | <i>fan</i><br>place<br>(fusser 27)    |
|     |  |                     |                  |  | <i>hyn</i><br>this                    |

Where the two approaches would probably agree (assuming they had access to the whole *Siarad* corpus at bangortalk.org [10]) would be in classifying *stopio* in (9) as a borrowing. Poplack and Meechan would presumably consider that *stop* in *stopio* is a borrowing on the grounds that it is linguistically integrated into Welsh by the addition of the suffix *-io* and its use in a typically Welsh construction where the finite verb *dw* comes first in the utterance. Myers-Scotton would agree that *stop* in *stopio* can be classified as a borrowing, but this decision would be made on the grounds not of its linguistic integration, but on the grounds that *stopio* occurs 88 times in the corpus. However, turning to *recognise-io* as in example (10), its integration into Welsh by means of the derivational *-io* suffix would probably lead Poplack and Meechan [1] to classify it as a borrowing, but since *recognise-io* is infrequent in our data Myers-Scotton’s approach would doubtless consider *recognise-io* to be a switch on the grounds of the low frequency<sup>6</sup> of its usage and the fact that it is not listed in any dictionary of Welsh.

<sup>6</sup> *recognise-io* occurs only once in our corpus of about half a million words.



As indicated above, not only do the definitions of borrowing used by Poplack and Meechan [1] and Myers-Scotton [3] differ, but also the hypotheses they propose. Having defined borrowings as linguistically integrated items, Poplack and Meechan predict that it will be more common for other-language items to be integrated than not, and thus that lone borrowings will be more frequent than lone switches in the data. For Myers-Scotton, however, frequency is part of the definition of borrowings rather than being a hypothesis to test. In addition, just as Poplack and Meechan's hypothesis is Myers-Scotton's definition, the reverse is also true since Myers-Scotton [2] hypothesises that borrowings will have a higher degree of linguistic integration than switches. The inverse relationship between Poplack and Meechan's and Myers-Scotton's definitions and hypotheses is illustrated in Table 1 below. Comparing the italicized material in the table will show how Poplack and Meechan's definition is a hypothesis for Myers-Scotton, while the underlined material shows how the reverse is the case. This suggests that the two approaches are incommensurable at present.

**Table 1.** Definitions *vs.* hypotheses: Poplack and Meechan *vs.* Myers-Scotton.

Lone Other-Language Items	Definitions of Borrowings/Switches	Hypotheses
Poplack and Meechan	<i>Linguistically integrated/not linguistically integrated</i>	<u>Borrowings more frequent than switches</u>
Myers-Scotton	<u>More frequent/less frequent</u>	<i>Borrowings more integrated than switches</i>

To attempt to overcome the incommensurability between the two approaches illustrated in Table 1, we shall dispense with the notions of switches and borrowings with their conflicting definitions and adopt the more neutral notion of *donor-language items* (a term also used by Poplack and Meechan) as our units of analysis instead. Next, in order to examine Poplack and Meechan's view that borrowings can be identified in terms of their linguistic integration, we shall consider whether it is possible to identify linguistic integration in an unequivocal way. The data used will be lone English verbs used in otherwise Welsh discourse. We will report on the results from three tests of the degree of linguistic integration of these English items into Welsh as follows:

1. Morphological: does each English verb inserted in Welsh have a Welsh derivational suffix *-(i)o*?
2. Syntactic: does each English verb inserted in Welsh appear in both synthetic and periphrastic constructions in a similar way to Welsh verbs?
3. Morphophonological: does mutation apply to English verbs inserted in Welsh in a similar way to its application to Welsh verbs?

If linguistic integration is indeed categorical, we would expect to find that the results of all three tests is the same.

