# Welsh-English code-switching and the Matrix Language Frame model[☆]

Margaret Deuchar [*]

*Department of Linguistics and English Language, University of Wales, Bangor, Gwynedd LL57 2DG, UK*

## Abstract

The main aim of this paper is to test the Matrix Language Frame (MLF) model of code-switching with Welsh-English data, and to determine whether Welsh-English is a ''classic case'' of code-switching in that a matrix language can be clearly identified in bilingual clauses. The MLF model is currently one of the most influential in this field, and has been successfully applied to several other language pairs. Three principles associated with it, the Matrix Language Principle, the Asymmetry Principle and the Uniform Structure Principle, were tested on data collected from informal conversations by bilingual Welsh-English speakers located in north-west Wales. The results of the analyses suggest general support for the three principles and for labelling Welsh-English code-switching ''classic''. These results are compared with those from other studies of classic and other kinds of code-switching. Finally, the implications for the MLF model are examined.
© 2005 Elsevier B.V. All rights reserved.

*Keywords:* Welsh; Code-switching; Matrix Language Frame model

## 1. Aims

The main aim of this paper is to test the Matrix Language Frame (MLF) model of code-switching with Welsh-English data. This will provide a further empirical test of this

[*] Tel.: +44 1248 382264; fax: +44 1248 382928.
*E-mail address:* m.deuchar@bangor.ac.uk.

influential model as well as determine whether or not Welsh-English is a case of ''classic code-switching'' (Myers-Scotton, 2002:9).

## 2. Terminology and unit of analysis

### 2.1. Code-switching

The term *code-switching* has been defined in various ways in the literature (see, e.g. Clyne, 2003:70–73). *Intra-sentential switching* is often differentiated from *inter-sentential switching*: the former involves a switch between languages within the clause, whereas the latter involves switching between clauses or sentences. In this paper we shall focus on intra-sentential switches within the clause and shall take as our unit of analysis the bilingual clause: this will be defined as a clause containing one or more morphemes from more than one language. The following are some examples of utterances containing bilingual clauses from Welsh-English code-switching data:[1]

(1)  yn fan'a     gafon        ni    **staff meeting** un  tro  ia        [MEP182]
     in  place-there have.3P.PAST PRON.1P staff  meeting  one time yeah
     'That's where we had a staff meeting once isn't it?'

(2)  oedd      o'-n          hollol    **ridiculous** yn  doedd        [MEG193]
     be.3S.IMP PRON.3S-PRT completely ridiculous   PRT NEG-be.3S.IMP
     'It was completely ridiculous wasn't it?'

(3)  dw        i'n    **love-**io  **soaps**                          [B12B8]
     be.1S-PRES 1S-PRT love-VBZ soaps
     'I love soaps'

(4)  mae       o-'n          reit  **camouflaged** yn  dydi          [MEW50]
     be.3S.PRES PRON.3S-PRT quite camouflaged   PRT NEG-BE.3S.PRES
     'he's quite camouflaged isn't he?'

(5)  mae       o-'n          fath-â **catching**                     [T112A2]
     be.3S.PRES PRON.3S.M-PRT sort-of catching
     'it's sort of catching'

(6)     oedd    o-'n       gar tua   **seven seater**         [MEG73]
        be.3S.IMP PRON.3S.M-PRT car about seven seater
        'It was about a seven seater car'

## 2.2. *Switches versus loanwords*

In identifying bilingual clauses as illustrated above, loanwords do not count as coming from a different language from the language into which they are borrowed. So, example (7) is not classified as a bilingual clause because the two English words it contains, *ffitio* 'fit' and *cwpwrdd* 'cupboard', are loanwords.

(7) doedd       'na  ddim byd   yn ffitio i  mewn i'-r    cwpwrdd 'ma [MEG164]
    NEG-BE.3S.IMP there NEG   world PRT fit    to inside to-DET cupboard here
    'There was nothing that fitted inside this cupboard'

However, clear criteria are necessary for distinguishing loans like *cwpwrdd* from single word switches like *ridiculous* in (2). There is in fact considerable controversy among code-switching researchers about exactly how to draw the line between single word loans and single word switches (see, e.g. Poplack and Sankoff, 1984; Poplack and Meechan, 1998; Muysken, 2000:72; Myers-Scotton, 2002:41). Poplack and Meechan (1998) consider most single-word switches not to be true switches, but to behave like loans: this position means that they exclude most ''lone other-language items'' from their analysis of code-switching. Myers-Scotton, on the other hand, considers single-word switches to be evidence of the validity of the Matrix Language and Asymmetry Principles (see section 4.1.1 and section 5.3 below), according to which one language is more activated than the other in bilingual speech production. Since our aim in this paper is to test the Matrix Language Frame model, we shall include single-word switches in our analysis, distinguishing them from loans in terms of their predictability (Myers-Scotton, 2002:41). This criterion is related to ''listedness'', or ''the degree to which a particular element or structure is part of a memorized list'' (Muysken, 2000:71). Loans are assumed to be listed in the vocabulary of monolingual speakers of the recipient language, whereas switches are not. However, a criterion of knowledge by monolingual speakers cannot be applied to a language, like Welsh, which has virtually no monolingual speakers. We shall therefore assume that, for Welsh, listedness is reflected in established dictionaries of the language, in particular the comprehensive *Geiriadur Prifysgol Cymru*[2] for Welsh. If an English-origin item is found in this Welsh dictionary, we shall assume that it is a loan, whereas, if it is absent, we shall assume that it is a switch. For example, we shall consider words like *ffitio* 'fit' and *cwpwrdd* 'cupboard' in (7) to be loanwords, since they are listed in the dictionary, whereas *ridiculous* in (2) is not considered to be a loanword, because it does not appear in the dictionary. This is admittedly an entirely arbitrary criterion, especially since we recognise that dictionaries reflect usage at an earlier point in time rather than the present. The main consequence of this fact for our analysis is that we may identify some single words as switches when they

---

[2] The University of Wales Dictionary, the most authoritative and comprehensive Welsh dictionary.

should be considered loans. This could have the effect of inflating the proportion of bilingual clauses containing a single-word switch as opposed to a switch of several words. However, we shall guard against this danger by performing an additional analysis which excludes all single word switches and comparing this with the more inclusive analysis (see section 5.3).

## 3. Bilingualism in the Welsh context

According to figures from the last census in 2001, Welsh is spoken by 582,368 people living in Wales, or 20.8% of the population of the principality (http://www.bwrdd-yr-iaith.org.uk). Virtually all Welsh speakers of school age and above are bilingual in Welsh and English, providing ample opportunity for language contact. Although the degree of bilingualism has increased among Welsh speakers throughout the twentieth century (see, e.g. Jones, 1993:550), there has been some English-Welsh language contact for more than a thousand years (see Aitchison and Carter, 1994) and this has been particularly intensive since the Act of Union of 1536, which gave English privileged status in Wales (see Jones, 1993:539).

Welsh lost ground after 1536, with temporary respite as a result of the Reformation and Renaissance (see Jones, 1993:541), and the census results from the twentieth century show that the proportion of speakers declined at least until the 1980s. However, the 2001 results show a slight increase which may show an arrest in that decline. This coincides with ever-increasing efforts at the revitalization of Welsh (see Williams, 2000), which include the availability of Welsh-medium education in most areas and enhanced legal status for Welsh. Thus, a period of relatively stable bilingualism may be said to characterize the last 10 years or so.

It should also be noted that the proportion of Welsh (and therefore bilingual) speakers in Wales varies considerably from region to region, but that Welsh is currently spoken by the highest proportion of speakers in Gwynedd, where 69% of the population speak Welsh according to the 2001 census. This is the region where most of the data to be analysed here were collected.

