

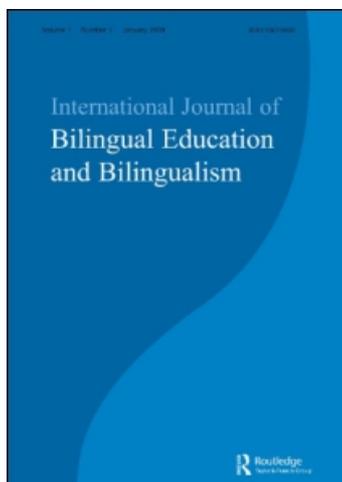
This article was downloaded by: [Bangor University]

On: 19 August 2010

Access details: Access Details: [subscription number 917208385]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Bilingual Education and Bilingualism

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t794297780>

'I saw the *madre*': evaluating predictions about codeswitched determiner-noun sequences using Spanish-English and Welsh-English data

Jon Russell Herring^a; Margaret Deuchar^b; M. Carmen Parafita Couto^b; Mónica Moro Quintanilla^{b,c}

^a Social Sciences Collections and Research, British Library, London, UK ^b ESRC Centre for Research on Bilingualism in Theory and Practice, Bangor University, Bangor, UK ^c Filología Inglesa, Universidad de Oviedo, Asturias, Spain

Online publication date: 16 August 2010

To cite this Article Herring, Jon Russell , Deuchar, Margaret , Couto, M. Carmen Parafita and Quintanilla, Mónica Moro(2010) 'I saw the *madre*': evaluating predictions about codeswitched determiner-noun sequences using Spanish-English and Welsh-English data', International Journal of Bilingual Education and Bilingualism, 13: 5, 553 – 573

To link to this Article: DOI: 10.1080/13670050.2010.488286

URL: <http://dx.doi.org/10.1080/13670050.2010.488286>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

‘I saw the *madre*’: evaluating predictions about codeswitched determiner-noun sequences using Spanish–English and Welsh–English data

Jon Russell Herring^a, Margaret Deuchar^{b*}, M. Carmen Parafita Couto^b and Mónica Moro Quintanilla^{b,c}

^a*Social Sciences Collections and Research, British Library, London, UK;* ^b*ESRC Centre for Research on Bilingualism in Theory and Practice, Bangor University, Bangor, UK;* ^c*Filología Inglesa, Universidad de Oviedo, Asturias, Spain*

(Received 29 January 2009; final version received 15 April 2010)

Previous work on intrasentential codeswitching has noted that switches between determiners and their noun complements are frequent in both Spanish–English and Welsh–English data. Two major recent theories of codeswitching, the Matrix Language Frame model and a Minimalist Program approach, make potentially competing predictions regarding the source language of the determiner in these mixed nominal constructions.

In this paper we evaluate the predictions of each theory with reference to comparable sets of Spanish–English and Welsh–English codeswitching data. Mixed nominal constructions are extracted to test the compatibility of these data with the predictions, taking into account coverage and accuracy.

We find that the data are broadly consistent with each set of predictions but do not find statistically significant differences between the accuracy of the predictions of the two theories. We examine in detail the counterexamples to the predictions of each theory to see what further factors may influence codeswitching patterns between determiners and their nouns, and also discuss the differences in observed patterns in the data from each language pair.

Keywords: codeswitching; bilingual corpora; evaluation of theories; Matrix Language Frame theory; Minimalist approaches to codeswitching

1. Introduction

Models of codeswitching proposed over the last 30 years have attempted to account for the fact that, contra Labov (1972), codeswitching is not random and un-systematic but follows certain predictable patterns. Two major recent theoretical approaches to codeswitching make competing predictions regarding the source language of the determiner in constructions consisting of a determiner in one of the bilingual speaker’s languages followed by a noun in the other language. In this paper the two theories we evaluate are the Matrix Language Frame theory (Myers-Scotton 1993, 2002) and the Minimalist Program approach (MacSwan 1999; Moro in press). We do this with reference to naturalistic data from two language pairs, Spanish–English and Welsh–English.

*Corresponding author. Email: m.deuchar@bangor.ac.uk

The Matrix Language Frame model predicts that in every codeswitched clause, either, but only one, of the languages may provide the morphosyntactic frame for that clause (the matrix language), and that the determiner should come from the matrix language. The Minimalist Program approach predicts that the determiner will always come from the language that has grammatical gender, for reasons to be explained below. Thus competing predictions arise for our Welsh–English and Spanish–English data in one critical context: if the matrix language of a clause in either dataset is English, the Matrix Language Frame model predicts that the determiner will be in English, while the Minimalist approach predicts that the determiner will always be in the other language, Welsh or Spanish.

In the next section we set out some basic definitions, review previous work on determiner-noun switches in the two language pairs, and give an overview of the two theories being evaluated. In the subsequent sections we explain our criteria for comparing the two theories, give full details of the predictions outlined above, and then present the data, analysis and results. Finally we discuss and interpret the results and comment on the implications for the two theories and our own future research.

2. Background

2.1. Codeswitching: definitions

Codeswitching is commonly taken to mean ‘the practice of alternately using two languages’ (Weinreich 1953, 5) or the ‘... rapid succession of several languages in a single speech event’ (Muysken 2000, 1). Examples (1) and (2) show codeswitching in Spanish–English and Welsh–English, respectively. (English words are given in bold typeface in the data from both language pairs.)

- | | | | | | |
|-----|---|---------------------------------|---------------------|--------------------|----------|
| (1) | Yo
PRON.1S ¹
el
det | lo
PRON.3S
doorway | puse
put.1S.PRET | allá
over-there | en
in |
|-----|---|---------------------------------|---------------------|--------------------|----------|

‘I put it over there in the doorway’.

