Pure

Bangor University

DOCTOR OF PHILOSOPHY

Identifying word-order convergence in the speech of Welsh-English bilinguals

Webb-Davies, Peredur

Award date:
2010

Awarding institution:
Bangor University

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 19. Jan. 2025

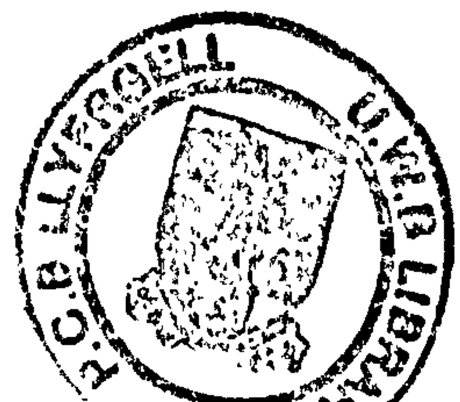
Identifying word-order convergence in the speech of Welsh-English bilinguals

Peredur Glyn Cwyfan DAVIES

Thesis submitted for the degree of Doctor of Philosophy (Ph.D.),
The School of Linguistics and English Language,
Bangor University 2010

Traethawd a gyflwynwyd ar gyfer gradd Doethur mewn Athroniaeth (Ph.D.),
Ysgol Ieithyddiaeth ac Iaith Saesneg,
Prifysgol Bangor 2010

Supervised by / Goruchwyliwyd gan: Prof. Margaret Deuchar



Abstract

This thesis presents a study of the speech of Welsh-English bilinguals to determine the extent and manner of the structural influence of English on Welsh, specifically the phenomenon of convergence, which is described as the increase in frequency of use of a construction (e.g. word order) in one language due to the prevalence of that construction in another language with which its speakers are in contact.

I take two approaches to measure convergence, using Welsh-English conversational data which were specially-collected for a 40 hour corpus. First, I adapt the Matrix Language Frame model (Myers-Scotton 2002), usable to identify the language from which clause morphosyntax is sourced, to identify convergence. I propose the concept of a dichotomous Matrix Language, which is where there is conflicting evidence for which language provides clause structure. In testing the model on speech from six speakers, I find that, with few exceptions, Welsh is the source of the structure in the majority of clauses analysed. I interpret this to show that word-order convergence in these data is limited insofar as using the Matrix Language Frame model indicates.

Second, I analyse the speech of 28 bilinguals for evidence of the deletion of the initial auxiliary verb in periphrastic constructions involving an auxiliary form of *bod* 'be' and a 2nd person singular pronominal subject *ti*. Auxiliary deletion (AD) in such clauses results in a clause-initial subject, which I compare to English SVO word-order. I find that AD in such contexts is very common in these data, and is also found in clauses with a different subject. Analysis of age variation in the data indicates that AD in Welsh has become more common in recent years. I propose that an increase to subject-initial clauses in Welsh may be a change in progress, which I interpret to be in part due to convergence to English.

Table of contents

List of tables and figures

Acknowledgments

First and foremost I thank Prof. Margaret Deuchar for her supervision, which has been frequent, prompt, unreluctant, perspicacious and unfailingly beneficial, and for her assistance and encouragement, particularly with regards to academic publication and presentation.

Second, I thank my parents, Gareth and Eleri, profusely for their constant support and, ultimately, for a room in the loft to finish the writing-up.

I also thank my colleagues at Bangor University, particularly at the School of Linguistics and English Language and at the ESRC Centre for Research on Bilingualism. During the course of writing this thesis individuals from these departments have provided of their time and knowledge to help me with the study, and thus deserve my thanks and respect. Specific thanks must go to Prof. Eddie Williams for being my internal examiner and for providing valuable comments, Dr. Dirk Bury for being my second supervisor, Dr. Pam Macdonald, Prof. Ineke Mennen, Dr. Panos Athanasopoulos, and Dr. Enlli Thomas.

Others now follow who have provided support in ways I can never hope to fully repay, so my gratitude will have to do for the time being. Diolch o galon i chi i gyd.

At Bangor: Dr. Maria del Carmen Parafita Couto, Dr. Jonathan Stammers, Dr. Tom Ristimaki, Elen Robert, Siân Wynn Lloyd-Williams, Dr. Jon Herring, Dr. Diana Carter, Alberto Rosignoli, Lexi Hindley, Kathryn Morgan, Marika Fusser, Dr. Fredrik Karlsson, Dr. Kara T. McAlister, Prof. Peredur Lynch, Prof. Hywel Wyn Owen, Gruff Prys, Dr. Sara Parry, Dr. Gwenith Hughes, Tegau Andrews and Hunydd Andrews, and of course to Dr. Owain Llwyd for buying the beer.

Worldwide: Dr. David Willis for examining me and providing insightful comments, Prof. Carmen Silva-Corvalán, Dr. Tina Hickey, Dr. Ad Backus and Dr. Patricia Shaw.

Not forgetting: Kelly Webb, Seiriol Cwyfan Davies, Matthew Hilton, Bryn Jones, Llio Evans, Beth Fox, Dr. Ben "Hunter" Snook, Matthias Ammon, Dr. Alaric Hall, Dr. Oliver Padel, Prof. Andy Orchard and Dr. Richard Dance. And everyone in Starfish.

List of abbreviations used in this thesis

1S: 1st person singular pronoun

1PL: 1st person plural pronoun

2S: 2nd person singular pronoun

2PL: 2nd person plural pronoun

3Sm: 3rd person singular masculine pronoun

3Sf: 3rd person singular feminine pronoun

3PL: 3rd person plural pronoun

-A: deleted auxiliary verb

+A: retained auxiliary verb

ACC: accusative case affix

AD: auxiliary deletion

B2S: Welsh auxiliary construction where the auxiliary verb is the 2nd person singular form of *bod* (*wyt*)

CS: code-switching

COMP: complementizer morpheme

CONDIT: conditional tense

DET: determiner

EL: embedded language

ibid.: *ibidem*, indicating the same reference as previously stated

IMP: imperfect tense

IMPER: imperative mood

L1: first language

L2: second language

L_A: Language A

L_B: Language B

ML: matrix language

MLF: the Matrix Language Frame model

MOP: Morpheme Order Principle

MP: the Minimalist Program

NEG: negative particle/affix

NOM: nominative case affix

-P: deleted aspectual particle

+P: retained aspectual particle

NONFIN: non-finite verb

PAST: past (perfect) tense

PERF: perfective affix

PREP: prepositional affix

PRES: present tense

PRT: a pre-verbal particle

PRT.PAST: perfective aspectual particle

Q.NEG: negative interrogative morpheme

Sm: singular masculine affix

SMP: System Morpheme Principle

Spec: specifier morpheme

VBZ: verbalizer suffix

viz.: *videlicet*, indicating a more detailed specification of a point

Chapter 1 — Introduction: An examination of the future of the Welsh language

In his famous radio address *Tynged yr Iaith* ("The fate of the language"), delivered in Welsh on the Welsh region of the BBC Home Service on the 13th February 1962, Saunders Lewis foresaw the future of the Welsh language as bleak.

> Mi ragdybiaf hefyd y bydd terfyn ar y Gymraeg yn Iaith fyw, ond parhau'r tueddiad presennol, tua dechrau'r unfed ganrif ar hugain, a rhoi bod dynion ar gael yn Ynys Prydain y pryd hynny.[1]

He also noted the words of schools commissioner Matthew Arnold, writing in 1852 in response to the recent reports published on the state of education in Wales,[2] that "sooner or later, the difference of language between Wales and England will probably be effaced". Referring obliquely to the likelihood that the 1961 census would reveal yet another decline in the proportion of the people of Wales who could speak Welsh,[3] Lewis was warning his listeners, and the nation in

---

[1] "I predict also that Welsh as a living language will cease to be, assuming present trends continue, by around the beginning of the twenty-first century, given that there are still people in the British Isles at that time."

[2] The so-called 'Blue Books', i.e. the *Reports of the commissioners of enquiry into the state of education in Wales* (1847); for a discussion see e.g. Davies (1994).

[3] It did. Though writing in 1962, Lewis would not have had access to the 1961 census results at that time, which would have informed him that the percentage of people in

general, that a continuation of what he saw as the flattery ("gweniaith") of English-medium governance in Wales by Welsh speakers, particularly those of the middle classes, would result in the hasty death of Welsh.

Lewis, naturally, writes in rhetorical terms and with a distinct political agenda. Though he writes of the potential death of Welsh, he foresees such an event as the result of socio-political processes and of, particularly, the usage of English instead of Welsh in the public and private sectors, rather than, as it might be seen from a linguist's perspective, as the result of the grammatical dominance of English over Welsh in terms of structure, lexicon and phonology, which could eventually lead to language death (see e.g. Thomason and Kaufman 1988:50, Myers-Scotton 1998). Presumably Lewis' warnings were effective, and Arnold's prediction that Wales and England would be linguistically homogenised has been falsified, since Welsh is still a living language at the beginning of the 21st century. The 2001 census results for Wales indicate that 20.8% of the population (582,368 people) over the age of 3 can speak Welsh,[4] which, whilst a lower number and proportion than 1961, represents the first increase in at least two centuries of the proportion of the population that can speak Welsh.[5]

Another notable trend over the last century has been the increase in bilingualism among Welsh speakers in Wales. The 1901 census indicates there were 280,900 monolingual Welsh speakers (Jones

---

Wales who could speak Welsh had gone down, from 28.9% to 26.0% (656,002 people), the continuation of a decline in the overall number of Welsh speakers that had begun in the 19th century (the original published census data is available at http://www.byig-wlb.org.uk/Cymraeg/cyhoeddiadau/Cyhoeddiadau/629.pdf or http://tinyurl.com/walescensus1961; see Jones 1993 for a discussion).

[4] This is the proportion/number of people who can speak Welsh. The number of people who have at least one skill (speaking, reading, writing, understanding) in Welsh is higher (797,717 people or 28.4%); see http://tinyurl.com/censuscomparison.

[5] It is an increase in 2.1% (74,270 people) since the previous census of 1991. See the Welsh Language Board documentation at http://tinyurl.com/censuscomparison.

1993:548), representing 14% of the population of 2,012,876.[6] This proportion had lessened to 1% by 1961 and effectively to zero by 2001.[7] Conversely the number of speakers who are bilingual in Welsh and English has increased over that century. The proportion of Welsh speakers who could also speak English was shown to be 69.8% in 1901; this had increased to 95.8% by 1981 (Jones 1993), and by 1991 it was assumed in the census that all Welsh speakers could also speak English (see e.g. Deuchar 2005b).

What these figures suggest in relation to Lewis' predictions are that, whilst Welsh has indeed declined (in terms of the number of speakers) since the beginning of the 20th century, and furthermore English has become the *de facto* 'other' language which Welsh speakers can speak, the spread of English has not caused the death of Welsh. In fact, the increase in the number of Welsh speakers in the last decade has been taken by some as an optimistic indicator of the firm foothold of Welsh (cf. Deuchar and Davies 2009). However, it is probable that the increase in Welsh-English bilingualism over the past century has had some result on Welsh which is not reflected in the statistics of the censuses. Lewis may have been proven wrong (a fact he would doubtlessly approve of) from a geolinguistic/anthropolinguistic perspective, but from a morphosyntactic point of view, perhaps he may yet be proven right. What if English has had a more fundamental effect on the Welsh language, specifically in its grammatical structure, which might result in Welsh becoming so much 'like English' that eventually

---

[6] My calculation using the statistics at http://www.statistics.gov.uk/census2001/bicentenary/pdfs/wales.pdf.

[7] Note that the census only counts members of the populace who were over the age of 3 at the time the census data were collected. It is probable that there were (and are) a number of very young monolingual Welsh children.

the two languages will, as Matthew Arnold might have it, become indistinguishable?

In this thesis I will respond to this question by undertaking a linguistic study into the extent of the structural effect of English on Welsh, specifically with regards to word order convergence. Convergence, as I will discuss in detail in chapter 2, is defined as the process whereby two languages in contact become more similar to one another in certain respects. From the point of view of considering the potential linguistic death of Welsh, and considering two extremes along a continuum, widespread convergence towards English across all aspects of Welsh syntax could be viewed as a precursor of imminent language death, whilst widespread lack of convergence could be seen as an indicator of English having had no effect on the grammatical structure of Welsh, thereby suggesting that language death is not a plausible scenario. I will report in this thesis as to where I believe Welsh currently sits along this putative continuum.