### 3.1. Earlier research on Welsh-English language contact

Ever since bilingual speakers have existed, it is likely that there has been at least some code-switching, although the switching of individual words is not distinguished from loans by earlier writers. For example, Fynes-Clinton (1913:ii), in a compilation of *The Welsh Vocabulary of the Bangor District*, includes English words "which have demonstrably formed part of the language for several centuries and those of very common and indispensable use" without distinguishing between loans and switches. Parry-Williams (1923:11) records examples of what he calls "loan-words" found in written (twelfth-century) Welsh texts from the Old English period as well as from the "Middle-and-New English" period following this. Sweet (1884) is possibly the first study which records extended examples of spoken (as opposed to literary) Welsh including "words taken directly from English" (Sweet, 1884 in Wyld, 1913:541). Again, Sweet does not

distinguish between loans and switches. (8) is an example of ''Spoken North Welsh'' from Sweet's own notes, in which he identifies English words with italics and provides his own translation:

(8)        bydd *season* brithylliaid yn dechreu oflaen *season samon*
           ''Is season trouts beginning before season salmon''

<div align="right">(Sweet, 1884 in Wyld, 1913:558–559)</div>

No distinction is made here between English-origin loanwords like *samon* which are listed in the dictionary as dating from before the time when Sweet was writing, and words like *season* which are not listed at all and which would be considered switches following the dictionary criterion used in this study.

## 3.2. Previous research on Welsh-English code-switching

Although she does not use the term *code-switching*, Thomas (1982) gives an example (9) of what she calls ''mixed language'' used in a radio interview:

(9)        Gallan       nhw        ddod  **in touch** efo  ni          yn yr   **Employment Office.**
           can.PRES.3P PRON.3P come in touch with PRON.2P in DET Employment Office
           'They can get in touch with us in the Employment Office.'

Lindsay (1993) explicitly uses the term *code-switching* in relation to the use of English and Welsh in Bangor. Lindsay (1993:15) says that ''This phenomenon is not necessarily restricted to high or low domains and can be found in the written language as well as the spoken''. Jones (1995) provides actual examples of Welsh-English code-switching from encounters in a post office and in a chemist's shop, and contrasts these with a religious event where there is an absence of code-switching. She also considers the relation between written and spoken texts, a theme she continues in Jones (2000), where she investigates bureaucratic Welsh-English discourse. In general, the emphasis in the few studies conducted so far on Welsh-English code-switching has been more social than structural.

## 4. The Matrix Language Frame model

The Matrix Language Frame (MLF) model was first proposed in detail in a book-length study by Myers-Scotton (1993), who acknowledges the contribution of Joshi (1985) in arguing for an asymmetry in the participation of the languages involved in code-switching. Joshi (1985:190–191) argued that ''despite extensive intrasentential switching, speakers and hearers usually agree on which language the mixed sentence is 'coming from'. We call this language the *matrix language* and the other language the *embedded language*.'' Joshi predicted that in a mixed language sentence all ''closed class'' items would come from the matrix language while ''open class'' items would come from either the matrix or the embedded language. Closed-class items included determiners, quantifiers, prepositions, possessives, ''Aux'', and ''Tense'', while open class items included, for instance, nouns,

verbs and adjectives. The open/closed-class distinction in Joshi's model has been replaced in the MLF by a distinction between ''content'' and ''system'' morphemes. The MLF model has been further developed in a series of publications including, for example, Myers-Scotton and Jake (1995, 2000), Myers-Scotton (2002), and Jake et al. (2002). Myers-Scotton and Jake (2000) introduce the 4-M ('four morpheme') model which distinguishes content from system morphemes as before, but divides the system-morpheme category into three subtypes: early system morphemes, bridge late system morphemes and outside late system morphemes. The distinction between these categories will be further outlined and illustrated in section 5.2 below. It is argued not only to describe ''classic'' code-switching adequately, but to do this in a way which accounts for bilingual language production. Clyne (2003:81) describes it as ''the most comprehensive and influential current framework'' in the field of language contact. The framework has also been extended to describe and account for language convergence and attrition (see Myers-Scotton, 2002) and to include, for instance, creole formation (see also Gross, 2000).

## 4.1. Principles of the MLF model

The MLF model is guided by several principles, three of which give rise to specific hypotheses to be tested here: the Matrix Language Principle, the Asymmetry Principle and the Uniform Structure Principle. While the Matrix Language and the Asymmetry Principles are specific to the MLF model, the Uniform Structure Principle is intended to apply to all bilingual speech (Carol Myers-Scotton, personal communication).

### 4.1.1. The Matrix Language Principle
According to this principle, it is always possible in classic code-switching to identify the Matrix Language in a bilingual Complementiser Phrase (CP), that is, a clause containing material from more than one language. The Matrix Language will be that of the morphosyntactic frame of the clause.

### 4.1.2. The Asymmetry Principle
This principle arises from the claim that ''bilingual speech is characterized by asymmetry in terms of the participation of the languages concerned'' (Myers-Scotton, 2002:9). We shall interpret this as suggesting that the Matrix Language will be unambiguously identifiable in most clauses. This principle is assumed to be different from the Matrix Language Principle in that while the latter allows for ambiguous identification in individual clauses, the Asymmetry Principle predicts that the majority of clauses in a set of data will be unambiguously identified.

### 4.1.3. The Uniform Structure Principle
According to the Uniform Structure Principle, ''the structures of the Matrix Language are always preferred'' (Myers-Scotton, 2002:8–9). This principle is based on the assumption that languages strive towards uniformity, and can be tested empirically by focusing on aspects of structure which may in the model be drawn from either the Matrix or the Embedded Language, but where a preference for Matrix Language elements is shown. This will be made more specific in section 5.4 below.

## 5. Application of the MLF model to the Welsh-English data

We shall now test the application of the three principles outlined above to Welsh-English data. This will involve exemplification and illustration of the principles in greater detail, followed by the results of an analysis relating to each principle. In general, we shall show that the principles do appear to be reflected in our data, and that they do indeed appear to represent a case of ''classic code-switching''. Nevertheless, we shall show that there are some problems of definition and some problematic examples, which will be discussed.

### 5.1. The data

Five hours of conversational data from a total of 30 bilingual Welsh-English speakers involved in 11 conversations were collected by bilinguals. Most conversations were recorded in north-west Wales, although one involved a telephone conversation between interlocutors from south Wales. All participants involved were adults, ranging in age from about 20 to middle age, and both men and women were included. In the analyses which follow, examples are drawn from all conversations. Where we present quantitative analyses, however, these are based on just one of the conversations.

### 5.2. Testing the Matrix Language Principle on the Welsh-English data

If this principle holds (see section 4.1.1), it will be possible to identify a Matrix Language in all bilingual clauses in the data. Two specific criteria will be used to identify the Matrix Language in each bilingual clause: (i) morpheme order; and (ii) the source of a particular type of system morpheme. These criteria follow from two additional principles to be outlined below. We shall not consider the Matrix Language Principle to be refuted if the Matrix Language is ambiguous in a few cases, for example, if morpheme order is compatible with both languages and there are no relevant system morphemes in a clause.

However, if morpheme order is partially Welsh-specific and partially English-specific, or if the relevant system morphemes come from both languages, then the Matrix Language Principle could be refuted.

### 5.2.1. The morpheme-order criterion
This follows from what Myers-Scotton calls the ''Morpheme Order Principle'':

In Matrix Language + Embedded Language constituents consisting of singly occurring Embedded Language lexemes and any number of Matrix Language morphemes, surface morpheme order. . . .will be that of the Matrix Language. (Myers-Scotton, 1993:83; 2002:59)

The morpheme-order criterion will be applicable wherever the two languages involved in code-switching have distinct surface orders. This is true of English and Welsh, since English is an SVO language while Welsh is VSO. The VSO order of Welsh is shown in example (10):

(10)     Agorodd     y     dyn y     drws.     (King, 1993:21)
         open-3S.PAST DET man DET door
         'The man opened the door.'

In addition, Welsh has a AuxSVO order in an alternative periphrastic construction which is used in informal speech, as in (11):

(11)     Mi     wnaeth     y     dyn agor y     drws.
         PRT do-3S.PAST DET man open DET door
         'The man opened the door.'