(Jake, Myers-Scotton, and Gross 2002, 82; appendix example 262)

- | | | | | |
|-----|--------------|-------------------|-------------|---------------|
| (2) | Lle
where | mae
be.3S.PRES | (y)r
DET | laptop |
|-----|--------------|-------------------|-------------|---------------|

‘Where’s the laptop?’

(Fusser5, 741²; appendix example 82)

2.2. Mixed nominal constructions

Our unit of analysis in this study is the nominal construction, a term chosen because it is theory-neutral. A nominal construction is also referred to as a determiner phrase (DP) in the Minimalist approach, and a noun phrase (NP) in other approaches (including earlier versions of generative grammar). For this analysis we focused only

on nominal constructions that were ‘mixed’ in the sense that they included a codeswitch between the determiner and the noun. The following are examples:

- | | | | |
|-----|----------------------------|-----------------|---|
| (3) | el
DET
‘the doorway’ | doorway | (Spanish Det + English N)
(appendix example 262) |
| (4) | y
DET
‘the rings’ | rings | (Welsh Det + English N)
(appendix example 61) |
| (5) | the

‘the mother’ | madre
mother | (English Det + Spanish N)
(appendix example 274) |

For our analysis we focus specifically on the prediction that each theory makes regarding the language of the determiner in the nominal construction.

2.3. *Review of the literature*

Previous work on codeswitching in nominal constructions includes the examination of data both from Spanish–English bilinguals (Gingràs 1974; Jake, Myers-Scotton, and Gross 2002; Pfaff 1979; Poplack 1980; Timm 1975; Woolford 1983) and Welsh–English bilinguals (Deuchar 2005, 2006). Timm (1975) found that in informal speech, in written language and in grammaticality judgments from Spanish–English bilinguals, mixed nominals were highly frequent. Pfaff (1979) similarly noted that such mixes were frequent and suggested that these were possible because no structural conflicts arise between the two languages. Poplack (1980) also found frequent Det–N switches (175 of 1835 switches in her data) and proposed an equivalence constraint: ‘Code-switches will tend to occur at points in the discourse where juxtaposition of L1 and L2 elements does not violate a syntactic rule of either language i.e. at points around which the surface structures of the two languages map onto each other’ (Poplack 1980, 586). And Woolford (1983), working within a GB framework (Chomsky 1981), argued that because the syntactic structure is parallel in Spanish and English (NP → Det N’), this allows a Det and an N’ to be inserted from the lexicon of either language. Finally, Jake, Myers-Scotton, and Gross (2002) proposed the Bilingual NP Hypothesis, according to which: (i) determiners in mixed nominal constructions should come from the matrix language of the clause (see Section 2.4), and (ii) nominal constructions in the embedded or other language (e.g. English Det + English N with Spanish matrix language) are permitted but dispreferred. Their results were compatible with both predictions.

Work on Welsh–English bilinguals has also reported that mixed nominals are frequently found in naturalistic data. Deuchar (2005) noted that NPs were the most common type of English insertion into Welsh, and suggested that the high degree of

paradigmatic and syntagmatic congruence of the categories across the two languages accounts for this. Deuchar (2006) tested three principles of the Matrix Language Frame model on Welsh–English data. Nominal constructions containing definite determiners were used to test one prediction, that determiners in bilingual clauses would be in the same language as the matrix language of the clause, unless followed by an NP also in the non-matrix language. The results showed that 41 nominal constructions containing a determiner were mixed, all consisting of a Welsh determiner followed by an English NP. In all these cases the matrix language was determined independently to be Welsh.

2.4. Overview of the two models to be evaluated

Models of codeswitching tend to fall into two main categories, syntactic and processing. Within syntactic models, much of the discussion centres on whether existing aspects of the theory that can account for monolingual grammar can also account for the grammar of bilinguals.

The Minimalist Program approach (MacSwan 1999, 2000, 2005a, 2009) is a syntactic model while the Matrix Language Frame theory, developed by Myers-Scotton (e.g. 2002), incorporates both syntactic and processing aspects. The Minimalist Program approach as developed by MacSwan reflects current Chomskyan syntax and takes the view that no extra apparatus (additional to that needed to account for monolingual speech) should be needed to account for codeswitching. The Matrix Language Frame model is partially based on an earlier model of generative syntax and partially on processing models. Our aim is to compare the extent to which the two models cover our data on nominal constructions and the extent to which the predictions they make are accurate.

2.4.1. The Minimalist model

MacSwan's Minimalist approach to codeswitching (2000) incorporates two central components: a computational system presumed to be invariant across human languages, and a lexicon, to which the differences observed between languages are attributed (MacSwan 1999, 2000, 2005a, forthcoming). The Minimalist Program claims that language variation is reduced to specific morphological realisation features in the lexicon, instead of to different syntactic configurations. As MacSwan (2000, 45) says, 'if all syntactic variation is associated with the lexicon, as in the MP, then code switching may be seen as the simple consequence of mixing two lexicons in the course of a derivation'. An important component of the Minimalist Program is Checking theory, in which features are introduced into the computational system. These features are then checked against those of the lexical items inserted during sentence derivation. MacSwan applies this to bilingualism in that the entries in the lexicon come from different languages and will be differentiated by the features associated with them. See Section 4.1 for an example of how the gender, number and person features are assumed to differ in English and Spanish.