My research will examine spontaneous conversational data collected as part of a project based at Bangor University, north Wales, funded by the Arts and Humanities Research Council and resulting in the creation of a substantial corpus of recorded and transcribed speech of Welsh-English bilinguals of various backgrounds. By studying the informal speech of Welsh speakers I hope to be able to comment, both on the situation of living Welsh as it is spoken in 2009, and also on the possible direction in which the morphosyntactic structure of Welsh will take in the future.

The next chapter will be a review of the literature in the field of bilingualism and language contact. I will focus on language contact phenomena such as code-switching and contact-induced language change, particularly convergence. I will comment on the present

research on convergence in the field of bilingualism, highlight what research has been undertaken in this field with regards to Welsh, and note where I believe the gap in this research currently lies, and which I hope to fill.

The third chapter will be a continuation of the literature review, focusing on the concept of a matrix language as the morphosyntactic frame for bilingual speech. In particular I will discuss the work of Myers-Scotton (1993, 2002, etc.) and her concept of the Matrix Language Frame model. I will explain how I propose to use this model to analyse a dataset taken from the Welsh-English corpus, whereby the language identified as providing the source of the morphosyntactic structure of a clause produced by a bilingual is called the matrix language, and by identifying to what extent Welsh is the matrix language of clauses produced by speakers in the dataset, it can be shown how dominant English is in the grammar of those speakers. I also describe my innovation to the Matrix Language Frame model which sets out to identify clauses produced by bilinguals which show word-order convergence, viz. a dichotomous matrix language, where information in a clause indicate that both a bilingual's languages are providing morphosyntactic structure to that clause, and can be taken as an indicator that there are signs of word-order convergence in such a clause.

In the fourth chapter I describe the methodology used to collect the data for this thesis (i.e. for the larger corpus from which I extract data to analyse). I describe how participants were found and selected, how recordings of them in conversation were made and using what equipment. I also describe the transcription system used for transcribing the recordings, namely the CHAT system used for LIDES (Language

Interaction D*ata* Exchange System), the on-line bilingual corpus network centred around the Talkbank website (http://talkbank.org).

In chapter five I show the results of the application of the MLF model to a dataset of transcribed data from six speakers. I discuss the matrix language distribution in the data, distinguishing between the distribution in finite, nonfinite and verbless clauses, and also in monolingual and bilingual clauses, before discussing whether it is Welsh or English which is the predominant source of morphosyntax for clauses produced by the speakers analysed. I then examine the subset of clauses in the data which has a dichotomous matrix language to see whether word-order convergence is found. I then discuss the extent to which this indicates that word-order convergence towards English is present in the speech of these bilinguals. As a comparison I then analyse a small dataset from a corpus collected by Eppler (2004) from Austrian German-English bilinguals, and compare the proportion of dichotomous matrix language clauses found in those data to the proportion found in the Welsh-English data.

In the sixth chapter I continue my study of word-order convergence in Welsh, but now move beyond the Matrix Language Frame model. I analyse clauses which would not be labelled as having a dichotomous ML according to that analysis, but which seem to show word-order convergence. These are Welsh auxiliary constructions which have the 2nd person singular pronoun *ti* as the subject, and where the clause-initial finite auxiliary verb (*wyt*) is often deleted by speakers, resulting in the pronoun being the first surface element of the clause. I analyse a second dataset of 28 Welsh-English bilinguals from the same corpus and show what proportion of their output of these kinds of clauses show auxiliary deletion. I then consider to what extent this indicates a predominance of subject-initial clauses in the output of these

speakers, and whether this can be interpreted as a sign of convergence from verb-subject (VS) to subject-verb (SV) word order in Welsh through the influence of English main clause order. As a comparison, I discuss work by Shaw, Campbell and Grant (2008) on the Canadian Amerindian language Central Salish, wherein a similar phenomenon of auxiliary deletion is found, and which is argued to show a word-order shift in Salish towards SV through the influence of English word order.

In the final chapter I then return to the results presented in the thesis, and discuss what they contribute overall to our understanding of the extent of convergence in Welsh towards English. I will also make suggestions as to further research that could stem out of the work I present here, and describe where I foresee my own research focusing next.

Chapter 2 — Review of the literature on convergence and contact-induced language phenomena

1. *Introduction: Language contact and bilingualism*

This thesis is concerned with identifying the synchronic signs of language change, specifically externally-motivated change deriving from language contact in bilingual communities. In this chapter I will discuss the existing literature relating to this topic, focusing on the process of convergence, and I will also outline my own theoretical position on contact-induced language change.

When speakers of one language come into contact with speakers of another language, those languages can affect one another. The result is language change in one or both of those languages (Weinreich 1953:1, Winford 2003:23), where parts of a language's grammar are lost, added to, or otherwise modified or restructured. For example, Romansh lost gender agreement in predicate adjectives as a result of contact with German (see Thomason 2003:690), and Kupwar Urdu acquired an inclusive and exclusive 'we' distinction due to influence from Kupwar Marathi and Kupwar Kannada (Gumperz and Wilson 1971:160–1). A language's main word-order can also change due to language

contact. For example, Finnish has shifted from SOV to SVO under influence from English and other Indo-European languages (Thomason 2001:11), and Central Coast Salish has mostly shifted from VSO to SVO, also under influence of English (Shaw et al 2008).

Certain forms of contact-induced language change[8] only usually occur due to intense contact. Indeed, according to Thomason (2003:689) and Winford (2003:23), for contact-induced structural change to occur, intense language contact is a requirement.[9] Such a situation exists in bilingual communities where the majority of speakers are able to speak both the affected language and the affecting language. Bilingualism, therefore, which is a result of language contact, often leads to language change. I define bilingualism as the ability to speak and understand more than one language, though complete fluency in both those languages is not a requirement (cf. Thomason 2001:3, Weinreich 1954:1). In the situation explored in this thesis, speakers of Welsh in Wales are also almost always speakers of English, and I will be examining the extent of language change in Welsh as a result of this bilingualism.

Thomason and Kaufman (1988:74–6) note that fluent bilingualism is by no means a prerequisite for contact-induced language change to occur. Different types of change can occur in languages depending on the intensity and types of language contact in evidence. They propose a "borrowing scale" which illustrates the correlation between the intensity of language contact/bilingualism and the form of language change which occurs. I summarise it below in table 1.

---

[8] This is the term which I adopt as the overall blanket term for this phenomenon in this thesis. I also use sometimes use borrowing in a general and non-specific sense to refer to some manner in which one language changes under influence from another language.

[9] Nevertheless, contact is not necessarily a sufficient condition by itself: other factors, such as congruence between the two grammars, are usually considered a requirement for change to occur. I discuss these factors later in this chapter.

1.	**Casual contact.** Bilingualism in source language not essential. Borrowing of: nonbasic vocabulary, e.g. nouns; also some verbs, adjectives and adverbs.
2.	**Slightly more intense contact.** Some bilingualism in source language expected. Borrowing of: some function words; some structure, but nothing that would alter structure types already in the borrowing language.
3.	**More intense contact.** More bilingualism, and social factors that favour borrowing. Borrowing of: basic and nonbasic vocabulary; moderate structural borrowing, e.g. word order, some inflectional affixes.
4.	**Intense contact.** Very extensive bilingualism, and social factors that strongly favour borrowing. Borrowing of: vocabulary of all types; heavy structural borrowing of any kind, including that which affects the typology of the borrowing language.

Table 1. The "borrowing scale", correlating intensity of language contact with degree of contact-induced change; adapted from Thomason and Kaufman (1988:74–6).

The greater the intensity of language contact, the greater the likelihood of extensive language change, and the more numerous are the different types of change (Thomason 2001:66). In situations where bilingualism is rare and most speakers are monolingual, lexical change can be expected but not structural change, whereas in situations where bilingualism is the norm, lexical or structural change can be expected

(cf. Croft 2000:207). The third and fourth levels of table 1 above represent situations of moderate to extensive bilingualism among the speakers, and in such situations, moderate to extensive types of language change occur. Bilingualism, then, can lead to extensive change, particularly when the form of bilingualism in evidence is simultaneous acquisition (cf. Field 2005:347–8, who notes that "profound change" can be exhibited in such situations).

In a situation of extensive bilingualism, such as among the Welsh speakers in Wales, language contact effects of the fourth type in Thomason and Kaufman's scale—extensive borrowing of practically any type—might be expected. External aspects play a role in the extent and intensity of change, such as speakers' attitudes to borrowing (Thomason 2001:69), the duration of language contact and the relative social dominance of the languages in question (e.g. Thomason 2001:66, Treffers-Daller and Mougeon 2005, Rista-Dema 2008). In the case of Welsh in Wales, language contact with English has been extensive for centuries, and monolingual English speakers in Wales outnumber Welsh-English bilinguals by about four to one, as the 2001 census shows.[10] Both these facts suggest that structural influence of English on Welsh could be both limitless in its form and extensive in its scope. The research presented in this thesis will contribute to examining the extent of this influence.

Language change stemming from contact takes many forms. This study is concerned with convergence, a mechanism of language change which I define and discuss in the next section.

---

[10] The UK census for 2001 show that 20.8% of the population of Wales can speak Welsh. This was an increase from the 1991 census, which indicated that 18.7% of the people of Wales spoke Welsh. These figures are taken from the table found on the Welsh Language Board's website: (http://www.bvig-wlb.org.uk/english/publications/publications/332.doc) or (http://tinyurl.com/censuscomparison).

2. Convergence

In this thesis I consider convergence to be a language change mechanism, a result of which is that languages can become structurally more similar to one another. Through convergence, inherent structural similarities between two languages in contact are enhanced (Bullock and Toribio 2004:91), with the result that those languages attain greater similarity with regards to such constructions (Backus 2004:179). Other authors use different terminology for this process, e.g. it is called simplification by Silva-Corvalán (1994:3).

As noted by Backus (2005:333), in convergence a construction or pattern (e.g. word order) that is congruent in L_A and L_B is used in L_A <u>in preference to</u> a construction or pattern which is not congruent in L_A and L_B. A result of this process in an increase in the frequency of that pattern in L_A (under the influence of its distribution in L_B), and in the subsequent decrease in the frequency of the related L_A pattern which is not congruent in L_B. Crucially, the L_A pattern undergoing convergence is not innovative to L_A—rather, it is <u>already present</u> in L_A (see e.g. Bullock and Gerfen 2004:96)—but the frequency of the occurrence of that pattern increases in L_A to be more similar to the equivalent pattern's distribution in L_B (Backus ibid.).

Convergence can be seen as pattern redistribution in one or both languages in contact, so it is system-preserving, although if the incongruent pattern ceases to be used entirely (due to the predominance of the congruent pattern which is undergoing convergence), then that redistribution would result in category loss/addition, so convergence can be the catalyst to system alteration (even though it is in itself not a system-altering process).

Convergence is one of several forms of contact-induced processes of language change, alongside e.g. direct transfer and calquing. I discuss these other processes briefly in section 5 below and compare and contrast them with convergence.

As noted by Thomason (2001:93), constructions which are similar in two languages, but not identically distributed (e.g. constructions that are marked in one language but unmarked in the other) are most prone to convergence. This concept is related to that of congruence, discussed in section 9 below. This is not the direct transfer of a construction from Language A into Language B which was not in Language A before, but rather the increase in the frequency of use of an existing construction (cf. Thomason 2001:93), and what Toribio (2004:167) described as "the preferential use of some structures over other options". As noted by Thomason (2001:89), the result of the convergence process in a language is "a change in frequency ... of previously existing constructions".

I interpret these proposals to mean that, when convergence occurs, a bilingual will choose a construction from one language (L_A) which has a similar counterpart in the other language (L_B) in preference to an L_A construction which does not; furthermore, the construction chosen is more prominent in L_B than in L_A. In brief, a speaker who uses an L_A construction that would usually be more common in that context in L_B than in L_A is producing convergence: in this scenario, L_A is becoming 'more similar' to L_B.

The elements of a grammar most prone to convergence, as noted by Backus (2004:180) and Thomason (2001:93), and as I infer from Sanchez's (2004:148) revised Functional Convergence Hypothesis,[11] are

[11] Sanchez formulates the Functional Convergence Hypothesis, which states that if two languages spoken by a bilingual have common features associated with the same

those which are similar across both languages in contact but not identically distributed. In order to identify word-order convergence in a bilingual dataset, word order which is similar (congruent) in two languages but with different frequency of distribution must be identified. I address this in the next chapter.

