Bilingualism: Language and Cognition 15 (3), 2012, 630–643 © Cambridge University Press 2011 doi:10.1017/S1366728911000381

Testing the nonce borrowing hypothesis: Counter-evidence from English-origin verbs in Welsh*

JONATHAN R. STAMMERS
MARGARET DEUCHAR
ESRC Centre for Research on Bilingualism in Theory and Practice, Bangor University

(Received: August 25, 2010; final revision received: May 27, 2011; accepted: July 1, 2011; first published online 15 December 2011)

According to the nonce borrowing hypothesis (NBH), "[n]once borrowings pattern exactly like their native counterparts in the (unmixed) recipient language" (Poplack & Meechan, 1998a, p. 137). Nonce borrowings (Sankoff, Poplack & Vanniarajan, 1990, p. 74) are "lone other-language items" which differ from established borrowings in terms of frequency of use and recognition. Lone other-language items are singly occurring words from the "donor" language which are preceded and followed by words or phrases from the "recipient" language. Whether such other-language words belong only to the donor language (and are classed as codeswitches) or to both the donor and the recipient language (and are classed as borrowings) is both a theoretical and a practical issue. Poplack & Meechan (1998a) suggest that this question can be settled by measuring the linguistic integration of donor-language words, so that infrequent donor-language words which behave like their recipient-language counterparts are categorised as (nonce) borrowings. This suggests that frequency of use need play no role in the extent to which other-language items are linguistically integrated into the recipient language. We challenge this hypothesis with an analysis of soft mutation on English-origin verbs in Welsh, which shows that integration is related to frequency.

Keywords: codeswitching, borrowing, Welsh–English, loanword integration, word frequency, mutation

1. Introduction

Researchers who study codeswitching have debated for more than half a century about how to identify linguistic borrowings as opposed to switches. Single "other-language" items inserted in an utterance may range from being a fully integrated borrowing like *adroit* (meaning "dexterous", borrowed from French in the seventeenth century) in English to a word like *bonjour*, which most English speakers would consider to be a French word or switch into French. However, there are many "in-between" examples which are hard to categorise. This is not only a theoretical problem but also a practical one for theorists of language contact because if borrowings have a different language membership from switches then one might expect them to behave differently. It is nevertheless quite tricky to find foolproof criteria for distinguishing between the two. Muysken (2000) proposes the notion of LISTEDNESS, according to which "[t]he dimension of listedness refers to the degree to which a particular element or structure is part of a memorized list which has gained acceptance within a particular speech community" (Muysken, 2000, p. 71). Part of the problem stems from the fact these two labels – switch and borrowing – are essentially theoretical constructs, operationally defined in different ways depending on the theoretical framework used by the researchers. The approach of Shana Poplack and associates treats borrowing and codeswitching as fundamentally different processes; as "two distinct phenomena" (Poplack & Meechan, 1998a, p. 132), whereas Carol Myers-Scotton's approach sees the two processes as being "part of the same developmental continuum, not unrelated phenomena" (Myers-Scotton, 1993, p. 163). Myers-Scotton argues that "they undergo largely the same morphosyntactic procedures . . . during language production" (ibid.).

Many criteria of various kinds have been proposed for making a borrowing versus codeswitching distinction.

* This research was funded by award no. 112230 from the AHRC (Arts and Humanities Research Council) in the UK to the second author. The work presented in this paper was part of the programme of the ESRC Centre for Research on Bilingualism in Theory and Practice at Bangor University. We would like to thank Gwen Awbery, Ad Backus, Phylip Brake, Peredur Davies, Marcel den Dikken, Eva Eppler, Carl James, Bob Morris Jones, Siân Lloyd-Williams, Kara McAlister, Carol Myers-Scotton, Shana Poplack, Elen Robert and Alberto Rosignoli for commenting on an earlier version of the paper, and two anonymous reviewers. In addition, we would like to thank Diana Carter, Edward Carter and Hans Stadthagen-Gonzalez for help with statistics.

Address for correspondence:
Margaret Deuchar, ESRC Centre for Research on Bilingualism in Theory and Practice, Bangor University, College Road, Bangor, Gwynedd LL57 2DG, Wales, UK
m.deuchar@bangor.ac.uk

Poplack & Sankoff (1984) (summarised by Muysken, 2000, p. 73) outline both linguistic and usage-based criteria. Their linguistic criteria are morphophonemic and/or syntactic integration while their usage-based criteria include frequency of use, native-language synonym displacement and acceptability. However, they caution that "[n]ot all of these criteria … will be satisfied in all cases which we may want to consider loanwords, and each of them may be satisfied by words which are not" (Poplack & Sankoff, 1984, p. 104).

Poplack and Sankoff (1984) not only use both linguistic integration and the usage-based frequency as criteria for identifying borrowings, as outlined above, but they also recognise a possible relation between linguistic integration and frequency. They observe that "[e]ven the degree to which the loanword is linguistically integrated has been attributed to the frequency of its use within the community" (Poplack & Sankoff, 1984, pp. 101–112). However, as we shall show below, this relation between frequency and linguistic integration is implicitly denied by the nonce borrowing hypothesis, introduced by Sankoff, Poplack and Vanniarajan (1990), according to which the degree of linguistic integration is claimed to be the same independently of frequency. Our contribution in this paper will be to provide evidence from a new language pair for the relation between frequency and linguistic integration, using a particularly sensitive indicator of linguistic integration, mutation in Welsh. We find that the application of mutation to the singly-occurring English-origin items we analyse is highly dependent upon the frequency of those items. The implication of this finding will be that the category of nonce borrowings is redundant.

In our study we will consider naturalistic Welsh–English spoken data, specifically single English verbal insertions into Welsh. We will treat these insertions as lone other-language (or "donor language") items, avoiding classifying them in advance as either switches or borrowings. Using variationist methodology as developed by Poplack and colleagues, we will consider the extent to which the frequency of occurrence of English items in Welsh affects their integration into Welsh. Frequency will be measured in terms of the number of occurrences in our spoken corpus of 456,266 words, while integration will be measured in terms of the application (or not) of the process of mutation where expected. We will compare our results for two categories of English items: those which are listed in the Welsh dictionary (which one might call "established borrowings") and those which are not. Non-listed English items are given in bold italics in examples (1) and (2) below, while listed items like *spwnjo* "to sponge" in example (1) and *tsieca* "check" in example (3) are not bolded and are spelt as in Welsh.

Examples (1)–(3) contain instances of English verbs inserted into otherwise Welsh sentences:[1,2]

(1) *maen nhw 'n* ***exfoliate****-io chdi*
be.3PL.PRES PRON.3PL PRT exfoliate.NONFIN PRON.2S
gynta ac yn spwnjo chi drosodd
first and PRT sponge.NONFIN PRON.2PL over
gynta [Fusser30]
first
"They exfoliate you first, and sponge you over first."

(2) *os wnei di* ***power-walk****-io*
if do.2S.NONPAST PRON.2S power-walk.NONFIN
fydda chdi fath â [Robert3]
be.2S.FUT PRON.2S kind with
"If you power-walk, you'll, like …"

(3) *tsieca fe ar y rhyngrwyd* [Fusser27]
check.IMPER PRON.3SM on DET internet
"Check it on the internet."

Exfoliate-io was not found in any dictionary of Welsh consulted, whereas *sponge-o*, spelt *spwnjo*, was found in the *Geiriadur Prifysgol Cymru* dictionary (Thomas, 1950–2002). Even where a sufficiently comprehensive dictionary does exist, it will tend to represent the more formal, standardised variety. Also, the contact linguist is effectively shifting the problem onto the lexicographer, who will consider somewhat similar factors in deciding which non-native origin items to include, but whose agenda may be very different. They may be highly prescriptive or conservative, or give greater weight to certain respected "literary" sources.