In this example, a form of the verb *gwneud* 'do' is used as an auxiliary in construction with a non-finite form of the verb *agor* 'open'. A similar contrast to that found between (10) and (11) is illustrated in the present tense versions in (12) and (13), where *bod* 'be' (and a particle *yn*) is used instead of *gwneud* 'do' as the auxiliary:

(12)     Agora          'r     dyn y     drws.
         open-3S.PRES DET man DET door
         'The man opens the door.'

(13)     Mae'r               dyn yn agor y     drws.
         be-3S.PRES- DET man PRT open DET door
         'The man opens the door.'

What (9)–(13) all have in common is that the finite verb appears before the subject instead of afterwards as it would in English.

Another way in which Welsh and English differ is in the relative order of head and modifier within an NP. Normal order in Welsh is head followed by modifier rather than modifier followed by head as in English. This can be illustrated by the Welsh NP *bag te* 'tea bag', in which the loanwords *bag* and *te* are placed in the Welsh order, head followed by modifier. The same applies to most adjectival modifiers, as is illustrated by the name of the charity *Y Groes Goch*, 'the Red Cross', where the head noun *croes*[3] is followed by the modifier. This example also illustrates that the definite article precedes the NP as in English. Thus, in a bilingual clause where the verb precedes the subject and where modifiers follow heads in NPs, we would identify Welsh as the matrix language according to the criterion of morpheme order. An utterance with subject-verb and modifier-head order, however, would indicate English as the matrix language according to this criterion.

We can illustrate how the morpheme-order criterion would apply to utterance (4), repeated here as (14):

(14)     mae     o-ᵗn          reit     **camouflaged** yn dydi               [MEW50]
         be.3S.PRES PRON.3S-PRT quite camouflaged  PRT NEG-be.3S.PRES
         'he's quite camouflaged isn't he?'

---

[3] The initial consonant of *croes* is voiced following the definite article because of its feminine gender.

In this utterance the verb *mae* 'is' appears clause-initially, reflecting Welsh VSO order, and we would thus identify this clause as following Welsh order. A similar conclusion would be reached in the case of example (15):

(15) oedd       gynnon ni        ystafell yn Plas yn Dref, ystafell **brilliant**    [MEG186]
     be.3S.IMP with-us PRON.3S room     in Plas yn Dref  room      brilliant
     'We had a room in Plas yn Dref, a brilliant room'

Here we have not only VSO order in Welsh, but also head-modifier order in the NP *ystafell **brilliant***. It would not be possible to identify the matrix language unambiguously if we were to find both Welsh VSO order and a modifier-head order that is deviant in Welsh. This happens very rarely, but we do have one example in our data. In example (16), English modifier-head is followed, but the rest of the word order is as in Welsh:

(16) ddaru ni gyfweld 'pnawn   'ma am ddeg awr **assistant** i DS       [MER542]
     did    we interview afternoon this for ten  hour assistant  to DS
     'we interviewed this afternoon for a ten-hour assistant to DS'

In this example the modifier NP *ddeg awr* 'ten hours' precedes its head *assistant* as in the English translation 'ten-hour assistant'. However, in the rest of the clause the word order is as in Welsh, with the auxiliary verb *ddaru* appearing at the beginning of the clause. The overall order of this clause thus cannot be classified as either English or Welsh, and the English order in the NP probably reflects a processing problem as a result of the need for on-line translation of a phrase taken from the English-speaking workplace, *ten-hour assistant.*
   An English type modifier-head order is found more regularly in utterances where English expletives are inserted as modifiers of Welsh heads, as in (17):

(17) mae'n           waeth na'r    **bloody** lle               [B110BRA]
     be.3S.PRES-PRT worse than-DET bloody  place
     'it's worse than the bloody place'

The position of English expletive adjectives contrasts with that of their Welsh equivalents as shown in (18):

(18) dan           ni 'di  cael tywydd **uffernol**, do?        [B10ALP]
     be.1PL.PRES we PRT get  weather hellish    tag
     'we've had hellish weather, haven't we?'

At first sight, the pre-head position of English expletives could pose a problem in identifying the matrix language if other aspects of word order are Welsh rather than English. However, we may note that a few adjectives occur pre-nominally in Welsh (cf.

Thorne, 1993:135) some of these arguably having emotive connotations.[4] If we were to consider English expletives to be used analogously to members of this class of Welsh adjectives then there would be no problem in identifying the word order in examples like (17) as Welsh. We shall adopt this approach in our analysis.

Another type of clause which may initially appear problematic for identifying morpheme order is where an English NP in English modifier-head order is inserted in an otherwise Welsh utterance with otherwise Welsh word order. This is illustrated in example (19):

(19)     achos  fod gen ti    dy   **silk handkerchief** yn dy   boced.      [MEW41]
         because be  to PRON.2s your silk handkerchief   in your pocket
         'because you have your silk handkerchief in your pocket'

The NP *silk handkerchief* follows the modifier-head order normal in English whereas the matrix language of this clause (see section 5.3.1 for identification of the matrix language in Welsh-English data) is Welsh. However, Myers-Scotton argues that this kind of example is not a problem for the MLF since the material around it follows the Matrix Language Principle. The NP *silk handkerchief* would be identified as an ''Embedded Language Island'' (EL). EL islands are defined as ''full constituents consisting only of Embedded Language morphemes occurring in a bilingual CP that is otherwise framed by the Matrix Language'' (Myers-Scotton, 2002:139). Their internal order is not taken account of by the morpheme-order criterion, which applies only to mixed-language constituents. EL islands are thought to reflect a heightened degree of activation of the EL in relation to the matrix language, which is assumed to be the most activated language in general. However, this activation may not be all that high if the EL islands are formulaic or composite expressions which might be listed in the mental lexicon of the EL, or if they are 'internal' EL islands in the sense of being part of a larger matrix-language constituent (Myers-Scotton, 2002:149). *Silk handkerchief* would presumably be classed as an internal EL island since it is part of the Welsh DP [*dy **silk handkerchief***] 'your silk handkerchief' where the determiner is Welsh. Myers-Scotton suggests that the frequent production of EL islands requires considerable proficiency in the embedded language (Finlayson et al., 1998), although there is some evidence to suggest that near equal proficiency in the two languages may lead to reduced frequency in the production of EL islands and more switching between clauses instead (Myers-Scotton, 2002:149; Backus, 1996:334).

In general, the word order of a clause is the same as that of the language which contributes the greater number of morphemes to the clause. However, this is not always the case. Example (20) is a bilingual clause where all the morphemes but one are from Welsh, but where the order is more compatible with English than Welsh:

---

[4] In fact, some Welsh adjectives always occur before the noun, e.g. *hen* 'old', whereas others vary in their position according to their meaning. One could consider English expletives either to belong temporarily to the *hen* class, or to occur before the noun when they have expletive rather than descriptive meaning.

(20)    fi     'di bod     i'r    **bus** lle                              [MSM10]
        PRON.1S PRT be.NONFIN to-DET bus place
        'I've been to the bus place.'

In this clause *bus* is the only word in English,[5] but it occurs in an NP *bus lle* 'bus place' where the head noun *lle* 'place' is followed by the modifier *bus*, following the normal modifier-head order found in English. However, in Welsh the normal order is head-modifier, so that we would expect *lle bus* in order to identify the matrix language as Welsh according to morpheme order. Instead, though, the order not only of the noun phrase but also of the rest of the clause is similar to English surface order, with a one-to-one correspondence between the Welsh morphemes occurring and possible English counterparts, as shown in (21):

(21)    fi 'di   bod i'r    **bus** lle[6]
        I  have been to-the bus  place

This would point to the matrix language being identified as English on the basis of the first criterion, morpheme order, although the second criterion has not yet been applied.

   In some bilingual clauses, the word order cannot be said to be uniquely Welsh or English because it is compatible with both. This applies, for example, to clauses where the initial element is focused in a way compatible with both languages, as in (22):

(22)    ni  oedd    y  **media stars**                              [MEM97]
        we be.3S.IMP DET media stars
        'WE were the media stars.'