MacSwan (2000, 43) applies general Minimalist Program mechanisms to account for codeswitching: an operation called Select, 'picks items from the lexicon and introduces them into the numeration, an assembled subset of the lexicon used to construct a derivation'. Then another operation, Merge, 'takes items from the numeration and forms new, hierarchically arranged syntactic objects'. Finally, 'the

operation Move applies to syntactic objects formed by Merge to build new structures'. These three operations are constrained by feature checking, a process that ensures the features of related lexical items match at each stage. In MacSwan's model, the grammatical requirements of specific languages are fulfilled by the language-specific features of the lexical items of those languages. MacSwan also suggests that bilinguals have discrete and separate lexicons for the languages they speak, each with its own internal principles of word formation, as well as separate phonological components. Codeswitching will involve the use of features from the lexicons of both languages.

2.4.2. The Matrix Language Frame model

An early article about the model (Myers-Scotton and Jake 1995) is entitled 'Matching lemmas in a bilingual language competence and production model'. Although this suggests that the Matrix Language Frame models (Myers-Scotton 2002; Myers-Scotton and Jake 1995, 2000) speakers' competence as well as being a production model, the article focuses on production. The model is conceived of as having three levels: the conceptual, the functional and the positional. The conceptual level is drawn upon by speakers to express their intentions, activating 'bundles' of features specifying semantic and pragmatic properties. These sets of features cause lemmas from both participating languages to be selected from the mental lexicon, and in turn the lemmas activate the 'formulator', which constructs and orders the utterance itself (Myers-Scotton and Jake 2000).

Myers-Scotton (2002, 14) emphasises the importance of the lexicon in this model, following Levelt (1989) in assuming that a given lexical entry will contain: (a) morphophonological information, (b) syntactic properties, and (c) a semantic and pragmatic representation. Because of the lexically driven nature of the model she argues it is 'not primarily a phrase structure model (i.e. not a syntactic model)' (Myers-Scotton 2002, 15). She also states that it is not a model of language production either, although she says it does 'presuppose [a] model of language production' (Myers-Scotton 2002, 23). What is distinctive about the Matrix Language Frame model, in our view, is that it incorporates assumptions about how language production works, rather than modelling competence alone.

2.4.3. Similarities and differences between the two theories

Both models assume that the clause or sentence is the maximal unit of analysis and therefore exclude intersentential codeswitching from their domain. But for the Matrix Language Frame model, the clause is the minimal as well as the maximal unit, since identification of the matrix language depends on the linguistic information from a full clause being available. The goal of MacSwan and colleagues is to account for bilingual speakers' competence using exactly the same apparatus as used for monolingual speakers, on the grounds of economy. Myers-Scotton's goal, on the other hand, is specifically to model the speech of bilinguals (even though she argues that it can also apply to monolingual speech).

As a result, the focus of the two models differs. Whereas MacSwan's model is a purely representational model of linguistic competence, with no claims as to how this relates to processing in production or comprehension, Myers-Scotton's model makes assumptions about production, as outlined above.

Regarding the data used by each approach to test their models, Myers-Scotton uses only naturalistic data while MacSwan (e.g. 1999) argues that naturalistic data and grammaticality judgements will complement one another. Naturalistic data can be used to test theories (cf. MacSwan 1999, 128) while it is argued that grammaticality judgements are a means of collecting ‘negative evidence’ and are thus necessary to determine what kind of constructions are ill-formed.

2.4.4. Debate between proponents of the two theories

Jake, Myers-Scotton, and Gross (2002) suggest that an integration of the notion of matrix language and current syntactic theory may be possible by reinterpreting the Matrix Language Frame model in the light of Chomsky’s Minimalist Program and vice versa. When addressing Minimalist Program explanations of codeswitching, Jake, Myers-Scotton, and Gross (2002, 69) claim that ‘such an approach may even only succeed partially if it incorporates a basic asymmetry between the languages participating in CS’. This is what they call ‘Modified Minimalism’, where uninterpretable features are conceived as properties of the matrix language, and mismatch of features does not cause any crash of the derivation. Instead, the conflict is resolved in favour of the matrix language, which would allow the derivation to converge.

On the other side of this debate, MacSwan is strongly opposed to the notion of a matrix language. From a theoretical point of view, he argues that the Matrix Language Frame model and the idea of a matrix language should be rejected ‘on grounds of scientific parsimony’ (MacSwan 2005a, 19), as it introduces theoretical constructs that do not play a role in grammatical theory and are not independently useful outside the explanation of bilingual data.

3. Evaluating alternative theories: criteria

While proponents of competing theories of codeswitching may argue for the merits of one theory versus another (see e.g. MacSwan 2005a, 2005b; Jake, Myers-Scotton, and Gross 2005), this is rarely done via a detailed comparison with reference to data. We compare the alternative approaches in terms of the following explicit criteria: (1) coverage; (2) absolute accuracy of predictions for data covered; and (3) comparative accuracy of predictions for data covered by both theories.

COVERAGE refers to the extent of the data about which each theory makes predictions. For our analysis, coverage is measured in terms of the percentage of the data containing determiners where the language of the determiner can be predicted. There may be constructions in the data about which the theory makes no predictions. For example, the Matrix Language Frame model does not make any predictions about the language of the determiner where the matrix language of the clause cannot be determined. For each theory, calculations are performed to determine what percentage of the data is accounted for.

Once we have identified which data are covered by each theory, we then outline the predictions from each theory as to what kind of mixed constructions will occur or not. These predictions can be checked against the data, to assess their accuracy. Another percentage figure is calculated, to indicate ACCURACY.