## 6. Tests of Linguistic Integration

### 6.1. Morphological Integration

Stammers and Deuchar [17] (p. 635) report on the use of three transcribed conversations<sup>7</sup> from the Bangor *Siarad* corpus. All verbs from these transcripts that were of English origin were extracted and classified as morphologically integrated in Welsh or not. There was a total of 184 tokens or 80 types<sup>8</sup>. Of these 184 tokens, 179 (97.3%) were morphologically integrated by means of a Welsh derivational suffix. For Poplack and Meechan [1], this fact would presumably lead them to categorize these 179 tokens as borrowings. Of course, if we look at the frequency of occurrence of these items they are far from uniform, including highly frequent loans found in the Welsh dictionary like *trio* 'try' (5 tokens), *ffonio* 'phone' (6 tokens) as well as low frequency verbs not found in any Welsh dictionary

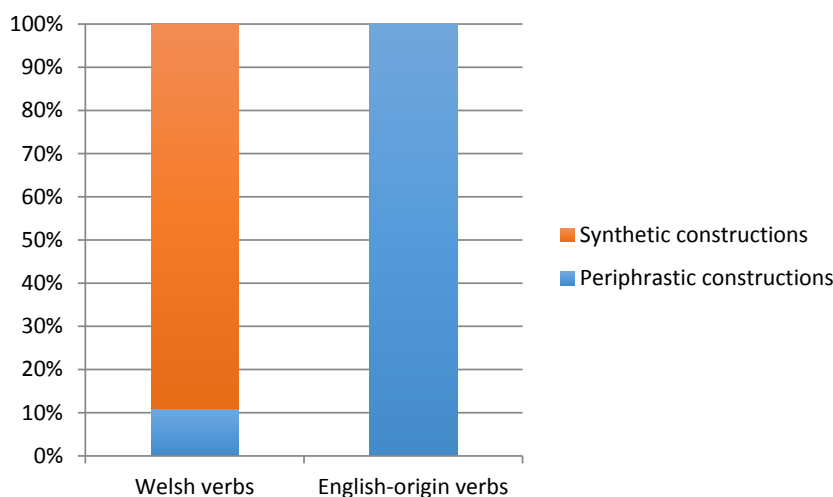
<sup>7</sup> davies2, fusser29, stammers4, in total just under 20,000 words.

<sup>8</sup> All the verbs are listed in Table 3, Stammers and Deuchar [17] (p. 637).

such as *stare-io* ‘stare’ (1 token) and *freak-o* (1 token). But this information about frequency would not be important for Poplack, who has argued elsewhere<sup>9</sup> that integration is abrupt<sup>10</sup> and that there is no relation with frequency<sup>11</sup> in their French-English data. For Myers-Scotton, the decision to draw the line between code-switching and borrowing would in contrast be based on frequency of usage rather than on the uniform morphological integration which we have discovered using our first test. We assume this would be considered by Myers-Scotton to be a central type of integration (see above), which would affect all donor-language items equally, regardless of any distinction drawn between code-switching and borrowing by investigators.

## 6.2. Syntactic Integration

One way of examining the syntactic integration of English-origin verbs was reported by Stammers in his comparison of their occurrence in synthetic *versus* periphrastic constructions with the distribution of native Welsh verbs in those two types of construction [11] (p. 82). Examples of the two types were given in (3) and (4) above. Stammers [11] extracted all English-origin verbs from finite clauses in the three transcripts used in the previous analysis, ending up with 111 tokens. Their distribution in verbal constructions was compared with a sample of 300 tokens of native Welsh verbs. As shown in Figure 1, 11% of native Welsh verb tokens appeared in synthetic constructions, compared with no English-origin verbs whatsoever, since 100% of the English-origin verbs appeared in periphrastic constructions. Stammers [11] considered the possibility that this dramatic difference might be due to the small size of his sample, so he went on to investigate a large sample of both English-origin and native verbs. He searched the entire Siarad corpus for all tokens of English-origin verbs in synthetic constructions, and found 35. He estimated that “35 tokens of synthetic constructions represents approximately 1.3% of all English-origin main verb finite clauses in the corpus” Stammers [11] (p. 88), in contrast with the results of a larger sample of 1082 tokens of which 12.6% were in synthetic as opposed to periphrastic constructions. He concluded that the distribution of native Welsh and English-origin verbs in syntactic constructions is not identical, so we may guess that this would lead Poplack and colleagues to conclude that the English-origin verbs cannot unequivocally be considered to be borrowings.



**Figure 1.** Proportion of Welsh and English verbs in periphrastic *vs.* synthetic constructions.

<sup>9</sup> Poplack and Dion [21].

<sup>10</sup> “Integration occurs abruptly, at the first mention of the nonce item” Poplack and Dion [21] (p. 308).

<sup>11</sup> “With respect to the criterion of plural marking . . . there is no evidence to suggest that integration increases as nonce nouns gain in frequency” Poplack and Dion [21] (p. 291).