In this example the word order is neither specifically Welsh nor English.

   Word order also cannot be identified as uniquely Welsh or English in some verb-initial questions like (23) (many questions in colloquial Welsh are distinguished from their affirmative counterparts only by intonation):

(23)    oedd    hi ddim yn **Social Services**?                     [MEG515]
        be.3S.IMP she not   in Social Services
        'Was she not in Social Services?'

As a final example of non-unique order, some utterances are incomplete clauses where the morpheme order is compatible with both Welsh and English, as in (24):

---

[5] Although the Welsh word for 'bus', *bws*, is somewhat similar, the vowel [ʌ] identifies it as English, since [ə] or [ʊ] would be expected in the Welsh word. However, the conclusions drawn about the matrix language would be the same even if the word had been classified as Welsh.

[6] I am told that this utterance sounds odd to native speakers, and it may be a 'performance error', particularly since a short pause is heard in the recording after *bus*. However, finding more examples of this kind could indicate convergence from Welsh towards English.

(24)     yn y gegin ar y **draining board**          [MEG160]
         in the kitchen on DET draining board
         'in the kitchen on the draining board'

From a morpheme order point of view the matrix language could be either English, with the insertion of the Welsh morphemes *yn y gegin ar y*, or Welsh, with insertion of *draining board* as an embedded-language island following English order.[7]

### 5.2.2. The system morpheme agreement criterion

This second criterion for identifying the Matrix Language of a bilingual clause follows from Myers-Scotton's "System Morpheme Principle": "In Matrix Language + Embedded Language constituents, all system morphemes which have grammatical relations external to their head constituent ... will come from the Matrix Language" (Myers-Scotton, 1993:83; 2002:59). This will involve identifying the relevant system agreement morphemes in a bilingual clause and noting which language they come from. Myers-Scotton distinguishes system from content morphemes on the grounds that content morphemes are involved in assigning or receiving thematic roles,[8] whereas system morphemes are not. Content morphemes are similar to Joshi's "open class" category and include nouns, verbs and adjectives. Examples of these in Welsh would be the noun *drws* 'door' and the verb *agor* 'open' in (10), and the adjective *uffernol* 'hellish' in (18).

System morphemes are then further subdivided into early versus late system morphemes, according to whether or not they are "conceptually activated" or directly linked to the speaker's intentions. The terms *early* and *late* reflect assumptions about how early or late in the production process the relevant morphemes are accessed. Early system morphemes have in common with content morphemes the fact that they are conceptually activated, but differ in that they do not receive or assign thematic roles. Examples of early system morphemes are determiners, e.g. *the* in English, *y* 'the' in Welsh, which add information about definiteness, and plural morphemes, e.g. *-s* in English, *-au* in Welsh, which add information about number. According to Myers-Scotton (2002:75). "early system morphemes are closely tied to their heads; they depend on the specific semantic/ pragmatic properties of a content head and add semantic/pragmatic information to that head".

Late system morphemes are then divided into "bridge" and "outside" late system morphemes, the latter being "coindexed with forms outside the head of their maximal projection" (Myers-Scotton, 2002:75) while the former[9] are not. In other words, outside late morphemes are the system morphemes mentioned above which have grammatical

---

[7] The inclusion of the system morphemes *y* 'the' make it more likely to be Welsh as a result of the Uniform Structure Principle: see section 4.1.3, but this kind of system morpheme cannot be used as a criterion for the matrix language.

[8] David Willis points out that thematic roles are not received by individual morphemes, but by phrasal categories such as DP, VP.

[9] Examples of bridge late system morphemes are prepositions like *of* in English or *de* in French when preceding a complement to a head noun as in *friend of John* or *beaucoup de gens* 'lot of people': cf. Welsh *o* in *llawer o bobl* 'a lot of people'). However, we shall not be concerned further in this paper with this type of morpheme.

relations external to their head. Examples are subject-verb agreement, clitics and case affixes.

For our purposes we shall re-define late system morphemes in terms of those involved in agreement processes rather than in terms of relations outside a morpheme's maximal projection. This is because the notion of maximal projection tends to be theory-specific (cf. Fukui, 2001), and also because in the case of a verb-first language like Welsh, current analyses would consider subject-verb agreement to take place within the same maximal projection,[10] i.e. within CP or FinP (see Roberts, 2004).

We shall select subject-verb agreement as the main means of identifying outside late morphemes in Welsh and English. This is because they are probably the most frequent kind of ''outside late morpheme'' that one can find in both languages, since they occur in most clauses which have finite verbs. Overt subject-verb agreement in English is of course mostly limited to the verb *to be* and third person singular *–s* in the present tense of regular verbs. However, in Welsh synthetic verbs have a rich system of inflection for person and number (see, for instance, King, 1993:180) in agreement with optional pronouns. Also, in the periphrastic construction with *bod* 'to be', the various finite forms of the verb (see King, 1993:145) mark person and number in agreement with a following personal pronoun. Thus, the language source of the morphology of finite verbs in bilingual clauses should enable us to identify the matrix language, and this criterion should lead to the same result for each clause as the morpheme-order criterion discussed in section 5.2.1.

To take an example, the matrix language of (14), repeated here as (25), was identified as Welsh in section 5.2 according to the order criterion:

(25)   mae      o-ⁱn       reit  **camouflaged** yn dydi              [MEW50]
       be.3s.pres pron.3s-prt quite camouflaged   prt neg-be.3S.pres
       'he's quite camouflaged isn't he?'

We can see that it is also Welsh according to the system morpheme agreement criterion, since the verb (with third person singular subject marking) and subject pronoun (third person masculine singular) also come from Welsh.

We saw in section 5.2 that it was sometimes not possible to apply the morpheme-order criterion where the word order of a clause was compatible with both Welsh and English. Just as there were some bilingual clauses to which the morpheme-order criterion could not apply, there are also some to which the system morpheme agreement criterion does not apply, in particular clauses not containing finite verbs, such as (19), repeated here as (26):

(26)   achos fod gen ti   dy  **silk handkerchief** yn dy  boced.        [MEW41]
       because be  to  pron.2s your silk handkerchief in  your pocket
       'because you have your silk handkerchief in your pocket'

In this case, the matrix language has to be identified on the basis of the morpheme order alone, which we identified above as Welsh.

---

[10] I am grateful to David Willis for pointing this out.

In section 5.2.1. above, we came across a problematic example (16), repeated here as (27), in which the morpheme-order criterion led to a conflicting result:

(27)  ddaru ni gyfweld 'pnawn 'ma am ddeg awr **assistant** i DS     [MER542]
      did   we interview afternoon this for ten   hour assistant to DS
      'we interviewed this afternoon for a ten hour assistant to DS'

Unfortunately the system morpheme agreement criterion cannot help us here since there is no finite verb: the auxiliary *ddaru* does not inflect for person.

In some cases neither criterion will apply, and it will not be possible to identify either Welsh or English unequivocally as the matrix language. This will happen in the case of incomplete clauses like (16) above, where morpheme order is compatible with both Welsh and English and where there are no subject-verb agreement morphemes.

We saw in section 5.2.1 above that the morpheme-order criterion led to the unexpected identification of English as the matrix language in example (18), repeated here as (28):

(28)  fi    'di bod    i-'r   **bus** lle                      [MSM10]
      PRON.1S PRT be.NONFIN to-DET bus place
      'I've been to the bus place'

What happens if we apply the second criterion, the system morpheme agreement criterion? In fact, we cannot apply it on the basis of subject-verb agreement, since there is no finite verb. However, we could consider *fi* to be a clitic indicating agreement with a missing auxiliary (especially in southern Welsh, of which this is an example[11]). The same could be said of the second-person clitic occurring without a finite verb in examples (29) and (30):

(29)  ti'n      gwybod, jest **draining board** a    um **filing cabinet**     [MEG160]
      PRON.3S-PRT know     just draining  board  and um filing cabinet
      'you know, just a draining board and a filing cabinet'

(30)  o ti'n          **gorgeous**                            [MEPB64]
      oh PRON.3S-PRT gorgeous
      'Oh you're gorgeous'

In examples like these we shall consider clitics to be indicative of subject-verb agreement.