Nevertheless, whether or not a word is included in the Welsh dictionary has proved a useful variable in our study. Where in the analysis below we compare the mutation rate of native Welsh verbs with those of English-origin verbs listed and not listed in a Welsh dictionary, we find clear differences. We will see that the rate of mutation on both native Welsh verbs and listed English verbs increases gradually according to the frequency of those verbs, but is also significantly higher than that of unlisted English verbs

[1] Key to glosses: 1S/2S/3S = 1st/2nd/3rd person singular (NB: 2S = 2nd person singular/familiar); 2PL = 2nd person plural/non-familiar; 3PL = 3rd person plural; 3SF = 3rd person singular feminine; 3SM = 3rd person singular masculine; CONDIT = conditional tense; DET = determiner (definite article); FUT = future tense (present, imperfect, conditional and future apply only to forms of the verb *to be*); IM = interaction marker (filler); IMP = imperfect tense; IMPER = imperative; NEG = negative/negative particle; NONFIN = non-finite; NONPAST = non-past tense; PAST = past tense; POSS = possessive pronoun; PRES = present; PRON = pronoun; PRT = particle; V = vowel.

[2] Except where indicated otherwise, references in numbered examples are to CHAT transcription files in the *Siarad* corpus (see Section 3 below).

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

of the same frequency. We will argue that this constitutes strong evidence against the nonce borrowing hypothesis.

2. The nonce borrowing hypothesis

One controversial category repeatedly used by Poplack and associates is that of NONCE BORROWINGS, a term originally used by Weinreich (1953, p. 11), though in a quite different context that preceded any notion of codeswitching as a phenomenon separate from borrowing. According to Sankoff et al. (1990, p. 71), "[n]once borrowings in the speech of bilinguals differ from established loanwords in that they are not necessarily recurrent, widespread or recognized by host language monolinguals". However, "they share the characteristics of morphological and syntactic integration in the host language and consist of single content words or compounds" (ibid.).

Sankoff et al. formulate what has come to be known as the nonce borrowing hypothesis (NBH hereafter), according to which there is "no difference between nonce borrowings and established loans ... with respect to their morphological and syntactic integration into host language contexts" (Sankoff et al., 1990, p. 94). From the definition of nonce borrowings versus established loans cited above, it would seem that the main difference between the two is frequency of use. We shall therefore refer to frequency explicitly in our own reformulation of the hypothesis:

> There is no difference between frequent and infrequent donor-language items in terms of their degree of integration.

We shall use this revised formulation in our own study of the insertion of English verbs in Welsh discourse. We shall measure integration by the application of mutation, which is a sensitive index, and will pay careful attention to the role of frequency in this integration.

The proponents of the NBH pioneered quantitative analysis in order to test their hypothesis. Sankoff et al. (1990) provide a quantitative analysis of Tamil–English codeswitching in which they compare the native Tamil and English-origin complements of Tamil verbs, concluding that "the English-origin material is shown, both morphologically and syntactically, to be virtually indistinguishable from Tamil (nonpronominal) nouns" (Sankoff et al., 1990, p. 71). Although English-origin objects are case-marked for accusative and dative slightly less than Tamil objects, they consider there to be enough similarity in the two patterns to provide strong evidence for the NBH (Sankoff et al., 1990, p. 87). Nevertheless, they report in a footnote (p. 99) that "the marking on English-origin material does seem to occur at a slightly lower rate in all the contexts we have examined quantitatively". This difference is not further discussed. Sankoff et al.'s method of quantitative analysis is developed further in a special journal issue edited by Poplack and Meechan (1998b) in which a more rigorous approach is followed. That is, some authors of the papers included there not only compare lone donor-language items inserted in a recipient language with equivalent items in the recipient language, but also with the same items in unmixed discourse. Sankoff (2001, p. 649) describes this approach as a "significant breakthrough in resolving the question of single-word tokens as 'code-switches' vs. 'borrowings' ... via the application of quantitative sociolinguistic methodology". However, we venture to suggest that there are two limitations to be found in the studies collected by Poplack and Meechan (1998b). One is that small differences between donor- and recipient-language items are sometimes overlooked, and the second is that the role of frequency is ignored.

Most of the papers deal exclusively with nouns. This focus is doubtless partly due to their sheer abundance, but also due to the position of nouns as the most highly borrowable elements in various proposed "hierarchies of borrowability" (see Muysken, 2000, pp. 73–74; Poplack, Sankoff & Miller, 1988). However, such a preoccupation with one grammatical category to the virtual exclusion of others would seem to be questionable. Another characteristic which all the studies in the volume share is the fact that they almost all conclude that the single-word donor-language insertions under consideration should all be classified as borrowings rather than switches. Only Turpin (1998) is a partial exception to this trend. With respect to discourse flagging, adjective placement in the NP, and determiners, Turpin did find very similar patterning between lone English-origin nouns and native French nouns, in a similar vein to those other studies. However, with respect to plural marking, Turpin unexpectedly found a quite different pattern, with English affixes appearing on several of the lone English-origin nouns, Therefore she had to conclude that a small proportion of the English-origin nouns were codeswitches, a conclusion that could easily have been missed had analysis been confined to the other linguistic variables. A further finding of interest to our study was that the tendency to use the English affix was related to relative infrequency in the use of the English word. Turpin cites Poplack et al.'s (1988) study as producing a similar finding and also reports Poplack and Sankoff's (1984) observation that "the more a loanword is diffused (and accepted) in the linguistic community, the more it is integrated into the recipient language" (Turpin, 1998, p. 231). These findings would appear to be contrary (as ours will) to the predictions of the NBH, according to which frequency does not play a role in linguistic integration.

Some of the other papers in the special issue seem at first sight to be more in line with the NBH. For example, Adalar and Tagliamonte (1998, p. 156) find that "when a lone noun, of either English or Turkish

origin, appears in contexts in which it is surrounded by the other language, it patterns systematically in accordance with its counterparts in that other language". Budzhak-Jones (1998, p. 161) finds that English-origin nouns in Ukrainian discourse "replicate the patterns of behaviour of monolingual nouns with respect to gender assignment, modifier-noun agreement, inflectional variability and flagging". Lastly, Samar and Meechan (1998, p. 203), studying lone English-origin nouns in otherwise Persian discourse, find "remarkable similarities between the treatment of native Persian nouns, attested loanwords, and unattested lone English-origin nouns". However, it must be noted that the similarities between attested and unattested English-origin nouns are greater than between either category and unmixed Persian nouns. Since attested loanwords (an undefined category) are assumed to be definite borrowings, the assumption is made that unattested words are borrowings too and that the NBH is supported. The more radical interpretation, based on first principles, that all English-origin nouns in Persian discourse are codeswitches, is not entertained.

Eze's (1998) paper is the only one in the special issue that deals with verbs as well as nouns. His focus is on lone English-origin nouns and verbs in Igbo discourse, comparing them not only with equivalent Igbo items in an unmixed Igbo context, but also with English items in an unmixed context and with English items in a multiword fragment of English within Igbo discourse. Analysing lone nouns, he considers the occurrence of definite and indefinite determiners, generic and non-generic nominal reference, and the structure of the NP. In each case lone English-origin nouns in Igbo discourse are found to pattern more like Igbo nouns in unmixed Igbo than English nouns in unmixed English. Turning to lone English-origin verbs, he firstly finds very similar patterning in the distribution of affixes on the verb (participial prefix, indicative affirmative suffix, etc.) between English-origin verbs and "monolingual Igbo verbs", and claims that this offers "strong grounds to conclude that the English-origin verbs are functioning as Igbo, irrespective of their etymological origin" (p. 188). This is illustrated in example (4) below, in which Igbo affixes (in italics) are attached to the English-origin verb *work* (in bold italics).