Table 2 summarizes the results of our tests of linguistic integration so far. As it shows, the English-origin verbs are almost all integrated according to the morphological test, whereas none are integrated according to the syntactic test. This suggests that identifying linguistic integration, which is crucial for Poplack and Meechan’s definition of borrowing, may depend on the test one uses and therefore is not an unproblematic criterion.

**Table 2.** Results of first two tests of linguistic integration.

English-Origin Verbs in Welsh	Proportion Showing Integration
Morphological test	97%
Syntactic test	0%

Perhaps the problem is that neither of the first two sets uses a sufficiently sensitive measure of linguistic integration. With the third test, however, we made use of a morphophonological process found in Welsh, soft mutation, to investigate the integration of English verbs into Welsh in a more subtle manner.

### 6.3. Morphophonological Integration

Soft mutation is a variable morphophonological process which affects certain consonants in the initial position of words in specific environments, for example following prepositions. Initial voiceless stops become voiced and voiced stops becoming fricatives, as outlined by Stammers [11] (p. 89) and Stammers and Deuchar [17] (p. 638). Table 3 below shows the changes that are undergone in word-initial consonants when they are subject to soft mutation. Soft mutation, unlike derivational morphology described above, is not one of the morphosyntactic phenomena which Myers-Scotton would predict to come from the matrix language (she would presumably class it as a ‘peripheral’ process), and hence it might be accepted by proponents of all approaches as a possible measure of linguistic integration.

**Table 3.** Initial consonants changes in Soft Mutation in Welsh<sup>12</sup>.

Initial Consonant (Phonetic)	p	t	k	b	d	ɫ	ɾ	m	g
Initial consonant (orthographic)	p	t	c	b	d	ll	rh	m	g
	↓	↓	↓	↓	↓	↓	↓	↓	↓
Mutates to (phonetic)	b	d	g	v	ð	l	r	v	
Mutates to (orthographic)	b	d	g	f	dd	l	r	f	(dropped)

Stammers and Deuchar [17] (p. 638) distinguish between lexically and syntactically triggered mutation. They state that “in lexically triggered soft mutation, the non-finite verb is directly preceded by a preposition, clitic or other particle causing soft mutation”. They give an example (reproduced as example (11) below) where the preverbal particle *i* (translated as ‘to’ in English) triggers soft mutation of the initial consonant of the following verb *costio*, which becomes *gostio*.

11. *WELL mae mynd i gosti pres*  
 well be.3s.PRES go.NONFIN to cost.NONFIN money  
 ‘well, it’s going to cost money’ (fusser 6)  
 Stammers and Deuchar [17] (p. 638), example (10)

Other environments involving lexically triggered mutation include where the non-finite verb is preceded by a second person or third person masculine possessive pronoun, the preposition *am* ('for/about'), *ar* ('on/about to'), *gan* ('by/while/with').

<sup>12</sup> Based on Table 4 in [17] (p. 638).

According to Stammers and Deuchar [17], in syntactically triggered soft mutation the non-finite verb is expected to mutate because of its position in the clause following the grammatical subject. This is illustrated in example (12) below:

12.    *wnest*                *ti*                *drio?*  
       do.2s.past        pron.2s        try.nonfin  
       ‘did you try?’        (stammers 5)

Stammers and Deuchar [17] (p. 639), example (14)

As they explain, the verb *drio* in example (12) is actually a mutated version of *trio* (‘try’). Stammers and Deuchar provide additional examples of environments for mutation, including one where mutation does not apply as expected. Although soft mutation is the most robust mutation type in Welsh (cf. Comrie) [18] (p. 81) its application is still subject to variation even in Welsh words. (cf. Ball and Müller) [19] (p. 256). This variation is advantageous for our analysis since it provides us with a fine-grained measure to compare the integration of English-origin verbs with the level of mutation found in native Welsh verbs, following the methodology advocated by Poplack and Meechan [1].