Naturally, if the result of convergence is a change in a construction's frequency, it is not possible in a synchronic study to identify an increase in the frequency of anything (cf. again Backus 2004:180: "mere changes in frequency are probably a common type of change, but almost impossible to demonstrate"). Nevertheless, from the linguist's analytical viewpoint, convergence is, I suggest, synchronically-observable: this is achieved by trying to find the evidence of the results of convergence. The evidence of word-order convergence will be where morphemes from L_A which are expected also to have structure (i.e. word

---

functional category then other features associated with the same category will be prone to convergence (2004:150). Sanchez examines what she terms convergence in the speech of Quechua-Spanish bilinguals. In Quechua, past tense features are associated with evidentiality, while in Spanish, past tense features are associated with aspect. In this regard, Quechua and Spanish are dissimilar. However, both languages associate mirativity (the expression of surprise or of new information) with past tense features. In this regard the two languages are similar (2004:148). Sanchez finds that speakers in her data appear to associate Spanish imperfect tense morphology with evidentiality, indicating convergence to Quechua in this category.

However, I define convergence differently to Sanchez. Because both Quechua and Spanish share a feature (mirativity) which is associated with a certain feature matrix (tense), Sanchez proposes that Quechua-Spanish bilinguals use other categories (such as evidentiality or aspect) associated with this matrix from the other language. However, since Quechua does not apparently have an aspectual distinction, nor does Spanish have an evidential distinction, I interpret Sanchez's findings to show borrowing of a feature from one language to another one which lacked that feature, rather than convergence as I have defined it. True, the end result is that both languages have become more similar as a result of contact, but it is not features which are similar in both languages which converge: the similar feature above, the one found in both languages, i.e. mirativity, remains apparently unchanged by this process. I therefore disregard the Functional Convergence Hypothesis in the current study, but I suggest it is useful to see how apparently similar outcomes—two languages becoming more similar—can derive from different processes: borrowing in Sanchez's (2004) data, but convergence as per e.g. Toribio's (2004) and my definition.

order) from L_A will instead have structure that is more expected of L_B, but still available in L_A. In such an instance, the L_A-L_B bilingual has extended the use of the L_A word order with those L_A morphemes under the influence of that word order in L_B. Myers-Scotton (2002) refers to a similar phenomenon when she states that a clause in which word-order convergence has occurred will be one where the morphemes come from one language but the grammar includes structure from both participating languages (2002:164). What Schmitt (2000:15) calls covert code-switching (a clause "with all morphemes from one variety ... but some abstract lexical structure from another variety") seems to be similar, as does what Johanson (2002a) terms combinational copying, which is the use of a construction from language A (the "donor language") in language B but using morphemes from language B (2002a:15). Note that both Myers-Scotton's and Schmitt's formulations imply that the observable results of convergence are particular to clauses which are monolingual at the surface level, but it indicates the likelihood that the results of convergence will also be observable in clauses which contain code-switching (see section 6 below for a discussion of the literature on code-switching), i.e. bilingual clauses; cf. Schmitt (2000), which shows Russian-English children producing English verbs without the expected Russian morphological suffixes, and Myers-Scotton's (2002) concept of composite code-switching, which appears to be similar to her definition of convergence (see chapter 3 for further discussion of this).

Therefore, a more general definition of the evidence of word-order convergence is required: a clause which has morphemes from one or both languages and also structure from both languages. Note that 'structure from both languages' here entails the meaning 'structure <u>available</u> in both languages' (see section 3 for examples), but there is

the assumption that the structure used is not usually used for the morphemes, or some of the morphemes, occurring in a given construction. Instead, its use has been extended.[12] An important aspect of my definition above is that convergence can be observed in both monolingual and bilingual clauses.

Thomason and Kaufman (1988) draw a distinction between "borrowing" and "interference", where borrowing represents language change that is caused by synchronic factors such as code-switching, and interference is caused by influence of the L1 on the target language (the L2) during second language acquisition (cf. Winford 2005:130). Convergence as I view it falls into Thomason and Kaufman's "borrowing" category, in that it occurs synchronically, possibly alongside code-switching, rather than being a form of interference of L1 on L2 during acquisition. I thus do not discuss Thomason and Kaufman's concept of interference in this thesis; cf. also van Coetsem's (1988) concept of interference and imposition. I discuss these approaches in more detail in section 5.

Convergence, it should be noted, can affect either one or both participating languages, being either unidirectional or multidirectional depending on the situation. In the case of the former, Montrul (2004) argues that frequently it will be the (psycholinguistically-dominant) majority language that affects the (psycholinguistically-weaker) minority language (2004:126–7), indicating that it is a one-way process, though Thomason (2001) points out that with convergence there is no source language and receiver language (or at least that such misleading

---

[12] Note that, therefore, since the morpheme order exhibited in such clauses will already be available in both languages, it follows that convergence (as I define it) not a system-altering process, since no new structures are borrowed from one language into another. The convergence process is a system-preserving one, where existing constructions in the language are redistributed.

terminology—and concepts—should be eschewed); rather, she posits that structures that result from the convergence "have no single source" (2001:89).[13] So-called 'new' structures occurring as a result of convergence were already either present in the languages or "resemble both languages in part but do not match one another completely" (ibid). Thus, as I described above, convergence differs from direct transfer, where a feature of one language is adopted into another which did not have that feature before. Heine and Kuteva (2005) also propose convergence to be a process whereby the participating languages are "mutually influencing each other", thereby becoming increasingly alike (2005:11). Convergence can sometimes result in both languages converging towards one another: cf. Thomason's (2001:148) evidence of Turkish-German bilingual children using Turkish phrase-final intonation patterns in German, and vice versa, but with different semantic functions. Furthermore two languages in contact can both undergo convergence but in different parts of the grammar, so that Part X of the grammar of L_A becomes more similar to Part X of L_B, and Part Y of L_A grammar becomes more similar to Part Y of L_B, e.g. in Welsh English.[14]

Nevertheless, it is quite possible that convergence will result in change in only one of the participating languages. This is indeed what is found in most of the studies that I will discuss below. Thus, when the concept of, for example, convergence in Welsh towards English is

---

[13] Though Thomason (2001:262) suggests that the term 'convergence' "is not usually used to designate unidirectional changes, except (sometimes) when one language changes very extensively to become more like another," I disagree. I give examples in section 8 which fit my notion of convergence but which result in only one language changing. My position is that convergence can be either unidirectional or multidirectional, depending on the specific situation.

[14] I give a brief discussion of the effects of English on Welsh word order in section 4 of this chapter.

discussed herein, this signifies the influence of English on Welsh to make Welsh, broadly-speaking, more similar to English.

It should be stated that what have been called convergence areas or linguistic areas or Sprachbünde (cf. Thomason 2001:99) are not related to the present discussion. Sprachbund refers to a geographical location where the close proximity of speakers of different languages has caused those languages to become more similar to one another. Whilst it is quite possible that the convergence process is behind changes in recognised Sprachbünde, I am not concerned with this issue, specifically with regard to the situation in Wales (which I do not suggest is a Sprachbund[15]).

In the next section I discuss existing research into convergence which supports my view of it as a process of language change.

3. Previous research on convergence

To recap, I define convergence as the increase in the frequency of a structure already present in a language on the basis of contact with a language in which that structure is more prominent. It is a structure-preserving not structure-altering process. Though I interpret convergence to be a process which can result in two languages in contact becoming more similar over time, I do not call the result itself convergence. In previous research, some writers have treated convergence as a process, some as a result, and some as both. I suggest, however, that previous work examining convergence shows data which support the interpretation of convergence as a process

[15] Primarily because a Sprachbund is usually defined as having more than two languages in close contact (cf. Thomason 2001:99), and this is not true for most of Wales.

(whether or not the original authors view it as such). I present an overview of key recent studies in this section, to show how convergence can be interpreted as a process.

Work by Sandalo (1995, reported in Thomason 2001:89) shows convergence in the speech of Portuguese-Kadiwéu bilinguals in Brazil. Portuguese is primarily an SVO language, whereas Kadiwéu has a large number of different main clause word orders, including SOV, VSO, OVS, VOS, OSV and SVO. Given sentences to translate orally from one language into the other, the speakers showed an increase in the use of SVO order when they produced Kadiwéu translations of Portuguese sentences. The inference is that the predominance of the SVO order in Portuguese has influenced an increase in the frequency of their use of SVO in Kadiwéu, because SVO is available in Kadiwéu but is only one option. Here, then, a construction which is available in Kadiwéu but is more prominent in Portuguese has increased in frequency in Kadiwéu because of this influence. This fits the definition of convergence I gave above, where pre-existing constructions in a language are redistributed (like SVO word order in Kadiwéu), rather than new constructions being introduced.[16] The result of clauses with convergence being produced frequently in the speech of Portuguese-Kadiwéu bilinguals is that Kadiwéu has apparently become an SVO language because of contact with Portuguese (cf. Lewis 2009).

Toribio's (2004) analysis of Spanish-English bilinguals provides another example of convergence. She demonstrates that contact with English has affected these speakers' suppression (or not) of subject pronouns ('pro-drop'), a typical feature of the Spanish of monolinguals.

---

[16] Note also that this example fits with my suggestion that convergence can occur in monolingual (in terms of what language the morphemes come from) clauses as well as bilingual clauses (section 2 above, where I note that "convergence can be observed in both monolingual and bilingual clauses").

Participants were asked to narrate a fairy tale in both (i) a monolingual Spanish mode based on a series of coloured pictures depicting the story and (ii) a bilingual CS mode[17] subsequent to them having been made to read aloud a "mixed-language" form of a story and then, presumably, paraphrasing it from memory; Toribio is unclear on this last point (2004:168). An example of data from the bilingual mode task is given in (5) below, where I have underlined the overt pronouns which would be unexpected in a monolingual Spanish speaker's output.

(5) *They had to return to his meager work* y a una casa que era muy humilde. Ella no estaba *accustomed to that kind of work*. Así es que ella sintió que esto no, que ella no iba a poder resistirlo. Tenía que *wash clothes* a mano, *she had to cook*. Era una vida totalmente diferente a la que ella estaba *accustomed to*.

"*They had to return to his meagre work* and a house that was very humble. She was not *accustomed to that kind of work.* So she felt that that was not, that she could not withstand it. She had *to wash clothes* by hand, *she had to cook.* It was a totally different life from what she was *accustomed to.*"

(from Toribio 2004:171)

Toribio's participants used overt pronouns in monolingual mode more often than was discourse-pragmatically expected of monolingual Spanish discourse. In the bilingual mode the use of overt pronouns was even more frequent than expected. In (5), note that after the first occurrence of the pronoun *ella,* one would expect subsequent subjects with identical reference to be null in the speech of a monolingual Spanish speaker, but

[17] See section 7 for a discussion of language mode (Grosjean 1998 etc.).

in this passage the subjects are frequently overt.[18] Toribio states that a speaker being in the bilingual mode and making use of code-switching "further favors the searching for parallels" between Spanish and English (2004:172); that is, being in bilingual mode emphasises the similarities between languages for the speakers.

These data imply that these bilinguals are influenced by the bilingual frame of the bilingual mode and showed convergence to English when it came to selecting whether or not to delete the pronoun. The lack of pro-drop in English influenced them to extend this pronoun retention to Spanish morphemes. This is a process which makes the overall discourse appear to be more English-like in its lack of subject pro-drop.

Montrul (2004), in a study with a similar focus to Toribio's, compared the speech of Spanish-English heritage bilinguals[19] living in the USA with the speech of monolingually-raised Spanish speakers from different Spanish-speaking countries, examining, among other features, their use of null subjects. The heritage speakers in Montrul's data used overt subjects (68.6%) more than null subjects (31.4%), compared to the reverse pattern in monolinguals' use (42.8% overt and 57.2% null). As Spanish allows such subjects to be null whereas English does not, Montrul argues that convergence is occurring in the Spanish morphosyntax of these speakers and that the dominance of English has led the speakers to select overt subjects instead of null subjects. She

---

[18] Note that there is nevertheless a null subject in the sentence *Tenía que wash clothes a mano* 'She had to wash clothes by hand.'

[19] Montrul uses the term "heritage speakers" to describe second- or third-generation bilinguals who come from a home where a non-English language (Spanish here) was spoken, and who have some level of bilingualism in both English and the heritage language.

ascribes this to the long-term English contact the adult bilingual heritage speakers have been exposed to (2004:138).[20]

Schmitt (2000), analysing data from Russian-English bilingual immigrant children in the USA, proposes the concept of covert code-switching, which, as I noted in section 2 above, represents convergence in bilingual clauses. Schmitt defines covert CS as what occurs when the lexical structure of one language is influenced by the lexical structure of another language, resulting in a composite morphosyntactic frame (2000:19), and thereby morphemes from one language are found with structure, or some structure, from another language (2000:15). Schmitt says that an example of this is bare forms, being, in her terms, "[Embedded Language] forms that are used without the required system morphemes of the [Matrix Language]" (2000:23). These are produced by some of the children in her data; an example is in (6) below.