To summarise, subject-verb agreement morphemes (plus clitics in the absence of finite verbs) are those which will help us identify the matrix language of a bilingual clause applying the second criterion. If the Matrix Language Principle is correct, then all system agreement morphemes in the same CP will come from the same source language; in addition, this source language will be the same as that identified by the first criterion, that of morpheme order (see section 5.2.1).

---

[11] One of the sets of data included a telephone conversation between two speakers who were originally from south Wales.

### 5.2.3. *A quantitative analysis*

In this section we describe a pilot quantitative analysis conducted in order to test the Matrix Language Principle. The analysis applies to a 40 minute conversation at a social gathering involving seven professional women. The transcription (made by one of the participants) was used to identify all bilingual clauses, i.e. those containing material from more than one language. English words listed in the Welsh dictionary which appeared in otherwise Welsh clauses were considered Welsh as outlined in section 2.2. Note that not all of the examples mentioned up to now were used in the quantitative analysis, but only those from one conversation, identified by reference numbers containing the letters 'ME'.

The two criteria for identifying the matrix language were then applied to each bilingual clause in order to test the Matrix Language Principle, according to which it is always possible in classic code-switching to identify the matrix language in a bilingual clause. The application of the first criterion, morpheme order, led to four possible outcomes:

(1) the label ''Welsh'' if the order was specifically Welsh;
(2) the label ''English'' if it was specifically English;
(3) ''either'' if the order was compatible with that of both languages;
(4) ''neither'' if it was compatible with neither language.

An example of a clause leading to each of these outcomes is given in (31)–(34):

(31)     oedden   nhw     mor **desperate though**                     [MEP130]
         be.3P.PAST PRON.3P so    desperate  though
         'They were so desperate though.'

(32)     Beauty is in the eye of the beholder, **ngwasi**                [MEP29]
         Beauty is in the eye of the beholder  dear
         'Beauty is in the eye of the beholder, my dear.'

(33)     yn y  gegin  ar y  **draining board**                          (= 30)
         in the kitchen on DET draining  board
         'in the kitchen on the draining board'

(34)     ddaru ni gyfweld 'pnawn  'ma am ddeg awr **assistant** i  DS    (= 22)
         did   we interview afternoon this for ten   hour assistant to DS
         'We interviewed this afternoon for a ten-hour assistant to DS.'

In (31), the order is identified as Welsh since the verb is clause-initial; in (32), the order is English: this is one of the few examples containing mostly English morphemes and following English SVO order. (33) is compatible with both English and Welsh surface order. (34) follows neither English nor Welsh order exclusively: it is verb-initial, indicating Welsh order, but the modifier *ddeg awr* precedes the head.

So far, only examples like (34) will be a problem for the Matrix Language Principle, because the impossibility of identifying exclusively English or Welsh order will prevent us from identifying a matrix language for each clause. (31) and (32) will not be a problem if

Table 1
Identification of Matrix Language in four examples

| Example number | Morpheme order | Source of agreement morphemes | Matrix language |
|---|---|---|---|
| (31) | Welsh | Welsh | Welsh |
| (32) | English | English | English |
| (33) | either | None | either |
| (34) | neither | None | neither |

application of the second criterion leads to identification of the same language; in the case of (33), we shall see that it will not be possible to apply the second criterion, and so the matrix language will be ambiguous between English and Welsh.

Having applied the first criterion to each bilingual clause, the second criterion was then applied. If the clause contained morphological marking of subject-verb agreement or a subject clitic without a verb as in (30) in section 5.2.2, this was identified as either Welsh or English. If there was no such marking, the outcome 'none' was recorded. Applying the second criterion to examples (31)–(34) above, we can see that subject-verb agreement in (31) (*oedden nhw*) is in Welsh; in (32) it is in English (*beauty is*); in (33) we cannot apply the criterion because there is no verb and so the clause is incomplete; and in (34) subject-verb agreement is absent because the auxiliary *ddaru* does not inflect for person (cf. King, 1993:165).

We can now demonstrate how the matrix language would be identified, where possible, in each of these four examples. This is done by determining whether the application of the two criteria leads to a compatible outcome, as indicated in the last column of Table 1.

We can see from Table 1 that where the morpheme order and agreement morpheme criteria identify the same language, as in (31) and (32), the matrix language is unambiguous. However, in (33), it is ambiguous because only the first criterion applies, leading to ambiguous outcome. In (34), one of the criteria leads to a conflicting outcome indicating that neither language is the matrix language, and so the matrix language is unidentifiable.[12]

Examples like (31) and (32) will be compatible with the Matrix Language Principle, whereas examples like (34) will not. Examples like (33) will not be directly in conflict with the principle since either language could be the matrix language, but the extent of this type of example will be relevant for testing a further principle, the Asymmetry Principle (see section 5.3 below). Our aim in the analysis which follows is to determine the relative frequency of clauses which are compatible with the Matrix Language Principle versus those, like (34), which are not.

The results of the analysis showed that there were a total of 163 bilingual clauses in the conversation, which consisted of 533 turns. The length of each turn ranged from one clause to 28 clauses but most turns were no more than one or two clauses long. Table 2 shows how many of the bilingual clauses turned out to have Welsh as their matrix language, how many English, and for how many clauses the matrix language could be either or neither language.

---

[12] A conflicting outcome would also result if each criterion indicated a different matrix language, but this does not in fact occur in the data.

Table 2
Matrix language in sample of bilingual clauses

| Matrix language | Number of bilingual clauses | Percent of bilingual clauses |
| --- | --- | --- |
| Welsh | 141 | 86.50 |
| English | 4 | 2.45 |
| either | 17 | 10.43 |
| neither | 1 | 0.61 |
| Total | 163 | 100 |

Table 2 shows that the overwhelming majority of the bilingual clauses analysed had a clearly identifiable matrix language, whether Welsh or English (86.5% + 2.45% = 88.96%). The matrix language in almost all of these clauses was Welsh. For 10.43% of the bilingual clauses analysed, however, the matrix language could be either language, and it was unidentifiable in one clause (0.61%).

All of the 18 clauses with ambiguous matrix language had in common that there was no finite verb or subject clitic to mark subject-verb agreement,[13] and the word order was compatible with both Welsh and English, as in (33).

The clause with neither language as matrix language has already been discussed as example (34) above. The problem in (34) was not the agreement morpheme criterion, which was not applicable, but the morpheme-order criterion. The modifier-head order within the noun phrase *ddeg awr **assistant*** was English rather than Welsh, while other aspects of the order of the clause were Welsh. As we can see from Table 2, this example accounted for less than 1% of the data analysed, and so can be considered rare. As suggested in section 5.2.1 above, the non-Welsh word order in the noun phrase may be attributable to on-line processing problems.

## 5.3. Testing the Asymmetry Principle on the Welsh-English data

As outlined in section 4.1.2, this principle predicts an asymmetry in the participation of the two languages in bilingual speech. We interpret this to predict not only that it will be possible to identify a matrix language in all clauses in the data (cf. section 5.2) but that, in addition, this language will be uniquely identifiable in the majority of clauses. This means that clauses where the matrix language could be either Welsh or English, while not directly contradicting the Matrix Language Principle, would contradict the Asymmetry Principle if they occurred in the data in large numbers.

To test this principle, we can make use of the analysis reported in the previous section and summarised in Table 2. If the Asymmetry Principle is to be upheld, then we would expect the vast majority of bilingual clauses to have either Welsh or English as unequivocally the matrix language. As we saw in the previous section, 86.5% of bilingual clauses had Welsh as the matrix language, and 2.45% English, making a total of 88.96% which had an unequivocal matrix language. In only 10.43% of the clauses was the matrix language compatible with both Welsh and English, and hence not unequivocally

---

[13] Myers-Scotton (personal communication) suggests that some of these clauses could be considered reduced CPs of the kind that commonly occur in informal conversation.