(4) *Ọ̀* ***work****-ụ-ghị*
it work-V-NEG
"It did not work." (Igbo–English; Eze, 1998, p. 187)

Eze points out that bare other-language forms have frequently been treated as codeswitches in the literature. However, the English-origin bare forms in Eze's data all occur in serial verbal constructions, the type of construction in which Igbo verbs also appear bare. This leads Eze to argue for a "borrowing analysis for bare English-origin verbs in otherwise Igbo discourse" (p. 190). However, the data suggest a less clear-cut conclusion. We may note that 29% (31 out of 106) of the Igbo bare verbs are in non-serial constructions: conditionals marked by tone rather than morphology. The author offers no explanation for why there should be no corresponding proportion of the bare English-origin verbs appearing as conditionals in non-serial constructions. Other peculiarities of English-origin verbs are acknowledged though. Where they appear bare in serial constructions, the majority of them are headed by the verb *me* "to do", while corresponding bare Igbo verbs are headed by a variety of verbs. This is explained as a "loanword incorporation device" (p. 190) and Eze concludes that "these bare [English-origin] verbs have been borrowed into Igbo as well" (p. 191). However, such exceptions make it difficult to see when an item would NOT be counted a borrowing, if both fully integrated items, and apparently less integrated items such as this both count. In fact, the Igbo construction with *me* is reminiscent of verb incorporation devices in other language pairs, such as Japanese–English, where an English verb may be incorporated into an otherwise Japanese utterance by juxtaposing it with the Japanese verb *suru* "to do". Scholars who discuss this phenomenon (e.g. Nishimura, 1995, p. 134 ff.) nevertheless appear to assume that this involves switching rather than borrowing. Furthermore, even Romaine (1989), who is cited by Eze in favour of his notion of a "loanword incorporation device", assumes that a similar construction in Panjabi–English discourse involves codeswitching.

Table 1 summarises the conclusions of the individual papers in the special issue and of a few additional papers reaching similar conclusions. The authors of all these papers share the assumption that English-origin words in other-language discourse will be classed as borrowings if they share characteristics with comparable words belonging to the recipient language but not with English words in unmixed English discourse. For example, if English-origin nouns appear with recipient-language affixes just like recipient-language nouns when in mixed discourse but not when in unmixed English discourse, this will be taken as evidence that the English-origin nouns are borrowings in mixed discourse. However, this ignores the possibility that codeswitched words may also be subject to linguistic integration by the host language, as the Matrix Language Framework (MLF) theory outlined by Myers-Scotton (2002) arguably predicts. The MLF can be interpreted as predicting the integration of switched (as well as borrowed) words in terms of word order and certain aspects of bound morphology. So in relation to example (2), rather than using the Igbo morphosyntactic frame as evidence that *work* is borrowed into Igbo, Myers-Scotton would use the same morphosyntactic frame to identify Igbo as the matrix language of the clause, and would expect donor-language words, whether switches or

Table 1. *Studies of lone donor-language items based on a variationist approach.*

Study	Language pair studied	Elements primarily analysed	Linguistic features studied in analyses	Conclusion
Sankoff et al. 1990	Tamil–English	Lone English nouns	Case inflections	All are borrowings
Poplack & Meechan 1995	Wolof–French; Fongbe–French	Lone French nouns	Definite/indefinite reference; NP word order	All are borrowings
Adalar & Tagliamonte 1998	Turkish–English	Lone English nouns	Vowel harmony; plural affixation; NP word order	All are borrowings
Budzhak-Jones 1998	Ukrainian–English	Lone English nouns	Case inflections	All are borrowings
Eze 1998	Igbo–English	Lone English verbs; lone English nouns	Affix distribution; serial constructions; vowel harmony (verbs); determiners; type of nominal reference; NP word order (nouns)	All are borrowings
Samar & Meechan 1998	Persian–English	Lone English nouns	Definite/indefinite reference; VP word order; case inflections	All are borrowings
Turpin 1998	(Acadian) French–English	Lone English nouns	Determiners; NP word order; plural marking; discourse flagging	Most are borrowings; minority are switches
Arroyo & Tricker 2000	Catalan–Spanish	Lone Spanish nouns	Definite/indefinite reference; plural marking; gender	All are borrowings
Shin 2002	Korean–English	Lone English nouns	Case inflections	All are borrowings
Cacoullos & Aaron 2003	Spanish–English	Lone English nouns	Determiners	All are borrowings

borrowings, to be morphosyntactically integrated, at least insofar as predicted by the System Morpheme Principle.[3] Therefore in order to achieve a "theory-independent classification of lone other-language items" (Poplack & Meechan, 1998a, p. 128) it may make sense to look for measures of linguistic integration which go beyond the predictions of the MLF. Myers-Scotton (1993, p. 183) argues that there are more central and more peripheral aspects of morphological integration and that "the more central ML [matrix language] inflections will appear with CS [codeswitched] forms while B [borrowed] forms may show relatively more peripheral ones as well".[4] In the analysis described below we will test a measure of integration which might be considered to be relatively peripheral in order to distinguish between switches and borrowings.

[3] According to the System Morpheme Principle of the MLF, "all system morphemes which have grammatical relations external to their head constituent ... will come from the Matrix Language" (Myers-Scotton, 2002, p. 59). Furthermore, according to the Uniform Structure Principle, there is a preference for all system morphemes to come from the Matrix Language.

[4] Central aspects of morphological integration are explained by Myers-Scotton (1993) to include subject–verb agreement and tense/aspect marking.

We shall also pay careful attention to the role of frequency. The papers in Poplack and Meechan (1998b) appear to ignore the possible role of word frequency in linguistic integration of lone other-language words in the recipient language. This is surprising, since there are indications of an opposite point of view in earlier work by Poplack and Sankoff (1984), Poplack et al. (1988), and also in more recent work, by Jones (2005). Poplack and Sankoff (1984, p. 101) argue that "the degree to which the loanword is linguistically integrated has been attributed to the frequency of its use within the community". They find a relation between usage frequency and measures of phonological integration which apparently include some morphological integration also. Poplack et al. (1988), in discussing their data, actually acknowledge that grammatical integration increases with frequency of usage. For example, they find that consistency of gender marking on English words in French discourse increases with the frequency of their usage, and that this is also true of the use of French zero plural affixation to English words in French discourse. Given the findings by Poplack and colleagues in the 1980s regarding the role of frequency, it is difficult to understand why this factor has been ignored in the studies in Poplack & Meechan (1998b).

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

In not one of the studies listed in Table 1, which argue for the NBH, is the issue of frequency addressed at all. Frequency plays a role in the study by Jones (2005) on Jersey Norman French in that she removes borrowings from her data in order to investigate codeswitching patterns. Jones uses an arbitrary frequency criterion as once proposed by Myers-Scotton (1993, p. 204) to eliminate borrowings from among the singly-occurring English-origin items in her data. Occurrences of three or more items were counted as borrowings, while occurrences of two or less were considered codeswitching. In order to validate this distinction Jones investigated variation in the discourse flagging of the items, and found (p. 13) that 34% of the "codeswitched forms" were flagged in some way (by means of hesitation, self-correction or "metalinguistic commentary"). Of the "borrowings" Jones found that just 19% were flagged. Then looking just at flagging by "self-correction", 26% of the "codeswitched forms" were flagged, as opposed to 0% of the "borrowings". This satisfied Jones that her category of "codeswitched forms" was sufficiently distinct from the other "borrowed" items to allow her to investigate codeswitching patterns using those data. The relevance of Jones's results to our study is that she has found frequency to make a difference in the occurrence of flagging. If flagging is an indicator of linguistic (non)integration then we may surmise that frequency does indeed make a difference in linguistic integration. This goes against the NBH, according to which infrequently produced donor-language items should be integrated just as well as those which are more frequent.