Finally, in order to COMPARE the accuracy of the predictions from the two theories on the basis of the same data, we look at only that subset of the data for which the two theories make predictions, and calculate comparative accuracy. A statistical comparison of any differences observed is also made.

4. The question to be addressed from each theoretical perspective: in mixed Det + N constructions, which language is the determiner predicted to come from?

4.1. Minimalist approach prediction

In codeswitched nominal constructions consisting of a determiner followed by an NP, the determiner will always come from the language which has grammatical gender – in other words, Welsh or Spanish, not English (see Moro in press).

The reasoning behind this is as follows: Chomsky (2000, 2001) proposes a Minimalist architecture in which features (such as the phi-features of D [or Det], but not N) enter the derivation without values specified. In Chomsky's terms, the phi-features set is a probe that seeks a goal, namely matching features (on the noun) that establish agreement.

The operation AGREE values and deletes these features. Only a probe or functional category (in this case the determiner) with a full set of phi-features is capable of deleting and valuing its own features. Moro (in press) follows Chomsky in that 'we take deletion to be a "one fell swoop" operation, dealing with the phi-set as a unit. Its features cannot selectively delete: either all delete or none' (Chomsky 2000, 124). In our grammatical examples (6a and 7a), the Spanish and Welsh probes (which contain a complete set of phi-features) delete and value their unvalued phi-set:

- (6) a. Spanish D, phi = {person, number, gender}
 English N, phi = {person, number}
 *b. English D, phi = {person, number}
 Spanish N, phi = {person, number, gender}
- (7) a. Welsh D, phi = {person, number, gender}
 English N, phi = {person, number}
 *b. English D, phi = {person, number}
 Welsh N, phi = {person, number, gender}

In the ungrammatical examples (6b and 7b), however, the deficient probe (the English D) lacks gender. Since it only has two phi-features, but is entering into a construction with a noun from Welsh or Spanish that has the additional gender phi-feature, it is incomplete and is unable to delete its uninterpretable features all at the same time. Consequently, the derivation does not converge. So the Minimalist approach predicts Spanish–English mixed nominals such as *el team* but not **the equipo* ('team') and Welsh–English nominals like *y television* but not **the teledu* ('television').

4.2. Matrix Language Frame prediction

According to the BILINGUAL NP HYPOTHESIS (Jake, Myers-Scotton, and Gross 2002, 78–9) the prediction is that determiners in mixed nominal constructions should come from the Matrix Language. This follows from the Uniform Structure Principle, according to

5.2. Data collection procedure

The two corpora may be considered to be comparable data for our purposes in this study given that they were planned, recorded and transcribed by associated teams of researchers at Bangor University using the same methods. The fieldworkers (six in Wales, three in Miami) carrying out the recordings were predominantly balanced bilinguals from within the community. Speakers in both locations were recruited either through the local press or online adverts, or using the friend-of-a-friend approach, and similar information letters in two languages were given to the participants.

All recordings were of pairs of speakers who already knew each other, in a location of their choice, using digital recording equipment. In each session, the fieldworker set-up the equipment and then left the room for about 35 minutes before returning to administer a detailed questionnaire (in the language of the participant's choice). For more details see Section 5.3. The recordings for both corpora were edited to remove the first 5 minutes of conversation, and transcribed in the CHAT format (MacWhinney 2007) according to the LIDES conventions for codeswitching data (Gardner-Chloros 2009, Appendix; The LIPPS Group 2000).

By using similar data collection methodologies, we aimed to maximise the comparability of the two datasets. There remain some differences between the two corpora in terms of size and time of collection. The Welsh–English corpus is larger (40 hours) and was collected between 2005 and 2008, while the Spanish–English corpus is approximately 20 hours and was recorded over 2 months in early 2008. With regard to the subsets of data used for the analysis, the number of speakers sampled reflects the relative size of each corpus as a whole (40 Welsh–English bilinguals, compared to 19 Spanish–English).

5.3. Profiles of speakers

Both datasets were more or less balanced for speaker gender: of the 40 Welsh–English bilinguals in the sample, 18 were male and 22 female, and in the Spanish–English corpus sample, nine were male and 10 female. In the Welsh–English dataset, the age range of speakers was wide (from 15 to 89), while in the Spanish–English one it ranged from 15 to 57.

We did not have detailed contextual data for the speakers in the compiled set of Spanish–English data. However, we do know that they come from bilingual speakers in both the USA and Spain.

For all the speakers in our own corpora, however, we had access to detailed biographical and sociolinguistic information gathered by means of a questionnaire after the recording. This provided information on extralinguistic variables including age of acquisition of each language, parental input, self-reported proficiency, the language of education, perceptions of identity and social network.

In comparing the questionnaire responses for the two sets of speakers, we found some similarities as well as differences. In both sets most speakers had acquired Welsh or Spanish under the age of 2 years, but this was not the case for the acquisition of English in the majority of speakers. While most speakers in both sets had acquired English at the latest in primary school, there were more Spanish–English than Welsh–English speakers who had acquired English in secondary school or as adults. Most speakers had heard Welsh or Spanish from their mothers or fathers or both, while a minority of speakers in both sets (Welsh–English set $N = 15/40$, Spanish–English

both covers this instance and makes the right prediction about it. The Minimalist Program prediction is also that the determiner should come from Spanish and so it too is correct for this piece of data.

The same process is shown for (11):

(11)	Y	después	de	hacer	el	research
	And	after	of	do.INF	DET	
	'And after doing the research'.					(Jake, Myers-Scotton, and Gross 2002, 83; appendix example 263)

In this example, no matrix language can be assigned because the clause only has a non-finite verb in it. Thus the Matrix Language Frame model cannot make a prediction regarding the language of the determiner. The Minimalist approach does cover this example, and correctly predicts that the determiner should come from Spanish.