(6) nu yest' baseball v <u>heaven</u>?
well is baseball in heaven
"Well, is there baseball in heaven?
[Alternative form:
nu yest' baseball v <u>heaven</u>-e?
well is baseball in heaven-PREP.Sm]
(from Schmitt 2000:23)

In (6) the English word *heaven*, which is a code-switched insertion in a clause which is predominantly Russian, does not have the expected

[20] Note that "heritage" bilinguals are not the same (in terms of relative fluency etc.) as the balanced bilinguals analysed in Toribio (2004), and therefore are arguably not directly comparable. However, it is worth noting that in both Montrul's and Toribio's studies, bilingualism of either kind (heritage or balanced) results in a convergence, albeit to a different extent, thereby further emphasising the link between bilingualism and convergence.

Russian prepositional singular case ending: in inserting the switched item from English, the speaker has incorporated with it the abstract lexical structure of English, which lacks overt noun inflection for case, other than genitive. Schmitt suggests that this is a case of convergence having occurred, since the Matrix Language is a composite of some structure from Russian, e.g. the verb morphology, and some from English, e.g. no noun inflection on *heaven*. She proposes that the convergence here occurs due to an increase in the level of activation of the Embedded Language to the point where it competes with the ML in the projection of the clause's morphosyntactic structure (2000:24). I interpret Schmitt to view convergence as a process, with a clause which has composite structure being the result (cf. 2000:25).

What is usually referred to as negative borrowing, which results in a category being lost in one language because of the absence of an equivalent category in the contact language (Sasse 1992a:16, Dorian 2006:557), can be considered to be a form of convergence. An example given by Sasse (1992b:68) concerns the optative mood in early 20th century Arvanitika, a variety of Albanian spoken in Greece. This mood was lost as a result of contact with Modern Greek, which lacks the same category. Sasse notes that Arvanitika speakers, as a result, extended the use of the subjunctive (what Sasse calls "conjunctive") mood instead of the optative, on a pattern with Greek. Here again I view convergence to be the process where these speakers used, on a clause by clause basis, the subjunctive mood in Arvantika constructions on the basis of the prominence of the same mood in Greek, rather than the optative mood, which Greek lacks. The overall effect of numerous instances of this convergence is that the speakers used the optative mood less and less frequently. Here, category loss was the eventual result of frequent occurrences of this convergence.

Note that many of the studies cited in this section—e.g. Schmitt (2000), Toribio (2004), Sasse (1992b)—feature data which show widespread occurrences of the same type of convergence, which results in those speakers' language changing. Nevertheless, convergence does not have to lead to change. If a type of convergence is not found very frequently in a dataset, then that indicates only that change has not yet resulted, not that convergence has not taken place. Convergence can be a 'one-off' process. I will discuss this point further in chapter 5 when I present the results of my analysis.

These studies demonstrate that convergence can be interpreted as a process deriving from language contact, and not in itself a result. Next, I will look at the research that has been undertaken so far on word-order convergence in the speech of Welsh-English bilinguals, specifically on the influence of English on Welsh word order.[21]

Thomas (1982) argued that Welsh speakers were at a stage of transitional bilingualism when he was writing, favouring the English language over Welsh, and he noted the demographic decline of Welsh speakers during the second half of the 20th century (1982:209-10)—a decline which has now arguably been reversed (see Deuchar and Davies 2009). Thomas bases his argument on his observations of the speech of bilinguals from North Wales. He identifies some structural changes in Welsh which he ascribes to English influence, such as the use of Welsh verbs on a pattern with English modals (e.g. *gallu* 'to be able', *dylu* 'ought', *medru* 'to be able'; cf. English *can*, *ought* or *shall* [1982:213]), an increase in the use of Welsh periphrastic constructions over synthetic constructions (1982:214), phonetic simplification through loss of features which do not have a parallel in English (e.g. vowel alteration in

---

[21] A brief discussion of contact-induced change of Welsh structure into English follows at the end of this section.

feminine adjectives, such as *het wen* 'white hat' > *het gwyn* 'white hat' [1982:215]), and simplification of the Welsh mutation system (e.g. the production of *yn Bangor* 'in Bangor' not *ym Mangor* 'in Bangor', where nasal mutation is not applied[22] [1982:216]; data by Jones [1998:97–8] reinforces this claim, as her respondents aged under 40 applied this mutation in fewer than 50% of the contexts appropriate to historical Welsh). Thomas' view of CS, meanwhile, is that it is found in the speech of less accomplished Welsh speakers, who are viewed as low-status by those speakers with a more extensive stylistic Welsh repertoire (1982:218).[23] Thomas hypothesises a later stage which follows on from extensive CS in the speech of Welsh-English bilinguals, where such speakers have reduced competence in Welsh grammar and are considered "inadequate even in vernacular style" (ibid.; cf. Deuchar and Davies 2009:19). This can be interpreted to refer to a form of convergence, where extensive interference from English in the grammar of Welsh would result in both English and Welsh structure being used within the same bilingual clause.[24] Thomason and Kaufman (1988)

---

[22] The preposition *yn* 'in' causes the nasal mutation of an initial plosive on a following noun, so *yn + Bangor* would be *ym Mangor*.

[23] Naturally, Thomas's observations reflect the situation of Welsh language as perceived by many in the 1980s to be bleak; cf. Jones (1981), who used unsubtle terms to describe intrasentential CS by Welsh speakers as the "indiscriminate use of English words and phrases in Welsh utterances" found only in the speech of "speakers below the age of fifty with a low level of formal education in the language" and in situations or communities with a "low to moderate Welsh intensity" (1981:49). In fact, as more recent studies show (e.g. Deuchar 2006, Deuchar and Davies 2009), CS by Welsh-English bilinguals is not restricted to such speakers, but is found in the speech of speakers from a variety of educational and regional backgrounds. The data from the 'Siarad' corpus, which I use for the research presented in this thesis, certainly upholds this point.

[24] As Deuchar and Davies (2009) point out, this concept is similar to a later concept by Myers-Scotton (1998), viz. matrix language turnover, where a shift in a speech community from the predominant use of one language as ML to the predominant use of another language as ML may have an intervening stage where both languages are being accessed as clausal ML, resulting in speakers producing clauses with convergence and/or what Myers-Scotton calls composite CS (see chapter 3 for a discussion).

comment on Thomas's findings, suggesting that they represent "first steps" on the continuum of language death, representing as they do simplification (e.g. stylistic levelling) and restructuring of Welsh on the basis of English influence (1988:102).

Jones (1998) also discusses whether or not Welsh is undergoing language obsolescence (i.e. death), noting that "[m]odern spoken Welsh is displaying reduction, simplification, increased linguistic transparency and quite prolific lexical borrowing" (1998:257). Examining two Welsh dialects—that of Rhymney in an Anglicised part of south Wales and the dialect of Rhosllannerchrugog in a more Welsh-speaking village in north Wales—Jones draws attention to contact-induced phenomena such as calquing (of which there are "numerous instances", particularly in the speech of speakers younger than 60 [1998:83]), e.g. the phrasal verb construction using a preposition *i ffwrdd* in (7) below, where the historical Welsh form would be *diffodd* 'extinguish'.

(7)		troi	e	i ffwrdd
		turn.2S.IMPER	3Sm	off
		'Turn if off'	(taken from Jones 1998:83)	
	[Hist. Welsh:	diffodd		e
		extinguish.2S.IMPER		3Sm
		'Turn it off']		

Calquing of this sort, Jones (1998:86) argues, is "an indication of language obsolescence". Jones also notes the use by schoolchildren she recorded of an emphatic *yn* aspectual particle in utterances such as (8) below.

(8) oedd e **yn**

be.3S.IMP 3Sm PRT

'He **was**' (taken from Jones 1998:85)

Jones ascribes this trend to influence from English emphatic *do* and *was*, and argues that in cases such as (8) *yn* "is being reinterpreted not as a predicate or as a verbal adjunct but rather as the verb itself" (1998:85). Jones' findings from comparing the two corpora indicate that there is some degree of language obsolescence and dialect convergence[25] (to Standard Oral Welsh) in spoken Welsh from both north and south Wales (1998:289), but that the southern dialect showed a more advanced stage of obsolescence, probably because it is in an area where English is a more prevalent language than the area of Rhosllannerchrugog (1998:102, 194).

Willis (2008) discusses instances of grammaticalisation of indefinite pronouns in Welsh. Weak negative polarity items such as *neb* 'anyone, no-one' occurred in both negative and non-negative contexts in Middle Welsh, as in the negative example in (9) and the affirmative example in (10) below.

(9) ny eill **neb** uynet drwydi

NEG can.3S.PRES anyone go.NONFIN through.3Sf[26]

"No-one can go through" (taken from Willis 2008:1)

[25] I.e. "convergence" in the general sense of one code becoming more similar to another code.

[26] My glosses to Willis' examples.

(10)	o	gelly	wneithur	da	y	**nep...**
	if	can.2S.PRES	do.NONFIN	good	to	anyone
	"If you can do good to anyone..."			(taken from Willis 2008:2)		

Over time, however, *neb* ceases to be used in certain contexts, such as the Yes-No question in (11), which Willis points out is unacceptable in contemporary Welsh according to grammaticality judgements, whereas *neb*'s usage has been extended to sentences which lack a sentential negation marker, such as in (12).

(11)	*oes	**neb**	wedi	ffonio
	be.3S.PRES	anyone	PRT.PAST	phone.NONFIN
	"Has anyone phoned"?		(taken from Willis 2008:2)	

(12)	dw	i	'n	gobeithio	gweld	**neb**
	be.1S.PRES	1S	PRT	hope.NONFIN	see.NONFIN	no-one
	"I'm hoping to see no-one"					
				(taken from Borsley and Jones 2005:79)		

Willis argues that the change in the distribution of *neb* in Welsh has influenced by English *no-one*, and that the innovative collocation *unrhyw un* 'anyone' (lit. *un* 'one, a' + *rhyw* 'some, any' + *un* 'one') has emerged in Welsh, as an adaptation of existing Welsh *unrhyw* 'same', to function in contexts parallel to English *anyone*. He comments that such a change probably occurred during the 20th century (2008:4). I interpret the changes to which Willis refers to be convergence, since the frequency of use of items within an existing category (indefinite pronouns) has been changed due to the influence of the distribution of similar items in English.

Deuchar (2006) examines Welsh-English bilingual utterances which possibly show word-order convergence. An example is given in (13).

(13)	fi	'di	bod	i	'r	<u>bus</u>	lle
	1S	PRT	be.NONFIN	to	DET	bus	place

"I have been to the bus place." (taken from Deuchar 2006:1996)

This is a bilingual clause. All morphemes are Welsh except for the CS insertion *bus* from English, but the word order in the NP *bus lle* is not the expected Welsh word order. The Welsh word order for these morphemes would be *lle bus*, where the modifier *bus* follows the head noun *lle.* Instead, the modifier is found preceding the head. Deuchar notes that the morphemes in this clause have a one-to-one correspondence with English, and suggests that it could be identified as convergence of Welsh towards English.

Deuchar and Davies (2009) analyse two similar models of language change proposed by Thomas (1982) and Myers-Scotton (1998a). Both authors view the process of contact between two languages in a community which can lead, ultimately, to speakers abandoning the original community language in favour of another, as having an intermediate stage (Deuchar and Davies 2009:26). At this stage, the other language provides some of the structure for the original community language. This is what Myers-Scotton terms composite code-switching (see chapter 3), which she suggests can lead to a "matrix language turnover" (Myers-Scotton 1998). Deuchar and Davies use Myers-Scotton's Matrix Language Frame model (also see chapter 3) to analyse Welsh-English conversation data for evidence of the influence of English structure on the speech of Welsh-English bilinguals. They find

that Welsh is the source of the morphosyntax in the great majority of clauses (94% of monolingual clauses and 95% of finite clauses), and furthermore that Welsh is almost always the supplier of morphosyntax in bilingual clauses (2009:33). Clauses which did not demonstrate a clear matrix language were considered to show possible convergence, but only one such clause was identified.[27] However, the overall results point to the absence of a structural shift, or matrix language turnover, from Welsh to English, since very little evidence of word-order convergence was found.

Davies and Deuchar (in press) again focus on the extent of word-order convergence in Welsh, examining more data than in Deuchar and Davies (2009), using a similar application of the Matrix Language Frame model as a method. They again find that Welsh is the source of morphosyntax for the structure of 99.7% of bilingual clauses produced by the speakers analysed. Davies and Deuchar argue for a continuity in the grammar of Welsh despite the influence of English structure.