3. Data

The data are taken from *Siarad*, a spoken corpus collected as part the *Code-switching and Convergence in Welsh* project.[5] The corpus consists of naturalistic recordings of informal conversations between Welsh–English bilinguals. Speakers were recruited by a variety of means, the most successful being through the extended social networks (i.e. friends of friends, etc.) of the researchers.[6] The aim was to obtain a wide range of speakers in terms of variables such as age, gender, geographical location, level of education, and language dominance.[7] They were all Welsh–English adults or teenagers who were fluent in both languages. A typical recording was of a conversation between two speakers who know each other well. Speakers were left to talk in private about whatever they wished. Typically, a recording lasted 30 minutes, with radio microphones and a Marantz digital recorder being used in most cases. A total of 40 hours of recordings have been collected and transcribed in CHAT format (see MacWhinney, 2000), using the LIDES system (see the LIPPS Group, 2000). The initial analysis in Section 4 below was based on extensive study of three fully transcribed conversations, lasting 1 hour and 45 minutes in total: two conversations between pairs of women in their mid-20s ("Fusser29" and "Davies2"), and one between a married couple in their early 40s ("Stammers4"). The analysis in Section 5 was carried out subsequently, and was based on exhaustive searches of the entire 40 hour (456,266 word token) corpus.

For more on Welsh–English bilingualism, see Deuchar (2005, 2006). See also Jones (1993) for further sociolinguistic background.

4. English-origin verbs in Welsh: Integration by derivational suffix

English verbs are typically incorporated into Welsh with the addition of a derivational verbalising suffix, as shown in example (5) below. There the English verb *emphasise* (which in this case already includes an English derivational suffix, *-ise*) has the Welsh derivational suffix -(*i*)*o* added to it to form a non-finite Welsh verb (or "verb-noun": see e.g. Borsley, 1993; King, 2003).[8] In (5), this form is inserted into a periphrastic construction using an inflected form of *bod* "to be" (in this example, 2nd person plural/non-familiar, present) as an auxiliary verb. Note from the first clause in (5) that native Welsh verbs appear in the same construction. The verb *defnyddio* "to use" here also has the -(*i*)*o* derivational suffix. It is a suffix commonly used in Welsh to derive a verb from a Welsh noun or adjective (see Deuchar, 2005, pp. 263–266; King, 2003, p. 132), and the only suffix now productive in incorporating English verbs.

(5) *pan* *dach* *chi* *'n* *defnyddio*
when be.2PL.PRES PRON.2PL PRT use.NONFIN
wide-angle lenses *dach* *chi* *'n*
wide-angle lenses be.2PL.PRES PRON.2PL PRT
emphasise*-io* *'r* ***foreground***.
emphasize.NONFIN DET foreground
"When you use wide angle lenses, you emphasise the foreground." [Fusser17]

Muysken (2000, p. 184) suggests that there are four main ways in which verbs from one language are integrated into another:

[5] The *Siarad* corpus can be accessed at http://www.siarad.org.uk/ or http://talkbank.org/data/BilingBank/Bangor/. The project was funded by award No. 112230 from the Arts and Humanities Research Council in the UK to the second author.

[6] These included Peredur Davies, Marika Fusser, Elen Robert and the first author.

[7] The majority of the speakers (131) come from North Wales, but 16 are from South Wales and three from mid-Wales.

[8] Borsley (1993) argues against the category of "verb-noun", and the term will not be used here.

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

a. The new verb is **inserted** into a position corresponding to a native verb, in adapted form or not
b. The new verb is **adjoined** to a helping verb
c. The new verb is a **nominalized complement** to a causative helping verb in a compound
d. The new verb is an **infinitive** and the complement of a native auxiliary.

As will be clear from the above description of how English verbs are integrated into Welsh using a Welsh derivational suffix, this process appears to conform with type (a) above, specifically involving an adapted form. Muysken provides examples of bare verb insertion, as in when Dutch infinitive forms are inserted into Chinese discourse without any adaptation. On the other hand, he shows how Spanish verbs are inserted into Quechua discourse with the addition of Quechua derivational affixes. Muysken points out that Quechua is agglutinative, which could explain the addition of Quechua derivational affixes to Spanish verbs, whereas Chinese, as an isolating language, has no derivational affixes to add to Dutch verbs. Welsh certainly has more derivational morphology than Chinese, and this could explain why English verbs are adapted by the use of Welsh derivational morphology, which we may consider to be a central aspect of its derivational morphology. As for why the derivational morphology of the host rather than the donor language should be used, this is predicted by the MLF approach (Myers-Scotton, 2002), outlined above, according to which integration into the bound morphology of the host language is expected. It should be noted that the integration of English verbs into Welsh appears to conform to type (d) above as well as type (a). The new, integrated form consisting of an English verb and a Welsh derivational affix is indeed comparable to a Welsh infinitive, and such forms occur as the complement of a Welsh auxiliary. This is because, as outlined above, English verbs are typically inserted into periphrastic constructions which are made up of an auxiliary and an infinitive. These constructions are more common in informal speech than the synthetic alternative, into which English verbs are not normally inserted.

We used three transcribed conversations from the Bangor Siarad corpus to assess the extent of morphosyntactic integration of English-origin verbs (the Davies2, Fusser29 and Stammers4 transcriptions). Each occurrence of every verb in the transcriptions that appeared to be of English origin was classified as either morphologically integrated by means of a derivational suffix or not.[9] As Table 2 shows, from the three transcripts examined, a total of 184 tokens (80 types) of lone English-origin verbs were identified in mixed contexts. These ranged from very well established borrowings to instances of verbs of lower frequency in our data which were not found in any Welsh dictionary. Table 3 shows all of the English-origin verbs identified. Of the 184 tokens, 179 (97.3%) were clearly morphologically integrated by means of a Welsh derivational verbal suffix. In the majority of cases this was the -(*i*)*o* suffix already mentioned, though some borrowings in the data have other suffixes (e.g. -*u* in *helpu* "to help"; -(*i*)*ad* in (*g*)*watsiad* "to watch"). The use of such non-productive suffixes is surely a sign that such borrowings are firmly established in the lexicon.

Table 2. *Morphological integration by affixation of English-origin verbs found in three transcripts.*

	Tokens	%	Types	%
Fully morphologically integrated (by Welsh verbal suffix)	179	97.3	77	96.3
Non–integrated or partially integrated	5	2.7	3	3.7
Total	184	100	80	100

Of the five tokens of lone English-origin verbs that were not clearly morphologically integrated (shown in bold capitals in Table 3), three were instances (all produced by the same speaker) of the verb-type *fancy*, apparently occurring in bare form. We see an example in (6), where *fancy* is inserted in a periphrastic Welsh verbal construction, to act as the non-finite complement to a finite auxiliary verb. (Whereas bold italics indicate an English word in our examples, underlined words indicate words which could be both English and Welsh.).

(6) *dw 'm yn* fancy *eistedd mewn*
be.1S.PRES.NEG NEG PRT fancy sit.NONFIN in
gornel efo hi trwy 'r nos timod?
corner with PRON.3SF through DET night you know
"I don't fancy sitting in a corner with her all night, you know?" [Fusser29]

However, it is debatable here whether or not this really is a bare form, or whether it is actually (partially) morphologically integrated. The attested dictionary form of this verb would be *ffansïo* [fan'sijɒ], but there may be other reasons why this form [fansi] is preferred by these speakers. For example, there are comparable verbs ending in -*i*, such as the native Welsh verb *poeni* "to worry", or the verb *profi* "to prove/try out", borrowed from English as early as the 13th century. Alternatively, one could argue that *fancy* is not truly a verb. *Ffansi* is indeed listed in Thomas (1950–2002) as a noun or adjective, but not as a verb. In example (7), it seems reasonable to analyse *fancy* in a similar way to *eisiau*/*isio* "want" or *angen* "need", which do not inflect and which some would analyse as nouns (or "pseudo-verbs") rather than verbs (King, 2003, pp. 238ff.).

[9] It was not always clear whether or not a verb was of English origin. See footnotes relating to some of the items in Table 3.