This process was repeated for all the mixed determiner-noun constructions extracted from the data of the two language pairs. The full data are included in an appendix.⁴

In the results presented in the next section, the percentage figures are presented so as to show the differences in coverage and accuracy between the two theories. For the Matrix Language Frame model predictions, the ability to make a prediction is contingent on there being enough linguistic context to assign a matrix language independently. To make the comparison of accuracy as balanced as possible, we distinguished between two potential reasons why the model was unable to predict if a matrix language was not assignable:

- If the Spanish–English data taken from published sources (the compiled data) comprised only the nominal construction cited out of context, this was seen as outside the testable evidence for or against the Matrix Language Frame approach, and so it did not count for the percentage calculations regarding coverage or accuracy.
- If there was a full clausal context but the main verb was not finite, as in (11), then this is deemed to be a legitimate lack of coverage on the part of the theory, and so it was taken into account in the calculations to determine coverage.

7. Results

7.1. *Quantitative analysis*

7.1.1. *Welsh–English data*

In Table 1 we give the results for the predictions about the language of the determiner taken from the Welsh–English data.

The difference in coverage between the two theories is quite marked since the Matrix Language Frame model is only able to make predictions for 69.78% of the data. Many nominal constructions occurred as fragments of clauses, which meant that there was no clause to use to identify a matrix language.

Table 1. Mixed Welsh–English nominal constructions: source language of determiner.

	Minimalist Program	Matrix Language Frame
Percentage for which the model can make predictions	225/225 100%	157/225 69.78%
Percentage where the model can make predictions and they are correct	225/225 100%	155/157 98.73%
Percentage of shared data for which predictions are accurate	157/157 100%	155/157 98.73%

The second row of the table indicates the accuracy of each theory in relation to those constructions about which a prediction could be made. The Minimalist approach was 100% accurate while the Matrix Language Frame model incorrectly predicts two instances in the data. We discuss these counterexamples in Section 7.2.

7.1.2. Spanish–English data

The Spanish–English data are made up of two sets which are presented together in Table 2. The two left-hand columns give the results from the compiled set of data taken from existing articles on Spanish–English codeswitching.

Of the 48 constructions in the compiled set, the Minimalist approach was again able to make predictions about all of them. However, for the Matrix Language Frame model, only 40 of the constructions were supplied with their full clausal context by the authors of the papers, and so we took this as the denominator of the fraction for our calculation. Of these 40, two of the constructions came from clauses where there was no finite verb with which to assign a definitive matrix language, so we took 38 as the numerator of this fraction. The corpus dataset (results given in the two right-hand columns) shows that, in parallel with the results from the compiled set, the Matrix Language Frame model has, at 90%, lower coverage.

Table 2. Mixed Spanish–English nominal constructions: source language of determiner.

	Compiled dataset		Corpus dataset	
	Minimalist Program	Matrix Language Frame	Minimalist Program	Matrix Language Frame
Percentage for which the model can make predictions	48/48 100%	38/40 ^a 95%	100/100 100%	90/100 90.00%
Percentage where the model can make predictions and they are correct	45/48 93.75%	36/38 94.74%	92/100 92.00%	88/90 97.78%
Percentage of shared data for which predictions are accurate	35/38 92.11%	36/38 94.74%	82/90 91.11%	88/90 97.78%

^aOnly 40 of the 48 constructions in the compiled set were supplied with their full clausal context by the authors of the papers, and so we took this as the denominator of the fraction for our calculation.

For the compiled set, the Minimalist approach made correct predictions for 93.75% of the constructions, while the Matrix Language Frame model did so for 94.74% of the constructions. The Minimalist approach is shown to have 92.11% accuracy for shared data compared with 94.74% for the Matrix Language Frame model. The patterns are the same for the corpus set, where the Minimalist approach has 91.11% accuracy for the shared data while the Matrix Language Frame model has 97.78%. Thus for all the Spanish–English data taken together, the results diverge from the Welsh–English data, as this time the Matrix Language Frame model appears to have more accurate predictions.

Testing for the theories' accuracy on more than one language pair was important, given the different outcomes observed for the two pairs. In our Welsh–English sample, the matrix language of the clause surrounding the mixed nominal constructions was Welsh in all but two examples (see 13 and 14). This meant that, for the vast majority of cases, although each model arrived at its prediction by different routes, the predictions for both models were the same: that the determiner should come from Welsh. The Minimalist approach led to this prediction because Welsh has grammatical gender while English does not, and the Matrix Language Frame model did so because Welsh was overwhelmingly the matrix language of the clause in these data. Data for which the predictions of each theory would be different in more cases were needed to provide a more stringent test, and this was achieved with the Spanish–English data.

In the Spanish–English data, a much higher proportion, 7.38% of the surrounding clauses (11/149), had English as the matrix language. This made these data more useful for comparing the two models, because clauses with English matrix language are those where the predictions of the two theories contrast with one another; here the Minimalist approach still predicts that the determiner should come from Spanish as it has grammatical gender, while the Matrix Language Frame predicts that it will come from English.