The inference from these studies is that there is limited evidence for word-order convergence in Welsh. Whilst English does apparently have occasional influence on Welsh word order, Welsh is very frequently the source of morphosyntax for the clauses produced, both monolingual and containing CS, by the speakers these studies analysed. Yet note that, these studies aside, convergence in Welsh has not been specifically targeted as a domain of study by previous authors. This thesis will attend to this issue.

In this section I discussed previous research in the field of contact linguistics which supports my interpretation of convergence as a

---

[27] Since I use some of the same data in this thesis as was analysed in Deuchar and Davies (2009), this clause showing possible convergence is presented and discussed in detail in chapter 5 where I discuss the results of my data analysed for this thesis.

mechanism of language change, and I defined it as the increase in the frequency of the distribution of a construction in L_A which is congruent with an equivalent construction in L_B over a construction in L_A which does not have an L_B counterpart. The term convergence has, however, been interpreted differently to this by some authors, and so I will consider these interpretations in the next section.

4. Other views of convergence

In this section I briefly consider other views of convergence which differ from the one I adopt in this thesis, explaining why I do not adopt these alternative frameworks. I also discuss other forms of language change (e.g. those proposed by Thomason and Kaufman 1988, van Coetsem 1988), and explain why I do not follow those approaches. To conclude the section I also briefly discuss some effects Welsh has had on English grammar in Wales.

Though I interpret convergence to be a process, following e.g. Backus (2004, 2005), other researchers have seem to interpret it partly as an effect (cf. Bullock and Toribio 2004:91, Aikhenvald 2006:45), where one language 'converges' towards another language (or both towards one another) because of some other process. Myers-Scotton (2002) sees convergence as both a process and as a result, with the process being where lexical structure from one language is combined with lexical structure from another language, and the result being restructuring of the grammar under the influence of the "stronger"[28]

[28] "Stronger" here presumably means in some sense dominant, e.g. sociolinguistically.

language (2002:164). Silva-Corvalán (1994) also views convergence both (implicitly) as a process[29] which leads to two languages becoming more similar to one another ("[t]he higher frequency of use of a form in language S ... in contexts where a partially corresponding form in language F is used either categorically or preferentially" 1994:4) and as a potential result of direct structural transfer, thus leading to "greater structural similarity in a given aspect of the grammar of two or more languages" (1994:4–5). Silva-Corvalán's interpretation echoes the definitions of convergence given above (e.g. Backus 2004:180, Thomason 2001:89), though I differ from both Silva-Corvalán and Myers-Scotton in that I do not define convergence as a result, but only as a process.[30] I propose that it is preferable to interpret convergence as a mechanism rather than as a result (or as both a process and a result), because, as seen by the framework of contact-induced change which I discuss in section 5, language change is best viewed as a sequence of process and effect, where certain processes result in certain types of change. According to Backus (2005:320), such processes include convergence, grammaticalisation and reanalysis, while the potential resultant changes of these processes can take the form of the addition, loss, replacement or redistribution of elements in one or both languages in contact (Backus 2004:179, 2005:320; Thomason 2001:85). Maintaining this distinction between process and result clarifies the progress of a particular language change (cf. Winford 2007:27, "[r]ecognition of the primacy of 'process' over 'result' leads ... to more meaningful classifications and analyses of the outcomes of contact-

---

[29] Silva-Corvalán (1994) uses the term "transfer" for this process.

[30] I concur, nevertheless, that the result of convergence can indeed lead to the linguistic situations they describe, where two languages end up being more similar to one another than they were previously. This, however, is merely a potential effect of convergence, not the convergence itself.

induced change"), and specifying convergence as a process allows for other processes to also result in two languages becoming somehow more similar to one another: e.g. in principle, borrowing of a lexeme from LB into LA in fact makes both languages slightly more similar to one another—because they now both share a lexeme where before they did not—but it is surely inappropriate to call such an event convergence. Classifying convergence as a process also allows for this process to be quantified efficiently, so that the result of convergence in a language can be measured by the extent to which that process has occurred in the speech of that language's speakers.

An instance of authors interpreting convergence differently to me, even though the end result is broadly similar to the convergence process, and even though they too view it as a process rather than as a result, is found in Bullock and Gerfen (2004). They analyse the phonology of French-English bilinguals in Frenchville, Pennsylvania, and claim to find an example of phonological convergence towards English. The authors explicitly discuss convergence in terms of a process, disputing others' claims that convergence involves "loss through reduction and simplification" (2004:96). They examine the loss of the allophonic or surface distinction between the mid front rounded vowels [ø] and [œ] in the speech of the French-English speakers they analyse: these two vowels have merged into a single variant, the rhoticized schwa (2004:101). They argue that a reason why French-English bilinguals converge to [ɚ] as "a close perceptual map" for the more standard French [ø] and [œ] (2004:102) is that all three vowels are formed with lip rounding, and that the closest American English vowel equivalent to mid and centralised and rounded is [ɚ]. The shared properties between French and English here are vowel height, roundness

and central position, and Bullock and Gerfen suggest that convergence in the French vowel system towards English has occurred here, with the speakers they analysed producing the vowel [ɚ] in the place of any mid central rounded vowel. They also note that the shift from French to English is consistent with the general predominance of English in this particular community.

I do not interpret, however, the process Bullock and Gerfen identify as convergence. The speakers may well produce [ɚ] in French words which have a mid central rounded vowel as some form of phonetic compromise, and the result of this process is that the speakers' vowel system is innovated to include [ɚ] rather than [ø] and [œ], but I defined convergence as the increase in frequency of an existing structure in a language, and, to my knowledge, [ɚ] does not exist in French, though it does in American English. A more satisfying explanation of the Frenchville phenomenon is that it is direct transfer[31] of American English [ɚ] into the French of these speakers to replace the lost vowels [ø] and [œ], based perhaps on the similar properties of the American vowel to the French vowels. This does not appear to be a matter of a change in frequency of an existing pattern, and so I do not interpret it as convergence. Instead I interpret it as direct transfer.

I do not adopt the distinction Thomason and Kaufman (1988) make between borrowing and interference through shift. In their work, borrowing, the incorporation of foreign features into a speaker's first language (1988:37), is linked to language maintenance, while interference is linked to language shift, where imperfect acquisition of the target language (the L2) is affected by knowledge of the speakers'

[31] See section 5 for my discussion of this process.

native language (the L1), and such "errors" as the learners make eventually spread into the target language as a whole, leading to change in that language (1988:38–9). In both borrowing and interference through shift, both schematic copies (construction types or forms) and substance copies (specific items such as words or phonemes) can be borrowed from one language into another (Thomason and Kaufman 1988; cf. Pakendorf 2007:28). The difference between which language is affected, and whether the situation involves borrowing or interference, is largely a sociolinguistic one (Pakendorf 2007:28). Whether it is substance or schematic copying which takes priority is also said to be determined by whether the situation is either one of borrowing or of interference through shift (Thomason and Kaufman 1988:37–9).

In the framework I adopt for this study, I will not incorporate the distinction Thomason and Kaufman draw between borrowing and interference. The reason for this is that I propose to focus my investigation on the grammatical categorization of contact-induced change, and Thomason and Kaufman's distinction seems to me to primarily based on extralinguistic factors such as the desire by a speaker community to "give up their native language and speak some other language instead" (1988:39). Such a description does not fit the situation in Wales, given the increase in the proportion of the population who speak Welsh (see chapter 1), and the types of change we might expect to see in the speech of Welsh-English bilinguals is a change in Welsh itself rather than a shift (imperfectly learnt or not) to English.[32]

---

[32] The questionnaires completed by participants in the study presented here (see chapter 4 for more details) included questions where respondents were asked to self-report on their speaking ability in both Welsh and English. The results do not indicate a speech community where speakers are abandoning Welsh for English, since speakers predominantly report either equal fluency in Welsh and English or slightly better ability in Welsh than English. Some speakers reported their English was slightly better than their Welsh. Another question asked for speakers' emotive attitudes to the Welsh and English

I also do not follow van Coetsem's (1988) framework of contact-induced language change. He differentiates between borrowing and "imposition", a distinction which is related to agentivity and relative language dominance (e.g. Winford 2007:26). Borrowing in van Coetsem's framework involves a speaker adopting elements from a non-dominant L2 into a dominant L1, whereas imposition involves a speaker adopting elements from a dominant L2 into a non-dominant L1. Thus, in a borrowing scenario the language which is more dominant for a speaker is affected by the less dominant language, and vice versa in an imposition scenario. With regards to the Welsh situation, speakers would either be dominant (i.e. more proficient) in Welsh or in English, and so if, for example, a speaker is dominant in Welsh and her Welsh shows influence from English, then this would be characterised as borrowing from English into Welsh according to van Coetsem's framework; if a speaker is dominant in English and her Welsh shows influence from English, then this would be characterised as imposition of English onto Welsh according to van Coetsem's framework. However, this framework seems to be applicable only to scenarios where the bilinguals are not balanced in terms of relative language dominance, whereas the speakers in the speech corpus I analyse are predominantly balanced bilinguals: 64% self-reported equivalent speaking ability in Welsh and English, and very few speakers reported that they had substantially more ability in one language than another. Given this, and also that my focus in this thesis is the effect of English on Welsh rather than Welsh on English, I propose that van Coetsem's concept of imposition is not relevant to the present study.

---

languages, and Welsh was predominantly viewed more favourably than English by the speakers in the corpus.

As noted in the previous section, convergence is not necessarily a one-way process: it can be either unidirectional (L_A converges to L_B but L_B remains unchanged) or bidirectional (L_A and L_B converge towards one another, albeit potentially in different parts of the grammars). The focus of this thesis is the effect of English on Welsh, I.e. where Welsh converges towards English. However, it should be noted that there have also been substantial contact effects of Welsh on English: this dialect is referred to as Welsh English. While I will not attend to such phenomena in this study, it is worth noting, to conclude this section, one or two effects that contact with Welsh has had on the grammar of English as spoken in Wales.

Thomas (1985) broadly categorises Welsh English into two groups: the eastern variety which is less affected by Welsh, and the western variety which, due to a higher frequency of Welsh-English bilingualism, in which "the traces of the Welsh language are more evident", and there is clear substratum borrowing from Welsh into the English dialect spoken in those parts of Wales (1985:215). Thomas (1985:214) notes the influence of the Welsh periphrastic construction of *bod* 'be' + an uninflected verb (e.g. *mae o'n mynd* 'he goes', where *mae* is an inflected form of *bod* and *mynd* 'go' *is* nonfinite) in unmarked Welsh English constructions like (14) below (which are "regularly attested"), where the continuous form of the present participle *is going* approximates the Welsh progressive form *mae...yn mynd*.

(14) WE: He's going to the cinema every week

Welsh: Mae ef yn mynd i

be.3S.PRES 3Sm PRT go.NONFIN to

'r sinema bob wythnos

DET cinema every week[33]

"He goes to the cinema every week"

[cf. Standard English: He goes to the cinema every week]

Thomas (1985:215) suggests that the above construction is typical of the Welsh English found of established western Welsh-English bilingual communities, whereas communities in south-east Wales, which are closer to England, display different patterns and show more overt influence from Standard English or Midland English dialects;[34] he takes this to indicate that Welsh grammar has had influence on this type of construction (cf. Penhallurick 2004:110–11).

Thomas (1985:216–7) and Penhallurick (2004:104) both draw attention to the influence of Welsh VS main clause word order in subordinate clauses of indirect questions in Welsh English (albeit "almost exclusively from south-west Wales", Penhallurick ibid.), so that instead of the Standard English construction shown in (15), a construction as shown in (16) is found.

(15) (Standard English) I wouldn't know **if there would** be any there now

[33] My gloss of Thomas's (1985:214) example.

[34] Thomas acknowledges that (as of 1985) a redistribution of these features is in effect, indicating that they are now stylistically, rather than regionally, marked.

(16) (Welsh English) I wouldn't know **would there** be any there now

The verb *would* in (15) is positioned after the subject *there* in Standard English SV word order, but in (16) it is positioned before it, in VS word order, which would be the word order in the Welsh equivalent, shown in (17).[35]

(17)	Faswn	i	ddim	yn	gwybod		os
	be.1S.CONDIT	1S	NEG	PRT	know.NONFIN		if
	byddai	yno	unrhyw		un	arall	yno
	be.3S.CONDIT	there	any		one	other	there
	rwan						
	now						

"I wouldn't know if there would be anyone else there now"

Despite the contact-induced effects shown here (among others), Thomas (1985:219) points out that the grammars of Welsh and English have remained quite distinct, and have "very largely maintained their structural integrity", that there are not many features of Welsh English which require an invocation of influence to Welsh to explain them, and, perhaps importantly from the perspective of the present study, that the same is true—according to Thomas—for the influence of English on Welsh structure. Thomas finds non-lexical borrowing of English into Welsh only "incidentally" (cf. Thomas 1982), and ascribes this to the

[35] The elision of *if* in (16) is also ascribed by Thomas to Welsh influence, since *os* or *a* in the same context in Welsh—as seen in (17)—is often elided in informal speech (1985:217).

notion that borrowing at the grammatical level is resisted more effectively by language systems (1985:219–10).