Stammers and Deuchar [17] describe how verbs used in Welsh periphrastic constructions can be divided into three categories: (1) native Welsh; (2) English-origin verbs listed in a dictionary of Welsh (‘listed’, e.g., in Thomas [20]) and (3) English-origin verbs not listed in a dictionary of Welsh (‘unlisted’). The following are examples of verbs<sup>13</sup> in the three categories:

- (I) Native Welsh verbs: Regular native verbs ending with the *-(i)o* suffix, e.g., *cofio* (to remember), *defnyddio* (to use), *cwyno* (to complain), *pwysu* (to push), *cneifio* (to shear), *treiglo* (to mutate), *twtio* (to tidy).
- (II) Listed English-origin verbs: Verbs of English origin ending with the *-(i)o* suffix and found listed in a dictionary of Welsh, e.g., *trio* (to try), *cario* (to carry), *clirio* (to clear), *dreifio* (to drive), *clarifffio* (to clarify), *pinsio* (to pinch), *bargeinio* (to bargain), *manejio* (to manage), *tsiecio* (to check), *cidnapio* (to kidnap).
- (III) Unlisted English-origin verbs: Verbs of English origin ending with the *-(i)o* suffix but not found listed in any dictionary of Welsh, e.g., TEXT-*io*, DOWNLOAD-*io*, BRIEF-*io*, QUOTE-*io*, BULK-*io*, CONNECT-*io*, BABYSIT-*io*, DECORATE-*io*, CONCENTRATE-*io*, MOLLYCODDLE-*io*, POWER-WALK-*io*.

Note that these last two categories are distinguished here in their orthographic representation, following the conventions used in transcribing the corpus. The reason for distinguishing not only between Welsh-origin and English-origin verbs but also between ‘listed’ and ‘unlisted’ English verbs in an analysis of linguistic integration is that it allowed us to determine whether in addition to frequency, there is a factor of ‘listedness’ which influences linguistic integration (cf. Muysken) [22] (p. 71). The analysis involved extracting all of the non-finite verb tokens found in the *Siarad* corpus that (i) ended in the *-(i)o* verbalising suffix; (ii) began with a consonant susceptible to soft mutation (subject to certain exclusions); and (iii) occurred in an environment where soft mutation could be expected to apply. Each of a total of 506 tokens was classified according to whether or not mutation actually applied.

The 506 tokens identified for the analysis (159 types) were an exhaustive selection of tokens of regular verbs meeting the criteria for mutation to occur and ending with the *-(i)o* suffix. This means that each began with a consonant which was subject to mutation, and each occurred in an environment where mutation was predicted to occur. For each token, we noted whether or not mutation had actually occurred. One-hundred forty-three tokens of native verbs were identified, or 44 types, of which an example is *defnyddio* (to use; either occurring with an initial [d] or in its mutated form with an initial

<sup>13</sup> Note that these are non-finite forms which occur in periphrastic constructions with finite forms of Welsh auxiliary verbs.

fricative [ð] as *ddefnyddio*). 302 tokens of listed English-origin verb were identified, or 81 types, of which an example is *trio* (either occurring with an initial voiceless stop [t] or in its mutated form with an initial voiced stop [d] as *drio*), and an example of an unlisted English-origin verb from the 61 tokens (34 types) found is *COPE-io* (either occurring with an initial voiceless stop [k] or in its mutated form with an initial voiced stop [g] as *GOPE-io*). The results of our mutation analysis are shown in Figure 2.

Figure 2 shows that there are differences between the three categories of verbs in terms of their behaviour in environments where mutation is expected. Native Welsh verbs show a 73% rate of mutation in environments where this is expected while English-origin verbs not listed in the dictionary show the reverse pattern: an 84% rate of non-mutation in environments where it is expected. The intermediate category of English-origin verbs found in the dictionary shows an intermediate pattern: The majority of tokens (66%) are mutated, meaning that they pattern more like the native Welsh verbs than the English-origin verbs not in the dictionary. These results show that integration measured in this way is not ‘abrupt’ as suggested by Poplack and Dion [21] but that listed English-origin verbs are considerably more integrated than unlisted English-origin verbs.