Table 3
Matrix language in sample of bilingual clauses excluding those containing switches of only one word

| Matrix language | Number of bilingual clauses containing switches of more than one word | Percent of bilingual clauses |
|---|---|---|
| Welsh | 73 | 92.40 |
| English | 3 | 3.8 |
| either | 3 | 3.8 |
| neither | 0 | 0 |
| Total | 79 | 100 |

identifiable. While it can thus be said that one-tenth of the data does not appear to follow the Asymmetry Principle, this is only a small amount, and arises only in incomplete or elliptical clauses where there is not enough information to assign a matrix language.

One might raise the objection that the data are likely to support the Asymmetry Principle anyway, given that some of the bilingual clauses analysed include single-word switches. In order to respond to this potential objection, we conducted an additional analysis on the data in which we excluded all bilingual clauses containing a switch of only one word. The results are shown in Table 3.

Table 3 shows that, in the reduced set of data, 92.4% of bilingual clauses had Welsh as the matrix language, and 3.8% English, making a total of 96.2% which had an unequivocal matrix language. In only 3.8% of the clauses was the matrix language compatible with both Welsh and English, and hence not unequivocally identifiable. Thus, in the data excluding single word switches we see that we have even stronger support for the Asymmetry Principle than in the data which included single word switches. This suggests that, if we have misclassified any loans as switches in the earlier analysis as discussed in section 2.2, this has not led to a bias in favour of the MLF model.

Overall, then, the data do support the Asymmetry Principle as defined by Myers-Scotton (2002:9). They also support a quantitative notion of asymmetry, in that Welsh is the matrix language in almost all bilingual clauses where this can be unambiguously identified. In our data this dominance of Welsh is also reflected in a morpheme count of a sub-sample, which showed that Welsh morphemes make up over 90% of all morphemes in the conversation.

### 5.4. Testing the Uniform Structure Principle on the Welsh-English data

As outlined in section 4.1.3, this principle predicts a preference for the structures of the matrix language. According to the Matrix Language Principle, as we have seen in section 5.2, outside late or agreement system morphemes can only come from the matrix language of a clause. However, the Uniform Structure Principle goes beyond this in predicting that other system morphemes, e.g. early system morphemes, will be drawn preferentially from the matrix language. This does not mean that early system morphemes from the embedded language are prohibited, but only that they are less likely than early system morphemes from the matrix language. This kind of prediction lends itself to a quantitative analysis, which we performed on the same sample of the data as that reported in section 5.2.

For the purposes of testing this prediction on the data, we decided to focus on a particular early system morpheme found in both Welsh and English, the definite determiner, *the* in

English, and *y/yr/'r* in Welsh (King, 1993:30). Only the definite determiner was selected, as indefinite determiners equivalent to English *a/an* do not occur in Welsh.

If the Uniform Structure Principle is correct, we would expect definite determiners in bilingual clauses to be drawn from the matrix language rather than the embedded language. If any definite determiners from the embedded language do appear, we would expect them to be part of an embedded-language island, that is, followed by an English rather than a Welsh NP. In general, though, we would expect a preference for English NPs to be inserted in Welsh utterances following Welsh rather than English determiners. In example (22), repeated as (35), this preference is reflected by the fact that the English NP is preceded by a Welsh determiner *y* 'the':

(35)    ni oedd      y   media stars                                    [MEM97]
        we be.3s.IMP DET media stars
        'WE were the media stars.'

The English determiner *the* is not prohibited from occurring instead, and ***the media stars*** could have been produced instead of *y **media stars***, but the latter type of occurrence is predicted to be more frequent.

In the analysis all examples of definite determiners in bilingual clauses were isolated, and their source language identified. This was then compared with the matrix language of the clause. In addition, the determiners were further classified in terms of whether they were followed by an English or a Welsh noun phrase head. This can be illustrated by example (36):

(36)    doedd      hi'm        yn mynd ar y  **motorways** na'r  **dual carriageways**
        be.PAST.NEG PRO.F.3S-NEG PRT go    on DET motorways nor-DET dual carriageways.
        'She didn't go on the motorways or the dual carriageways' [MER84]

In this clause there are two occurrences of definite determiners, *y* and *'r*, both from Welsh. The matrix language is also Welsh on the grounds of both morpheme order and agreement morphemes. In addition, the noun phrases following the determiners in both cases are in English. In (32), repeated as (37), the definite determiner is English, followed by an English NP, and the matrix language is also English:

(37)    beauty is in the eye of the beholder, **ngwasi**.                [MEP29]
        beauty is in the eye of the beholder  dear
        'Beauty is in the eye of the beholder, my dear.'

A contrasting example is provided by (38), where the determiner and following noun phrase are drawn from English, but the matrix language is Welsh:

(38)    dw i'n      poeni diawl o'm byd  am  un byd  ond am   the list,
        be1s PRES–PRT worry devil of-NEG world about one world but about the list
        list number one              [MEM241]
        list number one
        'I'm not worried about anything at all except the list, list number one.'

Table 4
Distribution of definite determiners in relation to the matrix language

| Definite determiners | In clauses with Welsh matrix language | | | In clauses with English matrix language | | | All |
|---|---|---|---|---|---|---|---|
| | +Welsh NP | +English NP | +All NPs | +Welsh NP | +English NP | +All NPs | |
| Welsh | 16 | 41 | 57 | 0 | 0 | 0 | 57 |
| English | 0 | 2 | 2 | 0 | 2 | 2 | 4 |
| All | 16 | 43 | 59 | 0 | 2 | 2 | 61 |

The results are summarised in Table 4.

There were a total of 61 definite determiners available for analysis, 57 Welsh[14] and 4 English. As Table 3 shows, all the Welsh definite determiners occurred in clauses where Welsh was the matrix language, 15 of them preceding Welsh NPs and 41 of them preceding English NPs. There were only four English definite determiners, two of them occurring in clauses where English was the matrix language, and two where it was not. This means that, of the 61 determiners analysed, only two did not match the matrix language of the clause. One of these occurred as part of an embedded-language island as shown in (39):

(39)     **in the second column** rhoi tic                       [MER228]
        in the second column put tick
        'In the second column put a tick'

The other appeared in example (38), given above. I understand from native speakers that examples like (38) are fairly rare. Table 3 shows that out of 43 English NPs appearing in clauses with Welsh as the matrix language, only two were preceded by an English determiner. This supports the prediction that system morphemes will be drawn preferentially from the matrix language.

Table 3 also allows us to observe that, in the data analysed, there were no occurrences of determiners from the embedded language other than in embedded-language islands. That is, in clauses with Welsh as the matrix language, English determiners only preceded English NPs, never Welsh NPs. This suggests that the occurrence of embedded-language determiners outside embedded-language islands could have been used as an additional criterion for the identification of the matrix language in examples like (24). This would have led to the identification of Welsh as the matrix language in four of the 17 clauses listed in Table 2 as having "either" English or Welsh as matrix language.

## 5.5. Summary and conclusion

In this section we have considered the application of three principles of the MLF model to the Welsh-English data and have found that all three are upheld in a quantitative analysis. This case, then, does seem to be a "classic" case of code-switching. In the next two

---

[14] An additional four Welsh determiners had to be excluded from analysis because they occurred in clauses where the matrix language was either ambiguous or unidentifiable.

sections we shall compare our conclusions to those on other language pairs, and consider the implications of our analysis for the MLF model.


## 6. Welsh-English compared with other studies in the MLF framework

### 6.1. Other "classic" cases

Myers-Scotton (1993) developed her model mostly with reference to her Swahili-English code-switching data collected in Kenya. The identification of the matrix language in these data can be illustrated by example (40) (Myers-Scotton, 2002:57, example (2)):

(40)     u-na-wez-a          ku-m-pat-a        a-me-va-a        nguo      ny-ingine
         2S-NONPST-able-FV   INF-OBJ-find-FV   3S-PERF-wear-FV  clothes   CL9-other
         **bright** kama **color**   y-a            **red**
         bright as       color   CL9-ASSOC red
         'You can find her (she is) wearing other bright clothes [such] as red [ones]'.