A chi-square test of independence was run for both the Welsh–English and Spanish–English corpus datasets in order to test whether or not there was a difference in the accuracy of the given predictions from each model. As might be expected from the percentage figures shown in Tables 1 and 2, there was no statistically significant difference between the two models with regard to the accuracy of their predictions for either the Welsh–English or the Spanish–English corpus data.⁵

In summary, we found that the application of the predictions made by the two theories to our two sets of naturally occurring bilingual data brought out the following differences. Firstly, in both language pairs, the Minimalist approach had a higher level of *COVERAGE* for all the corpus data considered. This can be attributed to the difference in the theories' minimum unit of analysis as remarked upon in Section 2.4. Where the Minimalist approach needs only to see the nominal construction (DP) to make its prediction, the Matrix Language Frame model requires a surrounding clause (CP) with an assignable matrix language in order to make its own predictions. This context is not always available because of the nature of spontaneous speech, in which clauses (CPs) are not necessarily complete or coherent and the assignment of a matrix language is not possible.

Secondly, for the language pairs under consideration, it was found that the *ACCURACY* of the two theories in predicting which Determiner–NP combinations are possible was different, but not significantly so. Neither theory made predictions that

The first of the counterexamples from the compiled Spanish–English data, as in (15), is the only one to deviate from the predictions of both theories and can therefore be considered highly idiosyncratic:

- (15) **the** pendejo no tiene huevos
 the asshole.SM NEG have.3S.PRES egg.PLM
 ‘the asshole doesn’t have balls’
 (Pfaff 1979, 313; appendix example 251)

This might be an example of triggering (cf. Clyne 1967). The speaker begins in English but appears to favour the Spanish word *pendejo* to express his or her communicative intention. This could be because the word *pendejo* is used frequently in Spanish monolingual contexts by the speaker, or because despite beginning the clause in English, the speaker cannot access the word in English at the moment of speech. It could also be because the Spanish word has an emotive or connotative force for the speaker that its English translation does not convey. However, once the switch has been made into the other language, the speaker completes the clause in Spanish.

Also found in the compiled data were (16) and (17), counterexamples to the Minimalist approach, but consistent with the Matrix Language Framework approach. Like the Welsh–English counterexamples they were both produced by one speaker in the same conversation and this time they contain the same noun:

- (16) **but I used to eat the** bofe, **the brain**
 but I used to eat the brain.SM the brain
 ‘but I used to eat the brain, the brain’. (Poplack 1980, 597; appendix example 229)

- (17) **I used to make some** bofe!
 I used to make some brain.SM
 ‘I used to make some brains!’ (Poplack 1980, 597; appendix example 230)

It is likely that *bofe* is a word that the speaker associates with Spanish, both culturally and in usage. This usage may have been increasingly infrequent for Poplack’s informant (who was living in the USA when the data were collected), given that it is employed after *used to* in both clauses. Like *pendejo* in (15) it may be a word that the speaker rarely uses in English linguistic and cultural contexts, and so the Spanish equivalent, more easily accessed, is inserted. For further discussion on how this noun might be considered a loan rather than a switch, and therefore might be accommodated in the Minimalist approach, see Section 7.3.

Example (18) is predicted by the Minimalist approach but not by the Matrix Language Frame model:

- (18) **we have it planned for October twenty-ninth a** las seis
 en el Methodist Student Center DET six
 in DET

‘we have it planned for October twenty-ninth at six o’clock at the Methodist Student Center’.

(Poplack 1980, 311;
appendix example 249)

In this example the matrix language is English, which predicts the determiner of *Methodist Student Center* to be in English too. However, the speaker switches to Spanish for the adverbial *a las seis* ‘at six o’clock’ and this switch may trigger the production of the beginning of the next prepositional phrase in Spanish. The switch back to English for the NP *Methodist Student Center* may be explained by the likelihood that the location is primarily known by its English name.

The remaining examples come from our Spanish–English corpus, where again we have access to supplementary information collected at the same time as the speech data, as well as the linguistic and discourse context. The first example (19) was produced by a 27-year-old female simultaneous balanced bilingual. She reported a negative attitude to codeswitching per se, but reported that she did it frequently herself in natural conversation. She was talking to a bilingual cousin of hers who had recently moved to Miami:

- (19) **when I went to Canada I saw the** madre
mother.SF
‘when I went to Canada I saw the mother’. (Herring1, 464;
appendix example 274)

Here she is referring to her (monolingual English) boyfriend’s mother, and thus contextually has no reason to refer to the woman or her identity in Spanish. However, she seems to be adding pragmatic information to the event, perhaps that she has an uneasy relationship to her partner’s mother. Such non-syntactic and indeed non-production-orientated features of speech are beyond the predictions of both theoretical approaches assessed in this paper.

Example (20) is something of a mirror image of (15):

- (20) pero mi **printer doesn’t work with the**
but POSS.1S **old computer that we have**
‘but my printer doesn’t work with the old computer that we have’.
(Herring 3, 131; appendix
example 284)

Here the English noun ‘printer’ seems to trigger a switch to English for the rest of the clause and hence for English as matrix language, which then mismatches the possessive *mi* in the mixed nominal construction. The lack of overt gender agreement on *mi* suggests that perhaps it should be treated differently from determiners *ella* in mixed Spanish–English nominal constructions. This is matter for future research.

Examples (21) and (22) are spoken by the same person, a 26-year-old male L2 English speaker who learned English at secondary school in Cuba, and then moved to Miami at the age of 19:

- (21) **go to go to Wikipedia tha(t)** enciclopedia **online**
 go to go to Wikipedia DEM.3S encyclopedia.SF online
 'go to go to Wikipedia that encyclopedia online'. (Herring7, 304;
 appendix example 298)

Here he inserts the Spanish (cognate) word for *encyclopedia* after the English determiner, which in principle contradicts the prediction of the Minimalist approach. However, note that the pronunciation of the determiner is itself quite marked (the deletion of the coda consonant [t]) and is also an indicator of L2 speaker status rather than balanced bilingual. Perhaps the speaker is not confident of pronouncing the word with English phonology, or treats the lexical item as the same in both languages given that both forms are structurally similar. His relative lack of fluency in English is also displayed in (22):

- (22) **you know** el el **bad guy the bad guy in part two?**
 DET DET
 'you know the . . . the bad guy in part (Herring7, 667; appendix example 301)
 two?'