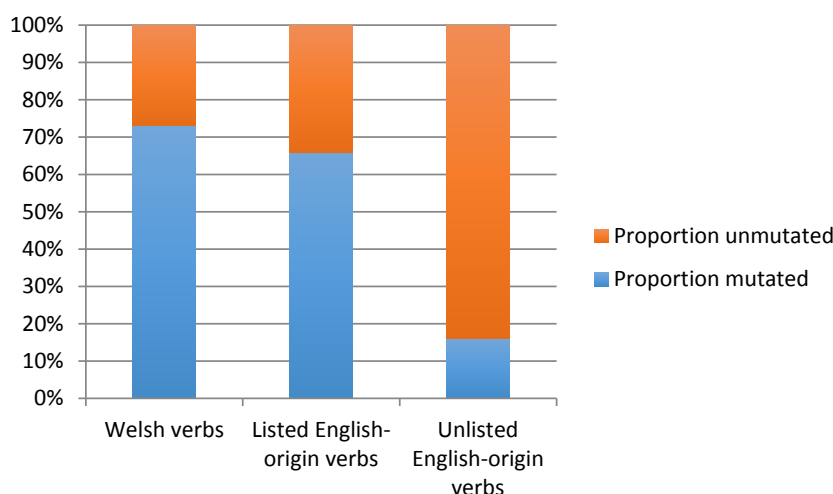


Figure 2. Application of mutation to three categories of verb.

## 7. Summary of Results of Applying Three Tests of Integration

Table 4 below summarises the results of the application of the three tests of linguistic integration.

Table 4. Results of application of three tests of integration.

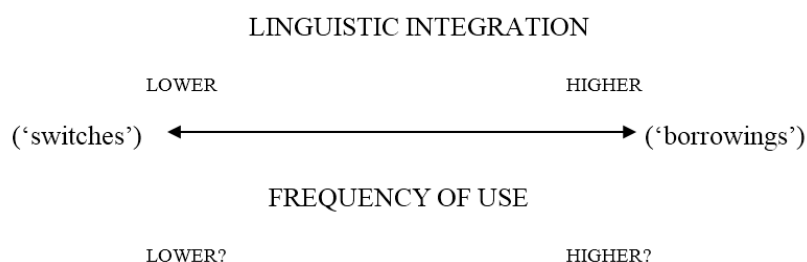
English-Origin Verbs in Welsh	Proportion Showing Integration
Morphological test	97%
Syntactic test	0%
Morphophonological test	66% of listed items; 16% of non-listed items

Table 4 shows that the three tests of linguistic integration have very different outcomes. If they were used to identify borrowings *versus* switches following Poplack and Meechan’s [1] approach, each test would lead to a very different classification of the same item.

## 8. Role of Frequency

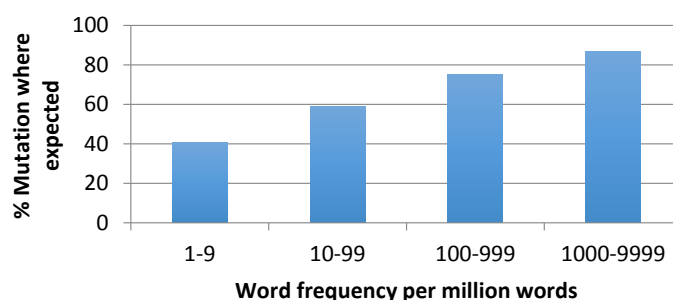
Stammers and Deuchar [17] raise the question of how we can explain the much lower level of integration (16%) of unlisted English verbs than the integration of listed English verbs (66%). In line with recent trends in linguistics (*cf.* Bybee [23]) they investigate the importance of frequency as a factor in accounting for this difference. They consider the possibility that the higher the frequency of

use of an item, the higher its linguistic integration may be. Figure 3 below is a representation of the hypothetical relation between the integration and frequency of donor-language items.



**Figure 3.** Hypothetical relationship between linguistic integration of donor-language items and frequency of use.

We therefore wondered whether frequency of use might be related to integration measured by the application of mutation. In order to investigate this relationship we next took all the verb types from all three categories included in our soft mutation analysis and calculated how many tokens of each type occurred in the corpus as a whole. We then grouped the verb types into four frequency bands based on orders of magnitude of the frequency per million words as found in the corpus, so that, for example, the first category included 79 verb types occurring one to four times in the corpus, or 1–9 times per million words. The third category included seven verb types occurring 46–450 times in the corpus, or 100–999 times per million words. The details of the categories and their relation to the application of mutation where expected can be seen in Figure 4.