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

Table 3. *English-origin verbs found in mixed clauses in three corpus transcripts, with items not listed in a Welsh dictionary shown in bold italics, and number of tokens given. English-origin verbs that were not clearly morphologically integrated are in bold capitals.*

***activate**-io* ×1	*dilifro* **(deliver)** ×3	*jibio* **(jib)** ×1	*sgrechian*[b] **(screech)** ×1
adio **(add)** ×2	***download**-io* ×1	*licio* **(like)** ×27	*siafio* **(shave)** ×1
arestio **(arrest)** ×1	*drymio* **(drum)** ×1	*marcio* **(mark)** ×1	*sincio* **(sink)** ×1
astudio **(study)** ×1	*dympio* **(dump)** ×2	*meindio* **(mind)** ×7	*siwtio* **(suit)** ×1
***babysit**-io* ×1	***e-mail**-io* ×1	*mwydro* **(moider**[a]**)** ×1	*smwddio* **(iron)** ×1
bwcio **(book)** ×1	*enjoio* **(enjoy)** ×1	**NAME-DROPPING** ×1	*sortio* **(sort)** ×6
canslo **(cancel)** ×1	*esbonio*[b] **(explain)** ×1	*pacio* **(pack)** ×2	*sterio* **(stare)** ×1
carfio **(carve)** ×1	***excuse**-io* ×2	***panic**-io* ×1	*stopio* **(stop)** ×4
cario **(carry)** ×1	**FANCY** ×3	*pasio* **(pass)** ×1	*stwffio* **(stuff)** ×2
cleimio **(claim)** ×3	*ffafrio* **(favour)** ×1	*pigo* **(pick)** ×2	*swopio* **(swap)** ×1
clocio **(clock)** ×1	*ffeindio* **(find)** ×11	*planio* **(plan)** ×2	**TAKING** ×1
***concentrate**-io* ×1	*ffonio* **(phone)** ×9	*protestio* **(protest)** ×1	***text**-io* ×7
***convert**-io* ×1	*fforddio* **(afford)** ×1	*pwyntio* **(point)** ×1	*traenio* **(train)** ×1
***cope**-io* ×2	*ffotocopïo* **(photocopy)** ×1	*recordio* **(record)** ×2	*trio* **(try)** ×10
crio **(cry)** ×3	***freak**-o* x1 ×1	*rhentio* **(rent)** ×2	*tsiecio* **(check)** ×1
cwestiynu **(question)** ×1	***gazump**-io* ×1	*rhymblo* **(rumble)** ×1	*twrio* **(tour)** ×1
cymharu **(compare)** ×1	*gwatsiad* **(watch)** ×6	*sbwylio* **(spoil)** ×3	*twtsiad* **(touch)** ×2
cysidro **(consider)** ×2	*helpu* **(help)** ×2	*seinio* **(sign)** ×1	*wastio* **(waste)** ×1
dawnsio **(dance)** ×1	*insyltio* **(insult)** ×1	*setlo* **(settle)** ×1	*watsio* **(watch)** ×1
delio **(deal)** ×2	*iwsio* **(use)** ×3	*sgratsio* **(scratch)** ×1	***whine**-io* ×1

[a] Meaning, to blather or talk nonsense (at someone); used in some dialects of English, but etymology uncertain and may possibly be a native Welsh verb.
[b] Etymology unclear.

(7) oh *ti* fancy pint *nos* *Iau*?
IM PRON.2S fancy pint night Thursday
"Oh, do you fancy a pint on Thursday night?" [Fusser29]

Example (7) causes difficulties for this analysis since it includes the particle *yn*, not used with these "pseudo-verbs", but it did seem unusual in that respect to some native speakers who we consulted regarding it.

In (8) the English verb phrase *name-dropping* is part of a non-finite clause following the Welsh imperfective particle *yn*. This is unexpected, since Deuchar (2005, pp. 266–268) shows that English verbal participles such as this are blocked from switching in terms of paradigmatic and syntagmatic congruence. The fact that *name-dropping* in English occurs virtually exclusively in its gerund form may be relevant, but such examples are in any case very rare.

(8) *yli* *fi* *'n* ***name-dropping*** *fel* *'a*
see.2S.NONPAST PRON.1S PRT name-dropping like there
"Look at me name dropping like that!" [Davies2]

One other interesting case was found of an English-origin verb lacking morphological integration in Welsh: *taking* as shown in (9), where the verb *take*, as part of the expression *taking it day by day*, appears in a periphrastic construction (with a 3rd person singular of the Welsh verb "be" as auxiliary. Examples of this type are very rare, but since it appears here as part of a multi-word unit, the verb here is not a lone item, so is excluded from this analysis.

(9) *mae* *hi* *'n* ***taking it day by day***
be.3S.PRES PRON.3SF PRT taking it day by day
"She's taking it day by day." [Fusser29]

Aside from these rare counter-examples, we do see that the overwhelming majority of lone English-origin verbs in Welsh, in this data sample, are morphologically integrated. From the perspective of Poplack and her associates, this fact alone would presumably be sufficient to count them all as borrowings. Myers-Scotton would in any case predict the morphological integration of donor-language items, whether these are counted as switches or loans. The -(*i*)*o* suffix would be classified by Myers-Scotton (2002, p. 75) as an "early system morpheme" on the grounds that it has conceptual content but differs from content morphemes in not receiving or assigning thematic roles. (See Deuchar, 2006, p. 1997, for further information about how this category applies to Welsh.) Early system morphemes from the matrix language (Welsh in this

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

case) are preferred over those from the embedded or non-matrix language according to Myers-Scotton's (2002, p. 8) Uniform Structure Principle. We investigate further the extent to which these English-origin verbs are integrated into Welsh, proposing in the process broader and more sensitive criteria of linguistic integration, and not restricting ourselves to those morphosyntactic phenomena which would be predicted to come from the matrix language anyway. We originally tested variation in verbal inflection (or lack of it), comparing the distribution of English-origin with native Welsh verbs in two alternative verbal constructions, periphrastic and synthetic (providing a measure of what might be called "syntactic integration"), but the results here were inconclusive. This was mainly because participation in synthetic constructions in the corpus data was found to be overwhelmingly restricted to the most highly frequent and entrenched verbs, so this criterion was not sufficiently sensitive to variation. A more sensitive criterion was found in the application of a highly variable morpho-phonological process, mutation, to native Welsh versus English-origin verbs. Our study of this process is discussed in the next section, and in more detail in Stammers (2010).

5. English verbs in Welsh: Integration by soft mutation

The Welsh mutation system is highly complex (see e.g. Borsley, Tallerman & Willis, 2007, pp. 19–26; King, 2003, pp. 13–19). The initial consonants of Welsh words undergo mutations of three types (soft, aspirate, and nasal), in a wide variety of contexts. Mutation here is not merely a phonological phenomenon, but is part of the grammar of Welsh, being determined by lexical and morphosyntactic environments. Soft mutation is the most widespread of the three types, and affects the largest number of consonants. The consonant changes are summarised in Table 4.

The non-finite verb in Welsh may undergo soft mutation in various environments, triggered either by a particular preceding word (normally a preposition) or by a sentence construction. In this analysis we classify non-finite verbs in the *Siarad* corpus where they occur in various environments where soft mutation is expected. The dependent variable in our analysis is thus the application or not of soft mutation in these environments. This can be grouped into two main types as follows.

5.1 Lexically triggered mutation

In lexically triggered soft mutation, the non-finite verb is directly preceded by a preposition, clitic or other particle causing soft mutation. Example (10) illustrates the most frequent environment found in our data, where the pre-verbal particle *i* (roughly equivalent to *to* in English) triggers soft mutation of the following verb *costio* "to cost", so that it becomes *gostio*.

Table 4. *Initial consonants changes in soft mutation in Welsh.*

Initial consonant (phonetic)	p	t	k	b	d	ɬ	r^h	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	*(dropped)*
Mutates to (orthographic)	b	d	g	f	dd	l	r	f	

(10) well *mae* *mynd* *i* *gosti*
well be.3S.PRES go.NONFIN to cost.NONFIN
pres
money
"Well, it's going to cost money." [Fusser6]

Another fairly frequent environment is where the non-finite verb is preceded by the 3rd singular possessive element (or proclitic, Borsley et al., 2007, p. 23) *ei*, where the possessor is masculine; *ei* may occur either as an agreement marker with the verb's object, as in (11), or as part of a passive construction based on *cael* "to have", as in (12).