In this example, the morpheme order follows that of Swahili, where the head noun *nguo* 'clothes' is followed by its modifiers rather than being preceded by them as is normal in English. In addition, the affixes marking case and subject-verb agreement (outside late morphemes) can all be identified as coming from Swahili. Hence the matrix language of this CP is unambiguously Swahili. In 40 conversations, Myers-Scotton (1993:236) reports finding only one example where the modifier-head order unexpectedly reflects English in an otherwise Swahili utterance where the outside late morphemes apparently come from Swahili. In this case, she argues that there are pragmatic reasons for this violation. As in our study, then the Matrix Language Principle was upheld in the overwhelming majority of utterances analysed.

The Matrix Language Principle is also upheld in the study by Finlayson et al. (1998) on code-switching involving English and various Bantu languages in South Africa. They find that it is always possible to determine the matrix language in a given CP and that there are no counter-examples to this generalisation. They also found that speakers typically used the same matrix language throughout the clauses in their turn, and, as mentioned in section 5.2.1, that there was a link between proficiency in the languages involved in code-switching and the type of bilingual clauses produced.

Among other studies reported by Myers-Scotton (2002) to have applied the MLF model successfully to their data are those by Hlavac (2000) and Türker (2000) on Croatian-English and Turkish-Norwegian, respectively. Backus (1996) is a study of Turkish-Dutch code-switching which, although it does not set out explicitly to test the MLF model, makes considerable use of its conceptual framework in accounting for the data.

Boussofara-Omar (1999, 2003) reports applying the MLF model successfully to account for code-switching between a variety of standard Arabic and local Tunisian Arabic: she reports that Tunisian Arabic is the matrix language into which elements of the standard are inserted (Boussofara-Omar, 2003:33). However, she reports finding some problematic data for the model. Not all of these turn out to be problematic for the MLF if

only late outside morphemes are considered (cf. Myers-Scotton, 2004). However, the implications of some remaining problematic data will be considered in section 6.2 below.

The Uniform Structure Principle, found to be supported in our analysis, is also the focus in Jake et al. (2002) analysis of NPs containing English elements in Spanish-English code-switching data where Spanish is the matrix language. Where Spanish would require a determiner preceding these NPs, the majority of determiners appear in Spanish rather than English. So examples like *el **career fair*** 'the career fair' are much more common than English NPs with English determiners as in ***the first one***. Jake, Myers-Scotton and Gross found that 151 English NPs occurred with Spanish determiners, as compared with 21 occurring with English determiners. This proportion (12.2%) was somewhat higher than in our analysis, where only 4.65% (2/43) English NPs were preceded by English determiners in clauses with Welsh as the matrix language. However, it is difficult to compare the two directly since Jake et al. included indefinite determiners in their analysis whereas we did not. Nevertheless, both studies indicate support for the Uniform Structure Principle.

## 6.2. Code-switching and convergence

Some of the data reported by Boussofara-Omar (2003) as apparently problematic for the MLF model are not actually bilingual clauses, but clauses with all the morphemes coming from one language. An example of this is her (7.1), part of which is reproduced below as (41).

(41)      li?anna  laa  ya-nhad      al-?adab
             because  NEG  3MS.IMP-advance  DET-literature
             'because literary production does not progress...'

She reports that all the morphemes in (41) come from standard Arabic, whereas the word order is that of Tunisian Arabic ['because' + VP + NP] and not the expected standard Arabic order ['because' + NP + VP].

Since (41) is not a bilingual clause (i.e. with morphemes drawn from more than one language) the predictions of the MLF model do not apply. However, it could be considered to exemplify a kind of "covert code-switching" (cf. Schmitt, 2000:9) or convergence, in that the structure of the clause shows convergence towards Tunisian Arabic.

A third type of code-switching, which is distinct from both overt "classic" code-switching and covert code-switching is what Myers-Scotton (2002:105) calls "composite code-switching", "a phenomenon with morphemes from two languages within a bilingual CP, and with the abstract morphosyntactic frame derived from more than one source language". This is sometimes found, for example, in the code-switching of children whose speech is undergoing a process of "Matrix Language turnover" (cf. Myers-Scotton, 2002:247) as a result of moving from a country where one of their languages is mainly spoken to a new country where their other language is spoken in the community. For example, Myers-Scotton (2002:166) cites Spanish-English data from a child who has moved from Colombia to the USA and who produces an utterance, shown as

(42), which has Spanish subject-verb agreement but English modifier-head order in the object NP:

(42)    yo    quiero    manzana jugo
          1s.NOM want.1s.PRES apple    juice
          'I want apple juice.'

Like covert code-switching, composite code-switching involves convergence in the use of one language towards the other.

Another example of composite code-switching is provided by Bolonyai (1998:34) from her study of a Hungarian-English bilingual child aged 3:7; this is reproduced as (43):

(43)    játsz-ok               **school**-*ot*
          play-1S.PRES.SUBJ.CONJ school-ACC
          'I'm playing school.'

This poses a problem for the MLF model in that whereas –*ok* is an outside late morpheme drawn from Hungarian, the word order follows English, with the object *school* after the verb, "violating the pragmatic rules of Hungarian that would require the object NP to be in Topic position" (Bolonyai, 1998:34). So the word order and morphological criteria (see section 2.2) do not identify either English or Hungarian unambiguously as the matrix language. Bolonyai proposes that, as a result of the influence of English on the child's Hungarian after she had begun attending an English-medium nursery school, there is a gradual shift in progress from Hungarian to English as the preferred matrix language. Evidence of this gradual shift is an 'in-between' or composite matrix language, which combines aspects of both English and Hungarian.

Since language contact often leads to convergence, it seems reasonable to ask whether there are indications of convergence in the Welsh-English data. We could look for this in (a) bilingual clauses which are problematic for the MLF; and (b) monolingual clauses where all the morphemes are from one language but the word order from another. Examples (16) and (20), repeated here as (44) and (45), are examples of the first type.

(44)    ddaru ni gyfweld 'pnawn   'ma am *ddeg awr **assistant*** i DS    [MER542]
          did    we interview afternoon this for ten   hour assistant to DS
          'We interviewed this afternoon for a ten hour assistant to DS.'

(45)    fi    'di bod     i'r    **bus** lle                    [MSM10]
          PRON.1s PRT be.NONFIN to-DET bus place
          'I've been to the bus place.'

Both contain NPs (underlined) showing English modifier-head order whereas other aspects of word order in the clauses are Welsh. An additional example (46) is found in written Welsh on the internet:[15]

---

[15] Found by David Willis, http://www.fydd.org/curiad/grwpiau/eraill/a.html, accessed on 31 October 2004.

(46)    Fe  wnaeth    y   <u>byd   enwog  Super Furry Animals</u> chwarae ei [sic]
           PRT do-3S.PAST DET world famous Super Furry Animals play       its
           gig cyntaf. . .
           gig first
           'The world-famous Super Furry Animals played their first gig. . .'

Here the underlined NP includes the pre-modifier *byd enwog* 'world famous', itself a calque of the English expression, retaining the English modifier-head order, and placed before the head *Super Furry Animals* in English-style modifier-head order. Just as the order in the NP *ddeg awr **assistant*** in (45) may be explicable in terms of on-line translation, so (46) may also demonstrate the effects of translation from an originally English text, and translation is a possible mechanism in processes of convergence. This type of example was very infrequent in our data, hence we did not consider it a major threat to the MLF. However, if such examples were to become more common, or to be much more frequent in another set of data, we might wish to argue that there are signs of convergence of Welsh towards English.

The second kind of evidence for convergence, in monolingual clauses, is beyond the scope of this study but might well be an avenue for future research.