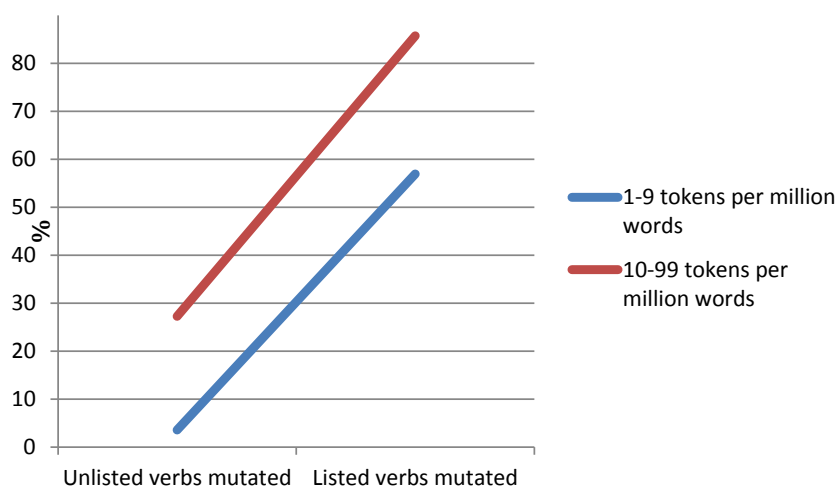
**Figure 4.** Results for mutation rate by word frequency grouping<sup>14</sup>.

As can be seen in Figure 4, the more frequent the verb, the more likely it is that mutation will apply in the expected environments. In fact, a remarkably clear and definite relationship is observed between overall frequency and rate of mutation where expected. Items of frequency one to nine per million only mutate in a small minority of instances (34.7%), whereas items of frequency 1000 or more per million mutate in the vast majority of instances (89.6%), with the mutation rate of intermediate categories increasing in line with the frequency bands. The relationship is a log-linear (rather than a linear) one, hence the groupings 1–9, 10–99, 100–999 and 1000–9999 per million words. A strong correlation of 0.99 between the rate of mutation and frequency was found using logarithmic values.

In Figure 2 we showed the rate of mutation of verbs in three categories, native Welsh, listed English verbs and unlisted English verbs. Stammers and Deuchar [17] (p. 642) report no significant difference in the application of mutation between the native Welsh and listed English verbs, but they did find a difference between these two categories and the unlisted verbs, where as we can see in

<sup>14</sup> Based on Figure 2 in Stammers & Deuchar [17] (p. 640).

Figure 2, mutation applies less. Thus not only frequency but also listedness appears to have an effect on the application of mutation. Figure 5 shows the proportion<sup>15</sup> of listed and unlisted verbs showing mutation at two comparable levels of frequency (1–9 and 10–99 per million words).



**Figure 5.** Application of mutation to listed and unlisted verbs of two levels of frequency.

In Figure 5 the listed verbs include both Welsh-origin and English-origin listed verbs, while the unlisted verbs are all English-origin. The Figure shows a similar relation between frequency and the application of mutation for both categories of verbs, as discussed above, but also shows that the frequency of mutation is lower for unlisted verbs. According to calculations by Stammers [11] (p. 116), this is not just an effect of frequency alone, but there is an interaction between frequency and listedness. In particular, listedness is a better predictor of the application of mutation for unlisted English-origin verbs only.

## 9. Discussion

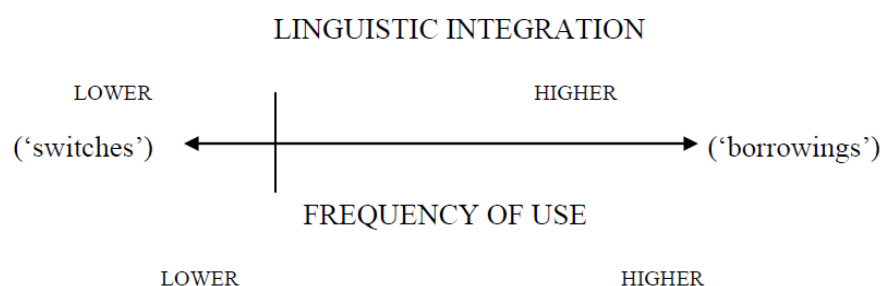
Previous attempts to evaluate competing theoretical approaches to code-switching *versus* borrowing have suffered from the fact that these terms have different definitions for different researchers. Furthermore, we have shown that what is a definition for one researcher may be a hypothesis for another. As shown in Table 1, Poplack and Meechan differ from Myers-Scotton in that they distinguish code-switching from borrowing according to the criterion of linguistic integration: a donor-language item which is linguistically integrated into the recipient language counts as a borrowing whereas an unintegrated item is seen as a switch. For Myers-Scotton, however, it is frequency rather than integration which distinguishes borrowings from switches. For her, rather than integration being a defining criterion it is the subject of a hypothesis which proposes that borrowings will be more integrated than switches. Myers-Scotton's hypothesis is thus equivalent to Poplack and Meechan's definition, and the reverse is also the case. Frequency is a criterion for Myers-Scotton but the subject of a hypothesis for Poplack and Meechan.