Thomas (1985), Williams (2000) and Penhallurick (2004) discuss predicate fronting in some dialects of Welsh English, where the predicate of a sentence is positioned before the subject in an affirmative or negative statement, e.g. *standing outside the door he was* (cf. *he was standing outside the door*). Williams (2000:226–7) argues that this feature of Welsh English has been influenced by the phenomenon in Welsh grammar whereby 'new' information is in clause-initial position ("into the 'verb-slot' goes whatever constituent the speaker/writer may consider new", 2000:218) , as illustrated in (18) below; cf. the (attested) Welsh English equivalent in (19) which shows fronting of new information.

(18)	Cymry	'dyn	ni,	welwch	chi
	Welsh	be.1PL.PRES	1PL	see.2PL.PRES	2PL
	"We are Welsh, see."				

(19) Welsh we are, see

[cited in Williams 2000:224]

Williams found that such constructions were produced primarily by L1 Welsh speakers from a region of Wales where "the Welsh language is more than holding its own" (2000:227), but was found less frequently by speakers from regions of Wales where Welsh is spoken by only a small minority. Note that the directionality of this change appears to be one-way: these authors do not suggest that predicate fronting in Welsh

is being affected (or has been affected) by the process of being borrowed into Welsh English.[36]

Other discussions of notable features of Welsh English syntax which derive from the substrate influence of Welsh grammar are available (e.g. Paulasto 2006), but I will not examine them further here. It will suffice to reiterate that word order change between Welsh and English is certainly not unidirectional, only that my focus in the study presented here is on the influence of English on Welsh.

5. Other processes of contact-induced language change

Convergence, I have argued above, is a process of language change, but it is not the only process that can occur. Other processes, which will not be the focus of this study but which are worth outlining for comparison, are discussed in this section.

Backus (2004) describes contact-induced change as encompassing various stages, namely (i) the causes of change (ultimate and proximal), (ii) the mechanisms of change (causal and processual) and (iii) the end effects/results of those mechanisms, being the change itself (2004:179). Backus defines ultimate causes of change as being primarily extralinguistic, encompassing such factors as social

---

[36] This returns to the point I made in section 2 above, that processes such as convergence can affect both languages in contact, but not necessarily in the same way or in the same part of the grammars: for instance, predicate fronting might be borrowed into Welsh English under the influence of Welsh, but different changes (e.g. adjective/noun relative word-order in NPs) might be borrowed from English into Welsh.

demographics of respective language communities, relative prestige, etc., but suggests that they are "generally too abstract" to be usable as explanations for specific instances of change. Of more direct relevance to this study is Backus's distinction between causal change mechanisms (also proximal causes) and processual change mechanisms.

Causal mechanisms are the effects of social factors on speech, frequently deriving from demographic asymmetry (e.g. in language dominance). CS, for example (see below for a discussion of this phenomenon), is argued to be a causal mechanism, in situations where speakers code-switch into a socially dominant language from their socially less-dominant language (Backus ibid.). Processual mechanisms, meanwhile, link causal mechanisms to language change (ibid.). Processual mechanisms link causal mechanisms to language change (2004:179). Backus sees the process of language change (at least partly) as a shift in entrenchment of words, constructions, etc. Degrees of entrenchment of such items fluctuate due to usage, particularly when an L_A construction is in competition with an L_B construction. In a scenario involving code-switching, for example, an increase in use of an L_B word by speakers of L_A would increase the entrenchment of that word in L_A. Extension of L_B usage into domains previously predominated by L_A would also change the degrees of entrenchment of L_A and L_B constructions (Backus ibid.). Backus does not, however, suggest that each specific mechanism leads to a specific change in a set of one-to-one mappings ("one change often leads to another, and often various changes interact" 2005:327).

Backus (2004, 2005) provides an overview of five main different types of processual mechanisms, as well as the potential effects they might have on language, allowing for the caveat that in actual data there may be overlap between mechanisms/changes. I represent his

framework in table 2 below (cf. the tables in Backus 2004:180 and 2005:329).

Mechanism	Processual mechanism of language change	Possible result of this process	The term I use
A	Use of LB expression realized with LA morphemes	System-preserving or system-altering calquing	*Calque*
B	Use of LB pattern *previously unknown in LA* realized with LA morphemes	System-*altering* changes in the *inventory* of grammatical categories	*Direct transfer*
C	Use of LB patterns *congruent with one of two or more LA patterns* realized with LA morphemes	System-*preserving* changes in the *distribution* of grammatical categories	*Convergence*
D	Use of one or more LA patterns congruent with an unmarked LB pattern	System-preserving changes in frequency of grammatical categories	*Redistribution*
E	(Continued) use of LA pattern which may or may not be congruent with an LB pattern	No change (stability)	*No change*

Table 2. Types of processual mechanisms of language change, and the possible resultant change.

Backus (2004:180) differentiates between system-altering and system-preserving changes. System-altering changes are the addition or loss of an entirely new category to a language, while system-preserving changes are a form of restructuring: they "alter the way in which a category is expressed" (2004:180) or add a novel term to an existing category (Aikhenvald 2003:2), rather than introducing new categories. Aikhenvald (ibid.) describes system-altering changes as the "significant restructuring of a grammatical system", such as a change in a language's typology from head- to dependent-marking (by acquiring dependent-marking properties), because they introduce new categories under the influence of the language(s) in contact. System-preserving changes, on the other hand, involve a degree of redistribution within existing categories in L_A, based on a differing distribution in L_B.

Broadly speaking, the processual mechanisms shown in table 2 above can result in either (i) no change, (ii) system-preserving change or (iii) system-altering change to the language in which they occur.

Mechanism A, the use of morphemes from L_A in a pattern from L_B ("translations of actual word combinations in the other language", Backus 2005:330), may be presumed to be either a system-altering change, if it introduces a new structure (such as a specific word order) into L_A, or system-preserving if the LB pattern used already exists elsewhere in L_A. Mechanism A is what is usually termed **calquing** (cf. Winford 2005:141, Aikhenvald 2006:27, Bowern 2008:202, etc.). An example of this mechanism is the semantic copying in the use of *denken* 'think' rather than *glauben* 'believe' in Australian German on the basis of the usage of English *think* (Clyne 1967).

Mechanism B involves the adoption[37] of a new construction from L_B into L_A, which is realized using morphemes from L_A ("importing a foreign combination without actually using any foreign morphemes", Backus 2005:322), and what Silva-Corvalán (1994:4) described as the incorporation of a form (lexeme, phoneme, construction, etc.) from L_B into L_A that was "previously absent" in L_A. I follow Silva-Corvalán (1994:4) in calling this process **direct transfer** in this thesis. Direct transfer is a system-altering process as it introduces a new form/category into L_A from L_B which previously did not exist in L_A. This mechanism B appears to overlap with Backus' mechanism A, in that they both involve L_B constructions realised with L_A morphemes, but differ in that mechanism B introduces a general pattern into L_A from L_B, whereas mechanism B introduces a specific pattern into L_A from L_B. Mechanism B is sometimes called structural borrowing (e.g. Field 2005, though cf. Backus 2005 who uses "structural borrowing" as a general term for any language change that is not internal).

Mechanism C is what Backus (2004) calls **convergence**, and I have discussed it above.

Mechanism D, the increase in frequency of an L_A construction on the pattern of a more frequent L_B construction, may be presumed to be system-preserving, in that there is no addition of a new construction, but a redistribution of existing constructions. Mechanism D differs from

---

[37] Note that adoption is a term used by Johanson (2002b:265) to describe a process very similar to what is described in this paragraph, where a form from one language which is socially dominant is borrowed into another language which is socially less dominant; he also calls this take-over. He gives an example of a native Swedish speaker such as himself using lexical and syntactic copies from English when speaking Swedish. My definition of direct transfer described here differs from adoption primarily in that Johanson's term stipulates a specific dominance disparity between the languages in contact—the 'opposite' procedure, where a form from a socially less-dominant language is borrowed into a socially more-dominant one, is called imposition or carry-over by Johanson (2002b:266; cf. the similar distinction Van Coetsem [1988] makes between "borrowing" and "imposition").

mechanism C in that it does not involve the necessary decrease in use of an incongruent pattern, and so probably cannot result in system alteration. An example of this mechanism is the increase in the frequency of the usage of diminutives in East Sutherland Gaelic under the influence of Scots English (Dorian 1993). I term this mechanism **redistribution**. Backus (2005:334) concedes, incidentally, that most processes of language change are probably "of the frequency-changing kind", but such changes over the course of a contact situation are, he argues, practically impossible to prove empirically.

Mechanism E does not result in change at all, and so is system-preserving.

It should be clear that the distinction between the mechanisms described in the above paragraphs derive from the different distribution of the borrowed elements in the languages in contact. For example, the difference between convergence and direct transfer is that the 'new' element was present in the source language before the process of convergence took place, but was not present in the source language before the process of direct transfer took place. This classification removes the requirement for differentiating contact situations on a sociolinguistic basis (e.g. Thomason and Kaufman 1988) or on a psycholinguistic basis related to dominance (e.g. van Coetsem 1988).[38]

As I specified above, the results of mechanism C, convergence, can either be system-preserving or system-altering, depending on what happens to the incongruent construction whose frequency is reduced in the process (i.e. whether it is lost entirely over time through decreased use, or whether its function is simply limited as a result). Given that I interpret convergence to be a process, whereas loss is best interpreted

[38] That is not to say, of course, that such methods of distinguishing types and forms of contact-induced change are not undoubtedly useful if the focus of the study warrants it.

as a result of certain linguistic processes, then I suggest that convergence should be viewed as a system-preserving mechanism through which the distribution of an existing pattern is affected. It is important to stress that 'system-preserving' does not entail 'no change'. There is indeed a change when processes such as convergence or redistribution occur, but the type of change does not involve the introduction of new structures or patterns from one language into another: the change is instead a restructuring of patterns that already existed in the language undergoing change.

Having explained the overall categorisation of contact-induced change adopted in this thesis, I return in the next section to discussing convergence, now turning the focus to three important aspects of convergence: code-switching, language mode and congruence.

6. Code-switching

A prominent contact-induced process found in bilingual situations is code-switching (CS), which is the use by a bilingual speaker of material from more than one language in the same discourse (e.g. Poplack 1980:224, Winford 2003:101, Thomason 2001:132; cf. Myers-Scotton 2002:7, who uses the term "bilingual speech" for the general phenomenon, and Muysken 2000:1, who uses "code-mixing"). Backus (2004) considers CS to be a synchronic reflex of language contact, specifically of certain social factors like the prestige of the respective language communities, and he notes that asymmetries in these factors between speakers holding a conversation have a direct effect on the extent and nature of the CS in that conversation (2004:179). Myers-

Scotton's (e.g. 1993:67) Markedness model claims that every community has a choice of code which is considered more unmarked than other code choices; e.g. a bilingual community might consider one of the community's languages more unmarked than another of the community's languages, and use that one in preference to the other.

CS is argued to play a role in language change (e.g. Backus 2005), which I discuss in detail in section 8 below. CS is of particular relevance to the study described in this thesis, since I will be proposing in chapter 3 that a model designed to describe the way languages are mixed in code-switching (Myers-Scotton's [2002] Matrix Language Frame model) can be used to identify synchronic instances of convergence.

CS can occur intersententially, where a speaker produces an utterance in one language and then produces another utterance in a different language, or intrasententially, where a speaker uses morphemes from more than one language within the same utterance (Myers-Scotton 1993:3, Thomason 2001:132, Winford 2003:105, Deuchar 2005a:255).[39] An example of intersentential CS is shown in (1) below, in which French words are underlined, whilst (2) shows an example of intrasentential CS, in which the English word is underlined.

---

[39] Some of these authors differentiate between these two phenomena by referring to one type as code-mixing and the second type as code-switching, or use other terminology. I will use herein the blanket term 'code-switching', and differ instead between intersentential and intrasentential CS, as per e.g. Thomason (2001). Note, however, that some authors (e.g. Winford 2003:105) use the term intrasentential ("or, more accurately, intra-clause") for CS between clauses; I do not follow this practice.

(1) Nadine est née au mois d'avril en dan in de maand oktober heb ik een winkel opengedaan in...