(11) *fyswn* *i* *licio* *ei*
be.1S.CONDIT PRON.1S like.NONFIN POSS.3SM
fenthyg *o*
borrow.NONFIN PRON.3SM
"I'd like to borrow it . . . " [Fusser9]

(12) *yr* *un* *capel* *'di* *cael* *ei*
DET one chapel PRT.PAST have.NONFIN POSS.3SM
***convert**-io*
convert.NONFIN
"the same chapel, having been converted" [Fusser29]

Note that in (11) the verb *benthyg* "to borrow" appears in its mutated form *fenthyg* [venθig] whereas in (12) the verb ***convert**-io* appears in an unmutated form.

There are various other environments, found less frequently in the data, where the non-finite verb is directly preceded by other lexical triggers. These include the 2nd singular/familiar possessive (or agreement clitic) *dy* "your", or the preposition *am* "for/about", as in

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

(13), *o* "of/from", *ar* "on/about to", or *gan* "by/while/with".

(13) *sut mae o am fihafio*
how be.3S.PRES PRON.3SM for behave.NONFIN
"how he's going to behave." [Fusser15]

In (13) the verb *bihafio* "to behave" appears in its mutated form, *fihafio* [vɪˈhavjɒ].

5.2 *Syntactically triggered mutation*

In syntactically triggered soft mutation, the non-finite verb will be expected to mutate due to its position in the clause, generally speaking because it follows the grammatical subject. Example (14) illustrates the most frequent such environment in the data, where the verb (here, *trio* "to try") appears in its mutated form (here, *drio*) due to its position in a periphrastic construction based on the *gwneud* auxiliary (here *wnest:* 2nd singular/familiar past), with subject following auxiliary and preceding the main verb.

(14) *wnest ti drio?*
do.2S.PAST PRON.2S try.NONFIN
"Did you try?" [Stammers5]

In (15) the non-finite verb (*deud* [deɨd] "to say") again mutates (to *ddeud* [ðeɨd]) following the grammatical subject, this time preceded by a modal auxiliary verb rather than a *gwneud* auxiliary:

(15) *sut fedra i ddeud*
how be.able.1S.NONPAST PRON.1S say.NONFIN
"how can I say" [Fusser4]

Finally, a non-finite verb will be expected to be mutated in a case such as (16), where it once again follows the subject in an infinitival clause introduced by the element *i* "to", in the sequence *i* + DP + VP.

(16) *well i ti bwco diwrnod* off
better to PRON.2S book.NONFIN day off
"You'd better book a day off." [Robert6]

Here note that the verb *bwco* "to book" is uttered in its unmutated form, whereas the mutated form *fwco* [vʊko] would have been expected in this environment.

In this analysis we quantify the application of soft mutation to three groups of verbs: (i) native Welsh verbs; (ii) English-origin verbs listed in a dictionary of Welsh; and (iii) English-origin verbs not listed in a Welsh dictionary (UNLISTED). Analysis involved classifying all 506 of the non-finite verb tokens found in the *Siarad* corpus that: (i) end in the -(*i*)*o* suffix; (ii) begin with a consonant susceptible to soft mutation, excluding the [ɬ]→[l], [rʰ]→[r] and [g]→[Ø] mutations; and (iii) occur in an environment where soft mutation is expected. Classifying only verbs ending in -(*i*)*o* is a good way of restricting analysis of native Welsh verbs to a manageable number of instances, as well as ensuring that irregular verbs are excluded, and that the native items are truly equivalent to the English-origin items. Excluding [rʰ]→[r] and [ɬ]→[l] is justifiable since [rʰ] and [ɬ] are not sounds that exist in English, and the [g]→[Ø] mutation is excluded both due to a lack of data for English-origin verbs, and also since it is an unusual case, strictly involving a deletion rather than a mutation. The results are shown in Figure 1 and in Table 5.

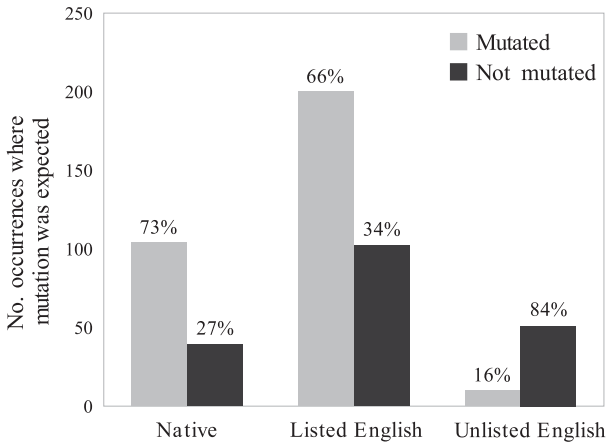

Figure 1. Results of analysis of soft mutation based on verb status.

As Figure 1 shows, mutation applies in the majority of expected environments in the case of the native Welsh (73%) and the listed English (66%) verbs. However, mutation only applies in the minority of expected environments in the case of unlisted English verbs (16%). These results appear to provide strong evidence against the category of nonce borrowings, as there should be (according to the NBH) "no difference between nonce borrowings and established loans … with respect to their morphological and syntactic integration into host language contexts" (Sankoff et al., 1990, p. 94). It is the unlisted English verbs which would be candidates for nonce borrowing status, but to achieve it they would need to show similar rates of the application of mutation to those found in the listed English and native Welsh verbs. This is patently not the case. A question not answered by the results presented in Figure 1 is whether the difference between the behaviour of the unlisted English and the other verbs can be explained by their unlisted status as such or by a factor such as frequency. We therefore conducted a frequency analysis of all the verbs analysed for Figure 1; the results are shown in Figure 2. Note that

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

Table 5. *Summary of results for the analysis of application of soft mutation. Percentages are in italics; summed totals are in bold.*

Frequency band		Native verbs			Listed English			Unlisted English			Overall			
Frequency per million words	Absolute verb frequency	n	n mutated	*% mutated*	n	n mutated	*% mutated*	n	n mutated	*% mutated*	n	n mutated	*% mutated*	*Types*
1–9	1–4	18	12	*66.7*	47	25	*53.2*	28	1	*3.57*	93	38	*40.9*	*(79)*
10–99	5–45	66	46	*69.7*	120	74	*61.7*	33	9	*27.3*	219	129	*58.9*	*(72)*
100–999	46–450	44	33	*75.0*	135	101	*74.8*	0	0	–	179	134	*74.9*	*(7)*
1000–9999	451–4500	15	13	*86.7*	0	0	–	0	0	–	15	13	*86.7*	*(1)*
Overall		143	104	***72.7***	302	200	***66.2***	61	10	***16.4***	**506**	314	***62.1***	***(159)***

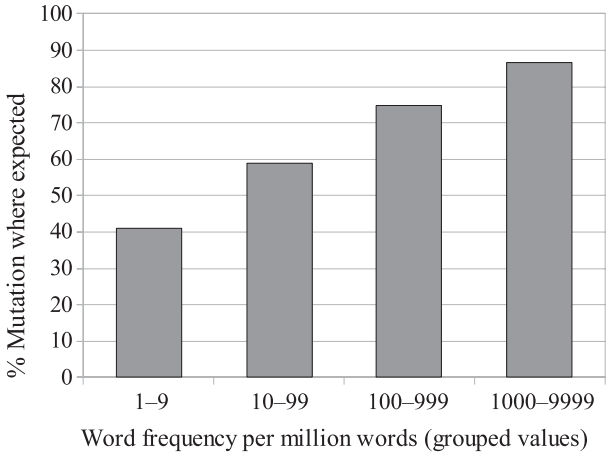

Figure 2. Results for mutation rate by word frequency groupings.