We have argued that the apparent incommensurability of these two approaches can be overcome by using a theory-neutral unit of analysis, donor-language items, and by testing the assumptions underlying the contrasting approaches. We first considered whether it is possible to identify linguistic integration in an unequivocal, categorical way, and found that different tests led to different results. However, using soft mutation as a particularly sensitive test of integration and comparing English-origin with native items, we found that integration by mutation was a matter of degree.

<sup>15</sup> Based on raw figures in Stammers [11] (p. 109).

Investigating then whether the extent of integration was related to frequency of usage, we found a strong relationship: the more frequent in our corpus a particular item (whether donor or native), the more likely it was to be mutated. In addition, listedness had an additional effect in that a listed verb (whether Welsh or English) was more likely to be mutated than an unlisted verb.

We can now return to the question of whether it makes sense to divide donor-language items into two categories, code-switches and borrowings. Is there a definable category of code-switches that we can distinguish from borrowings on the basis of one or more criteria? We have seen how integration as measured by the application of mutation is far from abrupt or instantaneous, but that it is related to both frequency and listedness. We have seen that items that are both low in frequency and low in degree of integration are also items that are not currently listed in the dictionary. These items we call switches, which thus do seem to be distinct from borrowings because of their unlistedness status. Note that while we have used inclusion in a dictionary<sup>16</sup> as a convenient measure of listedness, we are assuming some kind of listedness in the mental lexicon (*cf.* Muysken) [22] (p. 71) and future research may be able to capture this more effectively. Our results suggest that we should revise Figure 3 so that there is some discontinuity on the continuum between switches and borrowings. This is depicted in Figure 6.



**Figure 6.** Revised hypothetical relationship between linguistic integration of donor-language items and frequency of use (showing discontinuity).

We propose therefore that a low degree of integration is both a necessary and sufficient condition for identifying code-switches. Low frequency, on the other hand, is a necessary but not a sufficient condition for identifying code-switches. This is because some listed and native items may also have low frequency, but it seems that code-switches **MUST** have low frequency. These generalisations are summarised in Table 5. As Table 5 shows, code-switches will be those donor-language items which are low in both frequency and integration and they will fall on the left of the vertical line in Figure 6. The rest will be borrowings, and will fall to the right of the vertical line in Figure 6. Borrowings will be ‘listed’ and will be characterised by variable frequency and variable mutation within the same range as native Welsh items.

**Table 5.** Role of integration and frequency in identifying code-switches and borrowings.

Donor-Language Items	Frequency	Integration
Code-switches	Low	Low
Borrowings	Variable	Native-like

<sup>16</sup> We recognize that for the study of Welsh-English we are fortunate in having available comprehensive dictionaries for both languages, so that we can relatively easily distinguish switches from borrowings. Of course, we also recognize that there may be a time lag between the borrowing of a new word and its appearance in the dictionary, but our results to be reported below suggest that the criterion nevertheless works fairly well.



## 10. Conclusions

In this paper we have argued that although competing theories of the same phenomenon may appear to be incommensurable, they can nevertheless be evaluated if one is willing to redefine the unit of analysis and examine the assumptions underlying theory-specific definitions. In examining contrasting approaches to the distinction between code-switching and borrowing we have focused on donor-language items as the unit of analysis and have used English-origin verbs in Welsh to challenge Poplack and Meechan's [1] assumption that the distinction between code-switching and borrowing is categorical. Furthermore we have found our data compatible with the hypothesis (Myers-Scotton [2]) that linguistic integration is related to frequency. Overall, we have proposed that code-switches can be distinguished from borrowings on the grounds of low levels of both frequency and integration. Borrowings may or may not occur with low frequency but will have levels of integration comparable to those of host-language items.