"Nadine was born in April and then in October I opened a shop in..."

(taken from Treffers-Daller 1994:30)

(2) Maestro, y a ónde vamos a ir al swimmin', ónde, ónde?

"Teacher, where are we going to go swimmin', where, where?

(taken from Smith 2006:210)

The utterance shown in (1) was produced by a French-Dutch bilingual living in Brussels. It contains two clauses, with the speaker changing from speaking a monolingual French clause to speaking a monolingual Dutch clause. Each individual clause consists of morphemes and structure from just one language. The speech shown in (2) is from a Spanish-English bilingual child. All the morphemes in the utterance are from Spanish except for the English participle *swimmin(g)*, which has been inserted into the Spanish framework. Here, while the structure of the utterance is all from Spanish, the morphemes are from both Spanish and English, and the switch occurs within the subordinate clause *a ir al swimmin'*. Since the study I present in this thesis is clause-based, and that the analysis will focus on what exists within a clause rather than what happens between clauses, my focus here will be on intrasentential CS only.

Much research has been done in the past thirty years attempting to describe the extent to which CS conforms to any structural patterns. Some authors (e.g. Muysken 2000, Deuchar, Muysken and Wang 2007) have proposed a typological approach to CS, suggesting that different language pairs display variation across the different types of CS patterns

presented—insertion, alternation and congruent lexicalisation—arguing that such an approach reflects the great variety in code-switching found across various previously-studied language pairs (Deuchar et al 2007:299). Other authors (e.g. MacSwan 2005a and 2005b, Chan 2008), working from the perspectives of the Minimalist Program and Null Theory, argue that bilingual speech operates under the same parameters as monolingual speech, and that CS-specific patterns should not exist (cf. MacSwan 2005a:20, "nothing constrains CS apart from the properties of the mixed grammars involved").

In the past, however, the goal of many studies of CS was to identify whether or not there were universal grammatical constraints in CS. Poplack's (1980:227) free morpheme constraint proposed that CS could not occur between bound morphemes, such as verbal inflections: e.g. a Spanish-English bilingual affixing the Spanish verbalizing suffix *-iendo* to an English verb stem such as *eat* to create **eatiendo* would violate this constraint (1980:227). Poplack's equivalence constraint (1980:228), meanwhile, proposed that CS can only occur in loci where the juxtaposition of surface elements from two languages does not violate a syntactic rule of either language. An example cited by Poplack (1980) of a (fabricated) sentence which shows violation of the equivalence constraint is given in (3) below.

(3) el man que came ayer wants John comprar a car nuevo

"The man who came yesterday wants John to buy a new car."

(taken from Poplack 1980:228)

The use of the non-finite verb *comprar* 'buy' here follows the specifically-English rule of an infinitive complementizer (in Spanish the subjunctive form of the verb would be required, *compre*). Since both

languages are not equivalent here, a switch is not permitted according to the equivalence constraint. Poplack analysed data from Spanish-English bilinguals in a Puerto-Rican community in the USA, and found few counterexamples to her proposed constraints,[40] stating that the speakers she analysed generally adhered to the constraints proposed (1980:241). However, Poplack's constraints have been shown by other authors to be falsified when tested on data from bilinguals in other languages: e.g. Bentahila and Davies (1983) with Spanish-Hebrew data and Berk-Seligson (1986) with French-Arabic data. The inference is that while it is possible to formulate constraints on CS in the speech of some bilinguals, those constraints may not be universal, and it is likely it is not possible to formulate any absolute constraints on CS.

Nevertheless, it may be the case that there are CS tendencies to which some bilinguals generally adhere with their speech. Myers-Scotton (1993, 2002) calls this "classic CS". She posits that a bilingual's language production displays asymmetry between their two languages, and that, assuming certain sociolinguistic conditions are met (such as relative language stability), one language is always the source of certain parts of clausal morphosyntactic structure. That language is labelled the Matrix Language, and the other language, the Embedded Language, is limited in what kinds of morphemes it can supply to the clause (2002:8). In chapter 3 I discuss the notion of matrix language, as defined by Myers-Scotton and others, where I consider the usefulness of the concept as a means of identifying contact-induced structural effects in bilingual speech. In that chapter I will also motivate my decision to

---

[40] Less than 1% of Poplack's data showed apparent violation of the free morpheme constraint; this was where there was CS "within idiomatic expressions" (1980:240). A similarly low proportion of her data showed violation of the Equivalence Constraint (1980:241).

use Myers-Scotton's model to analyse the data for the present study, and discuss in more detail the CS frameworks outlined above.

In the next section I briefly describe the concept of bilingual mode, in which, some argue, CS is facilitated, before proceeding to discuss the connection between CS and language change, and the role CS possibly has in the convergence process.

7. Monolingual versus bilingual mode

Grosjean (e.g. 1998, 2000, 2001) discusses the concept of language mode, whereby the levels of activation of a bilingual's two languages vary along a monolingual-bilingual mode continuum. On one end of the continuum, the monolingual mode, only one language is active: this mode is used when talking to another speaker who shares only one of the bilingual's two languages. The other language is argued to be deactivated. On the other end of the continuum, the bilingual mode, the speaker has both languages activated, though Grosjean argues that one language is still more active than the other (1998:40). When both languages are activated in this way, processes such as CS are facilitated according to Grosjean. From a Matrix Language point of view (see chapter 3), the more-activated language is more frequently used as a Matrix Language source (Grosjean 2000:446). The less activated language is the one used for insertional CS (Grosjean 1998:42). Furthermore, Toribio (2004) found that the more activated both languages become, the more likely it is that one will have a structural effect (what Grosjean (2001:14) refers to as "interference" on the other. In Toribio's study of convergence in the speech of Spanish-English

bilinguals, speakers showed more discourse-pragmatic influence of English grammar in their frequency of subject pro-drop when they were code-switching (bilingual mode) than when they were not code-switching (monolingual mode).[41] This indicates that CS appears to increase the likelihood of structural influence of one of a bilingual's languages on the other. Therefore, a hypothesis could be formulated which suggests that convergence, or any other kind of contact-induced process of change, is facilitated or made more likely when there is CS, but that is not to say that CS is a prerequisite for convergence to take place. I discuss this concept in more detail in the next section.

It is not my specific intention in this thesis to discuss bilingualism from a psycholinguistic perspective, but it will be interesting to examine whether or not the Welsh-English bilinguals I examine here produce more convergence in their bilingual utterances than in their monolingual utterances. For this purpose I shall distinguish between the linguality of clauses in this study—that is to say, whether or not a clause is, in terms of the morphemes used, monolingual or bilingual.

8. Code-switching and language change

Backus (2005) notes that there is a connection between CS and language change, arguing that synchronic phenomena like CS, which are themselves reflexes of language contact, can result in lasting change in the language(s) concerned. For example, most authors agree that there is, for example, a close connection between CS and lexical borrowing (e.g. Poplack 1980:221, Field 2005:342, Thomason 2003:695)—even if

[41] For more details of Toribio's study, see my discussion of convergence in section 8 below.

it cannot be shown that every individual case of CS leads to lexical borrowing, or that every individual lexical borrowing derives from CS (cf. Field 2005:343)[42]—but it is also argued that there is also a link between CS and structural change (see below). In this section I discuss these considerations further.

According to Croft (2000), language change stems from speakers selecting one of three options when speaking: normal replication, where the speaker uses a familiar and established word or structure; innovation, where the speaker uses a word or structure that they have not used before; and propagation, where the speaker chooses a recently-innovated ('new') word or structure over an established ('old') one. Normal replication inhibits language change, and innovation by itself is not sufficient to create a lasting change in a language, but propagation of an innovation by sufficient speakers can result in that innovation becoming an established form, and thereby changing the language (cf. Thomason 2003:694, "any feature that can appear in a single person's speech at any time ... can turn into a permanent change in the entire language"). According to Croft's (2000) terminology, then, a speaker producing CS would presumably be considered to be making an innovation (the introduction of an innovative lexical item); if the same lexeme were then used more frequently by other speakers, it would be propagated and, eventually, would become an established lexical borrowing. CS and lexical borrowing are thus interpretable as two related points along a continuum.[43]

---

[42] Direct transfer (Silva-Corvalán 1994:4) is itself one process of language change (cf. Field 2005), but not all language change derives from direct transfer. Convergence, for example, is distinct from direct transfer, but is also a mechanism which can lead to language change.

[43] Note that Croft's primary criterion for differentiating between CS and lexical borrowing is based on the frequency of that lexeme's use. This makes a continuum-based interpretation

The relationship between CS and structural contact-induced change, however, is not as clear. Despite Thomason's (2003:697) assertion that there is no major link between CS and structural contact-induced change, Backus (2005) argues that, while single-word insertional CS has generally no effect on a language's grammar,[44] CS of more complex structures can have indirect effects on the structure of the language into which CS items are inserted (2005:334–5).[45] Backus cites as an example what he calls "chunks", complex lexical items, which can result in those constructions being calqued into the receiving language (2005:335), such as in Texas Czech, where switched English nouns marked with Czech instrumental case, e.g. *autobus-em* 'by bus', are shown to have led to an increase in the use of constructions with a preposition in the Czech of those speakers (Dutkova-Cope 2001). Treffers-Daller (2005) also proposes that CS of certain types of construction can lead to grammatical change. She finds that French-Brussels Dutch bilinguals use idioms in Brussels Dutch which have French structure and an innovative word order not usually available in Brussels Dutch (2005:503). She ascribes this change to the frequency of the insertional CS of such French constructions in Brussels Dutch

---

inevitable (presumably a lexeme produced once would thereby classify as a CS whereas that same lexeme produced e.g. 1000 times would classify as a borrowing).

[44] He says this is the case except when those insertions are derivationally marked, e.g. the use of a form of Turkish *yap* 'do' with inserted Dutch verbs in Turkish discourse, in which case this occasionally results in an increase in the frequency of such a declension/conjugation.

[45] Thus CS is often structure-preserving, in that it, put simply, adheres to the grammar of the languages in question. CS which is non-structure-preserving, where the structure for the code-switched elements is in some way unexpected (see e.g. what Myers-Scotton (2002) calls composite code-switching, described in chapter 3), can result in language change (see the rest of this section). In a situation where the CS is non-structure-preserving (cf. the bare forms produced in CS by bilingual children in Schmitt [2000]), this could be argued to sometimes lead to imperfect L2 acquisition ('through the back door', as it were), where the structure produced by children in CS spreads. Frequent instances of such non-structure-preserving CS would probably be necessary for this to occur.

speech, and suggests that over time Dutch items could be found in this word order, which would indicate word-order convergence (ibid.). This is reminiscent of Thomason and Kaufman's proposal that "structural borrowing is invariably preceded by lexical borrowing" (1988:113).

Field (2005) notes that the connection between CS and structural change is not a cause-and-effect relationship: while CS might model forms and structures from the other language, CS does not directly lead to lexical borrowing (2005:343). Nevertheless, he suggests that "the existence of one may increase the likelihood of the other" (2005:357). Field (2005:354) proposes a sequence by which CS can lead to change, which requires propagation of the innovated CS by multiple speakers until it becomes a part of the recipient language (cf. Croft 2000). I summarise Field's four stages below.

1. A speaker in bilingual mode "activates" a morpheme/structure from the "donor" language.
2. That speaker uses that morpheme/structure in the "recipient" language, i.e. as a code-switch.
3. When more and more speakers produce the same morpheme/structure, a "critical mass" of production is reached, the form becoming diffused into the general speech community.
4. The form is conventionalised in the "recipient" language (i.e. it becomes a "borrowing").

According to Field, speakers cease to associate the borrowed form with CS when in monolingual mode (see section 7 above), perhaps at stage 4 of the above process, when the form is used by the monolingual community or, presumably, by bilinguals in monolingual mode (2005:354). Thus CS can lead to language change, but only with the

requisite social preconditions in place, such as speakers' positive attitudes to language mixing and a lengthy and/or intense period of bilingualism in a community (2005:358).