Figure 2 does not differentiate between the categories of verbs (native/listed English/unlisted English) used in Figure 1 and that, like Figure 1, it includes the result for native Welsh verbs.

In Figure 2 we show the application of soft mutation in the expected environments to all verbs categorised this time according to the frequency of the verbs in the corpus. In Figure 2, a remarkably clear and definite relationship is observed between overall frequency and rate of mutation where expected. Items of frequency 1–9 per million in our corpus only mutate in a small minority of instances (40.9%, or 38 instances out of 93), whereas items of frequency 1000 or more per million mutate in the vast majority of instances (86.7%, or 13 instances out of 15), 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. Such a grouping makes good sense in view of the way the data are distributed, with very many tokens of low frequency verbs, and exponentially fewer tokens as the frequency increases. The data conforms, at least roughly, to Zipf's law (Zipf, 1935), which states that in a ranked frequency word list based on a natural text, frequency will be in inverse proportion to frequency rank. In grouping data based on frequency, equal sized linear groupings such as 1–200, 200–400, 400–600, etc. are generally not appropriate for the data. Grouping the data into equal linear groups would result in the vast majority of the data being contained in the lowest group(s), and probably some of the intermediate higher groups containing no data at all. Analysis based on logarithmic rather than raw values for frequency can also be justified theoretically in terms of theories of the effects of learning and practice. Lewis, Gerhand and Ellis (2001) discuss the cumulative-frequency hypothesis, re-analysing data from

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

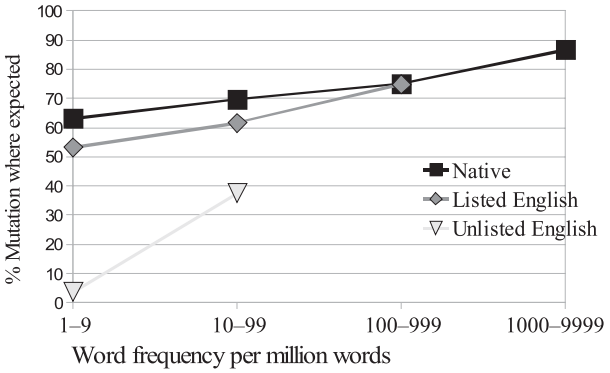

Figure 3. Rate of mutation for verbs in the three groups across frequency bands.

previous studies to show how log-linear values are far stronger predictors than linear values of certain effects. In psycholinguistic studies log values of word frequencies are often used in investigating word frequency effects on, for example, reaction times or performance in lexical or phonological tasks (Alegre & Gordon 1999; Lewis et al., 2001). This same effect was found in our data. A correlation of .78 was found between rate of mutation and frequency based on linear values, and although this is itself not a weak correlation, a markedly stronger correlation of .99 was found based on the logarithmic values.

Whereas Figure 2 shows the relation between the application of mutation and all verbs with no distinction being drawn as to category (native Welsh vs. English listed vs. English unlisted), Figure 3 shows the application of mutation according to both category of verb and frequency. As Figure 3 shows, the mutation rate is similar in the categories of Welsh native and listed English verbs, increasing gradually according to the frequency of those verbs. The behaviour of the unlisted English verbs can only be compared with the other two categories at the lower levels of frequency (i.e. 1–9 and 10–99 per million) but we can see that they behave quite differently, having a lower rate of mutation for the same level of frequency as the other two. If they were nonce borrowings then we would expect the rate of mutation to be the same as in the other categories.

Although Figure 3 suggests that frequency and verb category (unlisted vs. the rest) are the main predictors of the application of mutation, we also considered the possibility that the particular consonant to be mutated and the mutation environment may have made a difference.

Figure 4 shows the effect of the initial consonant to be mutated. Although one individual difference, between initial consonants /b/ and /k/, was found to be significant at the 5% level ($p = .012$), the initial consonant was not found overall to be significant in predicting the application of mutation. Table 6 allows comparison of the rates

Table 6. *Distribution of rates of soft mutation on verbs, according to type of trigger.*

		Lexically triggered	Syntactically triggered	Total
Mutated as expected	n	162	152	314
	%	64.0%	60.1%	62.1%
Not mutated	n	91	101	192
	%	36.0%	39.9%	37.9%
Total	n	253	253	506

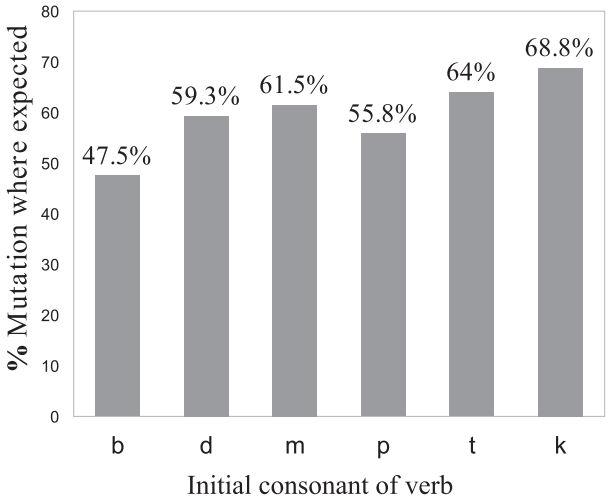

Figure 4. Distribution of mutation rates by initial consonant.

of mutation in lexically versus syntactically triggered environments. The rates of mutation are similar (around 60%) in both environments, with no significant difference being found.

However, though one individual difference, between initial consonants /b/ and /k/, was found to be significant at the 5% level ($p = .012$), these variables were found overall to be not significant in predicting the application of mutation.

6. Discussion

The process of incorporating verbs from English into Welsh speech, adapting them morphologically by the addition of the derivational suffix -(*i*)*o* has been shown to be highly productive. If we had used only this process of derivation as a measure of the integration of English verbs into Welsh, we would doubtless have concluded that all English-origin verbs are borrowings, some established (the listed English verbs) and some nonce (the unlisted English verbs). However, this would have been a

https://doi.org/10.1017/S1366728911000381 Published online by Cambridge University Press

contentious finding for advocates of the MLF, according to which this kind of derivational morphosyntactic integration might be expected to apply to all donor-language items, whether switches or borrowings. A fortunate aspect of working with Welsh was that we could draw upon an unusual (and arguably peripheral) aspect of morphosyntax, mutation, in order to measure integration. Furthermore, because of the variable nature of this process in even native Welsh material, we were able to compare the frequency of its application in Welsh and English-origin material, both listed and unlisted, making use of the variationist methodology pioneered in Sankoff et al. (1990) as well as in earlier studies. It may be because of the variability of this process and its sensitivity to both frequency and listedness that we have been able to find evidence against the NBH that was not apparent in the studies of the language pairs reported in Poplack and Meechan (1998b).

If we look again at our main results as displayed in Figure 3, we can identify two main findings: (i) the application of mutation increases with the frequency of the verb in the corpus, across all categories of verbs; (ii) for the same frequency of verb, there is no significant difference between the native Welsh and listed English verbs, but there IS a significant difference between those two categories and the unlisted verbs, where mutation applies less. This suggests that not only frequency but also "listedness" is a factor.[10] We may note that the slope of the line representing mutation in unlisted English verbs is steeper than in the case of the other categories, and it could be surmised that once unlisted English verbs reached the level of 100 per million or more, they would be showing the same level of mutation as the other verbs. Since no unlisted verbs reach this level, perhaps we can assume that when that happens they change their status to listed and enter the dictionary.

We used the results of our study to test a version of the NBH as formulated in Section 2 above: "There is no difference between frequent and infrequent donor-language items in terms of their degree of integration". The clear relation between frequency and integration into Welsh for both listed and unlisted English verbs unequivocally refutes this hypothesis. Since we do not find infrequent items that pattern like frequent ones, we also do not find evidence for the category of nonce borrowings.