**Acknowledgments:** This research was funded by award no. 112230 from the AHRC (Arts and Humanities Research Council in the UK to the first author.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Poplack, S.; Meechan, M. Introduction: How Languages Fit Together in Codemixing. *Int. J. Biling.* **1998**, *2*, 127–138. [CrossRef]
2. Myers-Scotton, C. *Duelling Languages: Grammatical Structure in Codeswitching*; Clarendon Press: Oxford, UK, 1993.
3. Myers-Scotton, C. *Contact Linguistics*; Oxford University Press: Oxford, UK, 2002.
4. Kuhn, T.S. *The Structure of Scientific Revolutions*; Chicago University Press: Chicago, IL, USA, 1970.
5. Feyerabend, P.K. *Against Method: Outline of an Anarchistic Theory of Knowledge*; New Left Books: London, UK, 1975.
6. Popper, K. *The Logic of Scientific Discovery*; Hutchinson: London, UK, 1959.
7. Pavlenko, A., Ed.; *The Bilingual Mental Lexicon: Interdisciplinary Approaches*; Multilingual Matters: Bristol, UK, 2009.
8. Corpus: MIAMI - BangorTalk. Available online: <http://www.bangortalk.org.uk/speakers.php?c=miami> (accessed on 6 May 2016).
9. Deuchar, M.; Davies, P.; Herring, J.; Parafita Couto, M.; Carter, D. Building bilingual corpora. In *Advances in the Study of Bilingualism*; Thomas, E.M., Mennen, I., Eds.; Multilingual Matters: Bristol, UK, 2014; pp. 93–111.
10. Corpus: SIARAD - BangorTalk. Available online: <http://www.bangortalk.org.uk/speakers.php?c=siarad> (accessed on 6 May 2016).
11. Stammers, J.R. *The Integration of English-Origin Verbs into Welsh: A Contribution to the Debate over Distinguishing between Code-Switching and Lexical Borrowing*; VDM Verlag: Saarbrücken, Germany, 2010.
12. King, G. *Modern Welsh: A Comprehensive Grammar*; Routledge: London, UK; New York, NY, USA, 1993.
13. Balam, O. Mixed verbs in contact Spanish: Patterns of use among emergent and dynamic bi/multilinguals. *Languages* **2016**. [CrossRef]
14. Labov, W. *Sociolinguistic Patterns*; University of Pennsylvania Press: Philadelphia, PA, USA, 1972.
15. Sankoff, G. A quantitative paradigm for the study of communicative competence. In *Explorations in the Ethnography of Speaking*; Bauman, R., Sherzer, J., Eds.; Cambridge University Press: Cambridge, UK, 1989; pp. 18–49.
16. Budzhak-Jones, S. Against word-internal codeswitching: Evidence from Ukrainian-English bilingualism. *Int. J. Biling.* **1998**, *2*, 161–182.
17. Stammers, J.R.; Deuchar, M. Testing the nonce borrowing hypothesis: Counter-evidence from English-origin verbs in Welsh. *Bilingualism* **2012**, *15*, 630–643. [CrossRef]

18. Comrie, B. Morphophonological alterations: Typology and diachrony. In Proceedings of the Morphology 2000: Selected Papers from the 9th Morphology Meeting, Vienna, Austria, 24–28 February 2000; Bendjaballah, S., Dressler, W.U., Pfeiffer, O.E., Voeikova, M.D., Eds.; John Benjamins Publishing Company: Amsterdam, The Netherlands; Philadelphia, PA, USA, 2002; pp. 73–90.
19. Ball, M.; Müller, N. *Mutation in Welsh*; Routledge: London, UK, 1992.
20. Thomas, R.J. Geiriadur Prifysgol Cymru, 1st ed. Available online: [http://www.cymru.ac.uk/geiriadur/gpc\\_pdfs.htm](http://www.cymru.ac.uk/geiriadur/gpc_pdfs.htm) (accessed on 15 April 2016).
21. Poplack, S.; Dion, N. Myths and facts about loanword development. *Lang. Var. Chang.* **2012**, *24*, 279–315. [[CrossRef](#)]
22. Muysken, P. *Bilingual Speech: A Typology of Code-Mixing*; Cambridge University Press: Cambridge, UK, 2000.
23. Bybee, J.L. *Frequency of Use and the Organization of Language*; Oxford University Press: Oxford, UK, 2007.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).