In this example English is the matrix language but the speaker switches to Spanish for the definite article *el*. However, note that this seems to cause him some disfluency which leads him to repeat *el* at first as if searching for a Spanish nominal complement. When he nevertheless produces the English NP *bad guy*, perhaps because the television programme he is referring to was in English, he apparently repairs this by repeating *bad guy*, but this time preceded by the expected English definite article *the*.

The last four examples are all produced by a 21-year-old female second generation Cuban American, who had lived in Miami all her life but both of whose parents spoke to her in Spanish at home. She only began learning English at primary school. All of the counterexamples spoken by her fulfil the predictions of the Matrix Language Frame model but contravene those of the Minimalist approach:

- (23) **she got the** manguera
 hose.SF
 'she got the hose'. (Herring9, 644; appendix example 302)

- (24) **he my mom got the** manguera
 hose.SF
 'he . . . my mom got the hose'. (Herring9, 651; appendix example 303)

- (25) **started hitting him with the** manguera
 hose.SF
 'started hitting him with the hose'. (Herring9, 652; appendix example 304)

This would mean that a word like *madre* would have lost its Spanish gender feature while becoming fully integrated into the English lexicon. Unlike the examples in (6b) and (7b) in Section 4.1, the determiner in (28) would not be deficient and so the derivation would not crash.

However, this approach ignores another generalisation resulting from our analysis, that there is a strong tendency in both sets of data for the language of the determiner to be the same as the language of the finite verb. This generalisation is supported by 280 cases, and only refuted by five. It is also compatible with the predictions of the Matrix Language Frame model, where the matrix language is the language of the finite verb. One striking finding here is that, whereas the Spanish data contained clauses with both Spanish and English as the matrix language (although only 32/149 or 21.5% of the clauses here had English as the matrix), the Welsh data contained almost exclusively clauses with Welsh as the matrix language.

This prevalence echoes the findings in Deuchar (2006) where of the 163 bilingual clauses analysed, 141 had Welsh as the matrix language, 18 clauses had no identifiable matrix language and only four had English as the matrix language. It is possible that syntactic factors (Chan 2009, 197), attitudinal factors, contexts of data collection and the like may have contributed to the predominance of Welsh as the matrix language. Further research is under way to determine whether these differences could be due to the language balance of the speakers, community norms or individual factors.

8. Conclusion

In this paper we have noted a contrast in the language of the determiner in mixed nominal constructions taken from our Welsh–English and Spanish–English data. We have explored two accounts of our data from two different theoretical standpoints.

The Minimalist account, predicting that the determiner would come from the language with grammatical gender, was shown to be successful in accounting for all the Welsh–English data and most of the Spanish–English data. However, when we examined the language of the verb in the clause containing the mixed nominal constructions as part of the process of testing the Matrix Language Frame model, we observed that the success of the Minimalist account was due to the fact that the language of the verb was almost always Welsh or Spanish, i.e. a language with grammatical gender. In the small number of clauses where the finite verb was in English, an English determiner was almost always found, contrary to the predictions of the Minimalist approach.

This led us to broaden our focus from predicting the language of the determiner to considering the counterexamples to the predictions of the theories. As suggested in the discussion above, both linguistic and extralinguistic factors need to be explored, and we plan to do this in future research.

Our study has demonstrated the value of naturalistic corpus data in helping to establish that codeswitching follows discernible patterns. It also allows us to reinforce the observation that codeswitching is a complex phenomenon which promises only to be fully understood by combining a narrow with a broader focus, and a syntactic approach with wider sociolinguistic, pragmatic and psycholinguistic approaches.

Acknowledgements

We are grateful to Peredur Davies, Siân Lloyd-Williams and Myfyr Prys for their help with the analysis and presentation of the Welsh–English codeswitching data, to Kara McAlister for creating a subsample of the corpus, and to Diana Carter for invaluable editorial assistance and for running the statistical analyses. We would also like to thank the following for their comments on earlier drafts of this paper: Katja Cantone, Michael Clyne, Marcel Den Dikken, Virginia Gathercole, Jeff MacSwan, Kara McAlister, Gerald Stell, Enlli Thomas and two anonymous reviewers.

The support of the Arts and Humanities Research Council (AHRC), the Economic and Social Research Council (ESRC), the Higher Education Funding Council for Wales and the Welsh Assembly Government is gratefully acknowledged. 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.