## 7. Implications of this study for the MLF model

As we saw in section 5, our analyses showed that our data were overwhelmingly compatible with the Matrix Language Principle, the Asymmetry Principle and the Uniform Structure Principle. This suggests that Welsh-English code-switching is indeed a "classic" case. We can derive additional confidence in this conclusion from Myers-Scotton's (2002:111–112) discussion of classic code-switching in terms of two extralinguistic factors, proficiency and stability. She suggests that it is found in situations of stable bilingualism, which we suggested above (section 3) is true of Wales. A second characteristic is that it occurs "between speakers who have sufficient proficiency in one of the participating languages to use it as the sole source of the morphosyntactic frame of bilingual CPs" (Myers-Scotton, 2002:110). This is clearly the case for the native speakers of Welsh in our recordings. They have all been exposed to both Welsh and English at an early age and are fluent in both.

Nevertheless, our study was conducted on only a small amount of data, and it would be desirable to extend it to a larger corpus. This would also allow us to check whether the very low frequency of problematic data is replicated in other datasets. If not, or if future data to be collected should not be compatible with the MLF, it may be possible to view it as evidence of language convergence in line with the suggestions made in section 6.2 above.

In our data we have thus shown that the MLF model has support from an empirical point of view. From a theoretical point of view, however, it should perhaps be a source of concern that the quadri-partite '4M' model of morphology it adopts is somewhat idiosyncratic, bearing little relation to current theories (cf. Spencer and Zwicky, 2001). We saw in section 5.2.2 that the definition of "outside late morphemes" was somewhat problematic from a theoretical point of view, and we redefined this category loosely in terms of agreement

morphemes. However, further work on the theoretical aspects of the MLF model might consider whether Myers-Scotton's quadri-partite model should be replaced by one which assumes a more standard bipartite distinction between lexical and functional categories (cf. Fukui, 2001:394). Perhaps the matrix language could be identified in terms of functional categories including subject-verb agreement, pronominal clitics, and determiners. As we saw in section 5.4, the use of determiners as a criterion would have made a difference to the results, in that it would have allowed us to identify the matrix language unambiguously in an additional four clauses in the data analysed. Further research on a range of datasets from various languages would be necessary to determine whether this would be feasible, but if so it would enhance the parallelism between the MLF and current morphosyntactic theory.

## 8. Summary and conclusion

In this paper we have tested the MLF model of code-switching with Welsh-English data and have concluded that Welsh-English can indeed be considered a classic case of code-switching, in that a matrix language can be clearly identified in bilingual clauses. This has been established by a qualitative and quantitative analysis, finding support for three principles associated with the MLF model: the Matrix Language Principle, the Asymmetry Principle and the Uniform Structure Principle. Further research on a larger corpus would allow us to pay more attention to any problematic examples which might be indicative of convergence in progress.

## References

Aitchison, J., Carter, H., 1994. A Geography of the Welsh Language 1961–1991. University of Wales Press, Cardiff.

Backus, A., 1996. Two in One: Bilingual Speech of Turkish Immigrants in The Netherlands. Tilburg University Press, Tilburg, The Netherlands.

Bolonyai, A., 1998. In-between languages: language shift/maintenance in childhood bilingualism. International Journal of Bilingualism 2 (1), 21–43.

Boussofara-Omar, N., 1999. Arabic diglossic switching: an application of Myers-Scotton's MLF model, Ph.D. Thesis. University of Texas, Austin, unpublished.

Boussofara-Omar, N., 2003. Revisiting Arabic diglossic switching in light of the MLF model and its sub-models: the 4-M model and the abstract level model. Bilingualism, Language and Cognition 6 (1), 33–46.

Clyne, M., 2003. Dynamics of Language Contact. Cambridge University Press, Cambridge.

Finlayson, R., Calteaux, K., Myers-Scotton, C., 1998. Orderly mixing and accommodation in South African codeswitching. Journal of Sociolinguistics 2, 395–420.

Fukui, N., 2001. Phrase Structure, In: Baltin, M., Collins, C. (Eds.), The Handbook of Contemporary Syntactic Theory. Blackwell, Oxford, pp. 374–406.

Fynes-Clinton, O.H., 1913. The Welsh Vocabulary of the Bangor District. Humphrey Milford, Oxford University Press, London.

Gross, S., 2000. When two becomes one: creating a composite grammar in creole formation. International Journal of Bilingualism 4 (1), 59–80.

Hlavac, J., 2000. Croatian in Melbourne: lexicon, switching and morphosyntactic features in the speech of second-generation bilinguals, PhD. Thesis. Monash University, Melbourne, unpublished.

Jake, J., Myers-Scotton, C., Gross, S., 2002. Making a minimalist approach to codeswitching work: adding the matrix language. Bilingualism, Language and Cognition 5 (1), 69–91.

Jones, R.O., 1993. The Sociolinguistics of Welsh, In: Ball, M., Fife, J. (Eds.), The Celtic Languages. Routledge, London, New York, pp. 536–605.

Jones, K., 1995. Code-switching, intertextuality and hegemony: exploring change in bilingual discourse. In: Proceedings of the Summer School, Code-Switching and Language Contact, Fryske Akademy, Ljouwert, Leeuwarden, pp. 108–118.

Jones, K., 2000. Texts, Mediation and Social Relations in a Bureaucratised World, In: Martin-Jones, M., Jones, K. (Eds.), Multilingual Literacies: Reading and Writing Different Worlds. John Benjamins, Amsterdam, pp. 209–228.

Joshi, A., 1985. Processing of Sentences with Intrasentential Switching, In: Dowty, D.R., Kartunnen, L., Zwicky, A. (Eds.), Natural Language Parsing. Cambridge University Press, Cambridge, pp. 190–205.

King, G., 1993. Modern Welsh. Routledge, London.

Lindsay, C., 1993. Welsh and English in the city of Bangor: a study in functional differentiation. Language in Society 22 (1), 1–17.

Muysken, P., 2000. Bilingual Speech: A Typology of Code-Mixing. Cambridge University Press, Cambridge.

Myers-Scotton, C., 1993. Duelling Languages: Grammatical Structure in Code-Switching. Clarendon Press, Oxford.

Myers-Scotton, C., 2002. Contact Linguistics: Bilingual Encounters and Grammatical Outcomes. Oxford University Press, Oxford.

Myers-Scotton, C., 2004. Research note and erratum. Bilingualism, Language and Cognition 7 (1), 89–90.

Myers-Scotton, C., Jake, J., 1995. Matching lemmas in a bilingual language production model: evidence from intrasentential codeswitching. Linguistics 33, 981–1024.

Myers-Scotton, C., Jake, J., 2000. Four types of morpheme: evidence from aphasia, codeswitching, and second language acquisition. Linguistics 38 (6), 1053–1100.

Parry-Williams, T.H., 1923. The English Element in Welsh: A Study of English Loan-Words in Welsh. Honourable Society of Cymmrodorion, London.

Poplack, S., Meechan, M., 1998. How languages fit together in codemixing. International Journal of Bilingualism 2, 127–138.

Poplack, S., Sankoff, D., 1984. Borrowing: the synchrony of integration. Linguistics 22, 99–135.

Roberts, I., 2004. Principles and Parameters in a VSO language: A Case Study in Welsh. Oxford University Press, Oxford.

Schmitt, E., 2000. Overt and covert codeswitching in immigrant children from Russia. International Journal of Bilingualism 4, 9–28.

Spencer, A., Zwicky, A., 2001. The Handbook of Morphology. Blackwell, Oxford.

Sweet, H., 1884. Spoken North Welsh. In: Wyld 1913. pp. 499–574.

Thomas, C., 1982. Registers in Welsh. International Journal of the Sociology of Language 35, 87–115.

Thorne, D.A., 1993. A Comprehensive Welsh Grammar. Blackwell, Oxford.

Türker, E., 2000. Turkish-Norwegian codeswitching: evidence for intermediate and second generation Turkish immigrants in Norway, Ph.D. Thesis. University of Oslo, Oslo, unpublished.

Williams, C. (Ed.), 2000. Language Revitalization: Policy and Planning in Wales. University of Wales Press, Cardiff.

Wyld, H.C. (Ed.), 1913. Collected Papers of Henry Sweet. Clarendon Press, Oxford.