Our results as summarised in Figures 2 and 3 offer clear empirical evidence of a continuum of integration, from verbs that are not integrated to verbs that are fully integrated. This is an important finding, since Myers-Scotton (1993), Treffers-Daller (2005) and others have argued that there is a continuum between codeswitching and borrowing. Although our results give empirical support to the idea of a continuum of integration, it is not clear whether the continuum runs from "switches" to "borrowings". Where the variable of listedness is applied, we do appear to find a categorical distinction. Thus, in Figure 3 a gap is seen between the maximum integration of the unlisted verbs (less than 40%) and the minimum integration of the listed verbs (over 50%). So the unlisted verbs may be candidates for the category of codeswitches, despite the acknowledged limitations of a Welsh dictionary in providing a criterion for identifying borrowings.

However, the next step in determining whether or not there is a categorical distinction between switches and borrowings may be to make use of neuroscientific techniques which could provide an insight into the status of candidate "switches" versus "borrowings" in the mental lexicon. We note Myers-Scotton's (1993, p. 207) claim that "B [borrowed] forms and CS [codeswitched] forms differ in their status in relation to the ML mental lexicon. B forms are entered in this lexicon, but CS forms are not". This claim could be tested by conducting ERP (event-related potential) studies of bilingual speakers being exposed to "other-language" words in order to examine whether their brain responses fall into two distinct categories. In the Welsh–English context, Welsh–English speakers could be exposed to a range of English and Welsh words while ERP measurements are made. English monolingual speakers as controls would be exposed to the same English words to provide a baseline and reduce the potential influence of phonetic differences between words. There is currently considerable controversy about the structure of the bilingual mental lexicon (see e.g. Dong, Gui & MacWhinney, 2005; Pavlenko, 2005) but a neuroscientific study of the kind outlined might contribute to our understanding of how the boundaries between languages are stored in a bilingual's lexicon, and what the status is of borrowed and switched words.

7. Conclusion

We have explored the role of frequency in the linguistic integration of donor-language items with reference to the application of soft mutation to English-origin verbs inserted in Welsh. We have shown that if the frequency of native, listed English and unlisted verbs is taken into consideration, listed English verbs are as well integrated as native Welsh verbs of the same frequency, this level of integration increasing as the frequency of the verb increases. In the case of unlisted English verbs, however, we have seen that they are less well integrated than listed English-origin or native Welsh verbs with the same frequency. This is contrary to the predictions of the NBH, and suggests that the category of nonce borrowings is

[10] Muysken (2000, p. 71) argues that "[t]he dimension of listedness refers to the degree to which a particular element or structure is part of a memorized list which has gained acceptance within a particular speech community".

redundant. The outcome is a more parsimonious theory in which no intermediate category between codeswitches and borrowings is needed, and in which a notion of "listedness" is the main way to distinguish between them.

References

Adalar, N., & Tagliamonte, S. (1998). Borrowed nouns; bilingual people: The case of the "Londrali" in Northern Cyprus. *International Journal of Bilingualism, 2 (2),* 139–159.

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40 (1),* 41–61.

Arroyo, J. L. B., & Tricker, D. (2000). Principles of variationism for disambiguating language contact phenomena: The case of lone Spanish nouns in Catalan discourse. *Language Variation and Change, 12,* 103–140.

Borsley, R. D. (1993). On so-called verb-nouns in Welsh. *Journal of Celtic Linguistics, 2,* 35–64.

Borsley, R. D., Tallerman, M., & Willis, D. (2007). *The syntax of Welsh*. Cambridge: Cambridge University Press.

Budzhak-Jones, S. (1998). Against word-internal codeswitching: Evidence from Ukrainian–English bilingualism. *International Journal of Bilingualism, 2 (2),* 161–182.

Cacoullos, R. T., & Aaron, J. E. (2003). Bare English-origin nouns in Spanish: Rates, constraints, and discourse functions. *Language Variation and Change, 15,* 289–328.

Deuchar, M. (2005). Congruence and Welsh–English code-switching. *Bilingualism: Language and Cognition, 8 (3),* 255–269.

Deuchar, M. (2006). Welsh–English code-switching and the Matrix Language Frame model. *Lingua, 116,* 1986–2011.

Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition, 8,* 221–238.

Eze, E. (1998). Lending credence to a borrowing analysis: Lone English-origin incorporations in Igbo discourse. *International Journal of Bilingualism, 2 (2),* 183–201.

Jones, M. C. (2005). Some structural and social correlates of single word intrasentential code-switching in Jersey Norman French. *French Language Studies, 15,* 1–23.

Jones, R. O. (1993). The sociolinguistics of Welsh. In M. Ball & J. Fife (eds.), *The Celtic languages*, pp. 536–605. London: Routledge.

King, G. (2003). *Modern Welsh* (2nd edn.). London: Routledge.

Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating age-of-acquisition effects: Are they simply cumulative-frequency effects? *Cognition, 78 (2),* 189–205.

The LIPPS Group (2000). The LIDES Coding Manual: A document for preparing and analyzing language interaction data. *International Journal of Bilingualism, 4 (2),* 131–270.

MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. Mahwah, NJ: Lawrence Erlbaum.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

Myers-Scotton, C. (2002). *Contact linguistics*. Oxford: Oxford University Press.

Nishimura, M. (1995). Varietal conditioning in Japanese/English code-switching. *Language Sciences, 17 (2),* 123–145.

Pavlenko, A. (ed.) (2009). *The bilingual mental lexicon: Interdisciplinary approaches*. Bristol: Multilingual Matters.

Poplack, S., & Meechan, M. (1995). Patterns of language mixture: Nominal structure in Wolof–French and Fongbe–French bilingual discourse. In L. Milroy & P. Muysken (eds.), *One speaker, two languages: Cross-disciplinary perspectives on codeswitching*, pp. 191–232. Cambridge: Cambridge University Press.

Poplack, S., & Meechan, M. (1998a). Introduction: How languages fit together in codemixing. *International Journal of Bilingualism, 2 (2),* 127–138.

Poplack, S., & Meechan, M. (eds.) (1998b). *Instant loans, easy conditions: The productivity of bilingual borrowing*. Special issue of *International Journal of Bilingualism*, *2 (2)*. London: Kingston Press.

Poplack, S., & Sankoff, D. (1984). Borrowing: The synchrony of integration. *Linguistics, 22,* 99–135.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics, 26, (1),* 47–104.

Romaine, S. (1989). *Bilingualism*. Oxford: Blackwell.

Samar, R. G., & Meechan, M. (1998). The null theory of code-switching versus the nonce borrowing hypothesis: Testing the fit in Persian–English bilingual discourse. *International Journal of Bilingualism, 2 (2),* 203–219.

Sankoff, D., Poplack, S., & Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change, 2,* 71–101.

Sankoff, G. (2001). Linguistic outcomes of language contact. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (eds.), *The handbook of language variation and change*, pp. 638–668. Oxford: Blackwell.

Shin, S. J. (2002). Differentiating language contact phenomena: Evidence from Korean–English bilingual children. *Applied Psycholinguistics, 23,* 337–360.

Stammers, J. (2010). *The integration of English-origin verbs into Welsh: A contribution to the debate over distinguishing between code-switching and lexical borrowing*. Saarbrücken: VDM Verlag Dr. Müller.

Thomas, R. J. (ed.) (1950–2002). *Geiriadur Prifysgol Cymru* [University of Wales dictionary]. Cardiff: University of Wales. [Concise version of 1st edition: http://www.cymru.ac.uk/geiriadur/gpc_pdfs.htm, retrieved February 15, 2010.]

Treffers-Daller, J. (2005). Evidence for insertional codemixing: Mixed compounds and French nominal groups in Brussels Dutch. *International Journal of Bilingualism, 9 (3/4),* 477–508.

Turpin, D. (1998). "Le Français, c'est le *last frontier*": The status of English-origin nouns in Acadian French. *International Journal of Bilingualism, 2 (2),* 203–219.

Weinreich, U. (1953). *Languages in contact*. New York: Mouton.

Zipf, G. K. (1935). *The psychobiology of language*. New York: Houghton-Mifflin.