Notes

1. Key to glosses: PL, plural; S, singular; PRON, pronoun; POSS, possessive pronoun; DET, determiner; PRT, particle; NEG, negative/negative particle; NFIN, non-finite; PRES, present; IM, interactional marker; IMP, imperfect tense; CONDIT, conditional tense; FUT, future tense; PAST, past tense; NPAST, non-past tense; IMPER, imperative; 1S/1PL, first person singular/plural; 2S/PL, second person singular/plural; 3S/3PL, third person singular/plural.
2. All Welsh examples come from our own data. Labels give the name of the transcript file, and line number of the example.
3. The Welsh data were originally collected and transcribed by Marika Fusser, Elen Robert, Peredur Davies, Jonathan Stammers, Siân Lloyd-Williams and Margaret Deuchar as part of the project ‘Code-switching and Convergence in Welsh’, awarded by the AHRC to Margaret Deuchar. They are available via TalkBank (<http://www.talkbank.org>) in the BilingBank archive.
4. The appendix is available online at <http://bilingualism.bangor.ac.uk/publications/documents/HerringEtAl2010.pdf>.
5. We conducted similar analyses for coverage, even though an analysis of the coverage of the two models may seem superfluous, since the Minimalist approach has 100% coverage by definition, and the Matrix Frame model does not in practice. However, we argue that it may be seen as an additional way of quantifying the shortfall in coverage of the Matrix Language Frame. We found that χ^2 analyses confirmed significant differences in coverage for both the Spanish–English corpus ($\chi^2(1, N = 200), p < 0.001$) and the Welsh–English data ($\chi^2(1, N = 456), p < 0.0001$).

References

- Chan, B.H. 2009. Code-switching with typologically distinct languages. In *The Cambridge handbook of linguistic code-switching*, ed. B.E. Bullock and A.J. Toribio, 182–98. New York: Cambridge University Press.
- Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht, The Netherlands: Foris.
- Chomsky, N. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. R. Martin, D. Michaels, and J. Uriagereka, 89–155. Cambridge, MA: MIT Press.
- Chomsky, N. 2001. Derivation by phase. In *Ken Hale: A life in language*, ed. M. Kenstowicz, 1–52. Cambridge, MA: MIT Press.
- Clyne, M. 1967. *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia*. The Hague: Martinus Nijhoff.
- Deuchar, M. 2005. Congruence and code-switching in Welsh. *Bilingualism: Language and Cognition* 8, no. 3: 255–69.
- Deuchar, M. 2006. Welsh-English code-switching and the Matrix Language Frame model. *Lingua* 116, no. 11: 1986–2011.

- Gardner-Chloros, P. 2009. *Code-switching*. Cambridge: Cambridge University Press.
- Gingrás, R. 1974. Problems in the description of Spanish–English intrasentential code-switching. In *Southwest areal linguistics*, ed. G. Bills, 167–74. San Diego, CA: Institute for Cultural Pluralism, University of California at San Diego.
- Jake, J.L., C. Myers-Scotton, and S. Gross. 2002. Making a Minimalist approach to codeswitching work: Adding the Matrix Language. *Bilingualism, Language & Cognition* 5, no. 1: 69–91.
- Jake, J.L., C. Myers-Scotton, and S. Gross. 2005. A response to MacSwan (2005): Keeping the matrix language. *Bilingualism: Language & Cognition* 8, no. 3: 271–6.
- Labov, W. 1972. *Sociolinguistic patterns*. Oxford: Blackwell.
- Levelt, W.J.M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- MacSwan, J. 1999. *A Minimalist approach to intra-sentential code-switching*. New York: Garland.
- MacSwan, J. 2000. The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language & Cognition* 3, no. 1: 37–54.
- MacSwan, J. 2005a. Codeswitching and generative grammar: A critique of the MLF model and some remarks on “modified minimalism”. *Bilingualism: Language & Cognition* 8, no. 1: 1–22.
- MacSwan, J. 2005b. Remarks on Jake, Myers-Scotton and Gross’s response: There is no “matrix language”. *Bilingualism: Language & Cognition* 8, no. 3: 277–84.
- MacSwan, J. 2009. Generative approaches to code-switching. In *The Cambridge handbook of linguistic code-switching*, ed. B.E. Bullock and A.J. Toribio, 309–35. New York: Cambridge University Press.
- MacSwan J., ed. Forthcoming. *Grammatical theory and bilingual codeswitching*. Linguistics series. Cambridge, MA: MIT Press.
- MacWhinney, B. 2007. The TalkBank project. In *Creating and digitizing language corpora: Synchronic databases, vol. 1*, ed. J.C. Beal, K.P. Corrigan, and H.L. Moisl. Houndmills: Palgrave Macmillan.
- Moro, M. in press. The semantic interpretation and syntactic distribution of determiner phrases in Spanish/English codeswitching. In *Grammatical theory and bilingual codeswitching*, ed. J. MacSwan. Cambridge, MA: MIT Press.
- Moyer, M.G. 2000. Negotiating agreement and disagreement in Spanish–English bilingual conversations with “no”. *The International Journal of Bilingualism* 4, no. 4: 485–504.
- 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: Oxford University Press.
- Myers-Scotton, C. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Myers-Scotton, C., and J. Jake. 1995. Matching lemmas in a bilingual language production model: Evidence from intrasentential codeswitching. *Linguistics* 33, no. 5: 981–1024.
- Myers-Scotton, C., and J. Jake. 2000. Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics* 38, no. 6: 1053–100.
- Pfaff, C.W. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55, no. 2: 291–318.
- Poplack, S. 1980. Sometimes I’ll start a sentence in Spanish *y termino en Español*: Toward a typology of code-switching. *Linguistics* 18, nos. 7/8: 581–618.
- Poplack, S., and D. Sankoff. 1984. Borrowing: The synchrony of integration. *Linguistics* 22, no. 1: 99–135.
- The LIPPS Group. 2000. The LIDES coding manual: A document for preparing and analyzing language interaction data. *International Journal of Bilingualism* 4, no. 2: 131–270.
- Timm, L.A. 1975. Spanish–English code-switching: El porqué y how-not-to. *Romance Philology* 28, no. 4: 473–82.
- Weinreich, U. 1953. *Languages in contact: Findings and problems*. The Hague: Mouton.
- Woolford, E. 1983. Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14, no. 3: 520–36.