Note that while CS can facilitate (but not 'cause') convergence, it is not not necessarily the only process which does so. I commented above on phenomena centred around L2 acquisition, such as Thomason and Kaufman's (1988) notion of interference through shift and van Coetsem's (1988) notion of imposition. These too are routes by which convergence can take place, where incomplete learning of the target language (the L2) grammar results in "negotiation" by speakers (Thomason 2001), where they over-generalise parts of their L1 grammar to their L2, such as in the speech of people from the former Yugoslavian/Hungarian border, where, during language shift from Hungarian to a Serbo-Croatian dialect, speakers applied a fixed stress rule from Hungarian to Serbo-Croatian, which usually has free stress, resulting in a new form of the target Serbo-Croatian where the stress rule is different to both Hungarian and standard Serbo-Croatian but similar to both (Thomason 2001:143). In this situation Thomason views convergence as being a result where a language becomes more similar to another (in this case, the Serbo-Croatian produced by subsequent generations is more similar to Hungarian), and is a case of interference through shift. There is no indication that CS was involved in this process, so here the proximal cause for the convergence to occur was some degree of imperfect learning of the L2, where the stress rules of L2 were redistributed on the basis of stress rules present in L1: this is categorisable as convergence according to my framework, but occurs in a different context to convergence that is found in the speech of adult bilinguals in which there is CS.

Crucially, my framework does not make a categorical distinction between CS and language acquisition as conduits for the convergence process, because, first, this distinction is related to the different directionality of the borrowing versus interference process rather than any inherent difference in the processes themselves, and, second, CS is a reflex of the speech of bilinguals who probably have some level of fluency, and any convergence that might occur with this CS is what might then get transmitted so subsequent generations via processes such as interference on the L2 or imperfect L1 acquisition—that is, one follows the other, and it is not a mutually exclusive choice which concept is ascribed to, as convergence can arise (or not) in either situation. Since it is probably impossible to synchronically observe language change that has been caused by generational 'mistransmission', I propose that it is easier to identify convergence of the more spontaneous kind in the speech of adults, such as that which co-occurs with CS, and that it is therefore this kind that I shall focus on in this thesis.

Having discussed how CS and language change frequently go together, though not necessarily as a cause and effect, I now proceed, to conclude the chapter, to examining the notion of structural congruence, which plays a role not only in CS but also in contact-induced structural change.

9. Congruence

Authors such as Weinreich (1953), Clyne (1967) and Poplack (1980), and more recently Sebba (1998), Johanson (2002a), Aikhenvald (2006) and Matras and Sakel (2007), highlight the role of congruence (also

termed equivalence by some) in language contact phenomena, both with regard to CS and to language change. Congruence is defined as similarity in parts of the grammars of two languages, and can be subdivided into paradigmatic congruence, being equivalence in grammatical categories (word class etc.), and syntagmatic congruence, being equivalence in surface word order (Deuchar 2005a).

Sebba (1998) argues that the degree of congruence in categories between two languages affects the likelihood of bilinguals producing CS in those categories. He posits a number of congruence outcomes, including harmonization, where particular categories in two languages are "the same" (i.e. fully-congruent) and so a bilingual can code-switch freely between them; neutralization, where speakers identify congruent categories and use them in preference to a more expected incongruent category; and blocking, where incompatibility (non-congruence) between categories in the two languages prohibits CS of that category. The notion of congruence is also integral to Poplack's (1980) equivalence constraint, which posits that CS is facilitated (or "allowed") when the surface syntax of the two interacting languages is congruent. Indeed, Poplack also implies that CS in constituents which are not equivalent across two languages tend not to occur "in actual performance" (1980:229).

As noted above, congruence also appears to be related, via CS, to language change (cf. Sebba 1998:17, "codeswitching and convergence[46] may go hand in hand"). Weinreich (1953) also writes that contact-induced change is facilitated "between highly congruent structures" (1953:33). Backus simply frames it as: "greater similarity facilitates copying". Thomason and Kaufman (1988) suggest that

---

[46] Convergence in this case refers to a more general result of language contact than the process I describe below.

congruence is central in the genesis of pidgins and creoles, where shared structures between the languages in contact are those that are most likely to be adopted in the new mixed language (1988:154). Aikhenvald (2006:32) comments that "pre-existing structural similarity" between languages in contact facilitates borrowing (structural, lexical etc.) of forms or patterns, which matches the general assumption of all the authors above that features which are similar in two languages are somehow easier to borrow from one language into another than features which are dissimilar.

With regard to this, typological closeness between two languages might be considered to enhance the likelihood of structural change, whereas typological distance might obstruct it. Thomason and Kaufman (1988) propose that close matching of categories (e.g. word order) in two languages in contact is likely to result in category replacement, through interference through shift, in one or more of those languages; cf. e.g. the replacement of case and number suffixes in Čakavian dialects of Serbo-Croat by forms from standard Serbo-Croat, where both languages are typologically close (1988:31). In the case of English and Welsh, English has SVO word order and Welsh has VSO, so in this they are typologically distinct. It might therefore be the case that structural influence of English word order on Welsh is predicted to be unlikely, whereas influence of main clause word order of English on Spanish might be more likely, since both have SVO.

However, some authors have shown that typological distance and/or incongruence in fact increases the extent of CS. Berk-Seligson (1986) analysed naturalistic data from Spanish-Hebrew bilinguals, testing the CS constraints proposed by Poplack (1980). Spanish and Hebrew are typologically-distant languages, and are generally incongruent, yet Berk-Seligson found that these speakers frequently

produced CS at loci where the grammar of Spanish and Hebrew were incongruent, thus indicating that congruence is not a prerequisite for CS (1986:334). She also found that L1 Spanish bilinguals exhibited very similar CS patterns to L1 Hebrew bilinguals, which suggests that the degree of bilingualism is not related to CS activity (ibid.). Bentahila and Davies's (1983) analysis of Arabic-French CS, again two languages which are not typologically close, demonstrates similar findings, that general incongruence between the two languages does not block these bilinguals from producing CS. Further, Deuchar (2005a), looking at Welsh-English conversational data, also finds CS is present despite the two languages not being typologically close. She argues that CS is made easier when both paradigmatic and syntagmatic congruence are present in two languages, but is restricted or blocked when only one or neither type of congruence is present (2005a:268).

Given that there is evidence that bilinguals in both typologically similar and typologically dissimilar languages produce CS and that contact-induced structural change also occurs in such language pairs, it seems likely that sufficiently-intense language contact renders CS/borrowing of any part of grammar possible (cf. Thomason 2003:694; also Backus 2005:323: "typological distance is not ultimately a barrier to convergence"). Congruence only has to exist between some parts of the grammars of two languages for structural influence to be facilitated; cf. Winford (2003:93), who points out, for example, that the similar morphosyntactic patterns of Kupwar Urdu and Kannada has led to changes in Kupwar Urdu gender marking and agreement patterns. Further, Bentahila and Davies (1983) found that Arabic-French bilinguals produced CS clauses with VS order, which is common as an Arabic word-order, and is also available in French, though it is not the unmarked main clause word order (which is SV). Thus the congruence in

the relative position of subject and finite verb between Arabic and French mean that clauses such as (4) below are produced, where a French subject (*le contrôle*) is positioned after an Arabic finite verb (*ža*).

(4) ža le contrôle
came the checking-time
V S
"The checking-time came"
(from Bentahila and Davies 1983:319)

I infer from these studies that a degree of congruence between two languages facilitates both CS and contact-induced language change.[47] I am specifically interested in syntagmatic (i.e. word-order) congruence, since in this thesis I am measuring the extent of word-order convergence in Welsh.

[47] Comparable to the concept of congruence is entrenchment, whereby, according to some interpretations, certain parts of a language system are more resistant to change than other parts (cf. Croft 2000, Backus 2005). Parts of grammar which are more entrenched, e.g. inflectional morphology, are less susceptible to change as a result of contact than parts which are less entrenched, e.g. the lexicon (Thomason 2003:693-4). Typological similarity is a factor: highly-entrenched material can be exchanged more readily between languages that are typologically similar than between languages which are typologically dissimilar (Thomason 2003:694), but, as with congruence, a situation of sufficiently-intense bilingualism can mean that the degree of entrenchment ceases to be an important factor. Frequency is also arguably a factor; cf. Aikhenvald (2006:29), who suggests that "[t]he more frequent the category in one language, the likelier it is to diffuse into another", and cites the integration of borrowed French derivational morphology into Middle English as an example of this.

10. Conclusion

In this chapter I reviewed the literature on convergence and other aspects of contact-induced language change, including CS. I have identified that there is limited existing research on convergence in Welsh, and I propose to attend to this gap with the study I report in this thesis. I discussed the process of convergence, reviewing recent studies of it, and have shown that some argue that it is facilitated by CS, bilingual mode and congruence of the two languages' structures.

It seems possible, then, that speech by bilinguals exhibiting CS is more likely to also exhibit convergence than speech by bilinguals which does not exhibit CS. In this thesis I propose to explore this issue by using a framework of CS to test Welsh-English bilingual speech data to see the extent to which convergence is found, and whether it is found more frequently in CS (bilingual) or non-CS (monolingual) utterances. The model in question is the Matrix Language Frame model (Myers-Scotton 2002), and I will describe this model in the next chapter.

Chapter 3 — Theoretical and methodological issues relating to data analysis: the Matrix Language Frame model

1. Introduction

In the previous chapter I discussed the literature on language contact, language change and convergence. I defined convergence (chapter 2 section 2) as the use of a construction from Language A which is available in Language A but is more prominent in Language B, and proposed (ibid.) that the results of word-order convergence can be seen in a clause which has structure from both participating languages.

In this chapter I review the literature relating to the concept of a matrix language, focusing on the Matrix Language Frame model as outlined by Myers-Scotton (1993, 2002). I then show how I adapt this model to identify word-order convergence, by introducing an innovation to the MLF model in the form of the concept of a dichotomous matrix language. This describes how a clause produced by a bilingual contains some structure from both the bilingual's languages, which I propose show signs of convergence. I will describe the results of applying this model to Welsh-English bilingual data in chapter 5.

2. Concepts of a base language for a bilingual clause

In chapter 2 section 6 I mentioned the notion that clauses containing CS are argued to have a matrix language which provides the morphosyntax of such clauses. I discuss this concept in greater detail here.

Many authors (see below) have argued that when bilinguals produce clauses with code-switching, one language acts as the base or host or matrix language for that clause, supplying structure and some morphemes, whilst the other participating language provides some morphemes but little or no structure.

Appel and Muysken (1987) draw attention to various notions of a base language: syntactically, "the base language is the one which imposes a particular constraint on the switching involved", i.e. the one which governs the structure of the clause; psycholinguistically, the base language is the dominant one for that speaker, e.g. perhaps their first language, whereas sociolinguistically the base language is considered to be the unmarked linguistic code usual for a particular setting. They point out that, taking these three aspects into account, a given clause could be assigned a different base language depending on one's point of view. In Myers-Scotton's earlier work (1993), she makes the similar observation that "relative proficiency is a psycholinguistic means of identifying the matrix language" and that, sociolinguistically, the matrix language is the unmarked language in a discourse (1993:67). In some data, however, it is apparently not clear which language is the unmarked choice. Nishimura (1986), for example, experienced difficulty

identifying a base language to her data from Japanese-English bilingual speakers because "both Japanese and English are actively used" within the discourse (1986:179); presumably she means that both languages are used frequently in her data and neither is marked.

In a different approach, Myers-Scotton (1993) proposed a morpheme frequency criterion to identify a dataset's overall matrix language, whereby the language which provides the most morphemes in a discourse sample is that discourse's base language. This is an idea she has since discarded (cf. Myers-Scotton 2002:61) on the grounds that precisely what constitutes a "discourse sample" is ambiguous, and also that whilst the language acting as the source of grammatical structure for a bilingual clause usually supplies more morphemes for that clause, this is not always the case.

There is clearly a degree of difference (and disagreement), then, between views of what constitutes a matrix language. I do not view Myers-Scotton's (1993) frequency criterion as being a useful interpretation of matrix language, nor do I follow the assumptions of e.g. Nishimura (1986) that a dataset can have an 'overall' matrix language. I suggest, as I discuss below, that a matrix language is best-viewed as a clause-specific construct, which thus allows for the variation between language use that is found in many data.

Joshi (1985) argues for a clause-based notion of a matrix language, with the language which provides a clause's finite verb morphology being identified as the matrix language.[48] This concept is also utilised by Klavans (1983), on whose work Joshi partly draws. The

[48] Nevertheless, other authors have suggested that the finite verb morphology criterion for identifying a matrix language is not always useful. For example, Clyne (1987) notes that phonological and semantic similarity in English and Norwegian words used in his Norwegian-Australian English data, e.g. English *come* versus Norwegian *kom*, make it difficult to identify which language those words come from, and so argues that the criterion is unusable for his data.

other language participating in CS is labelled as the embedded language. Joshi recognises that, in clauses containing CS, "the two language systems are systematically interacting with each other" (Joshi 1985:191). Furthermore, Joshi sets out a number of syntactic constraints on code-switching, including the "non-switchability" of closed-class items such as determiners or prepositions (1985:193) and of complementizers (1985:196). According to Joshi, the matrix language for an utterance may be presumed to be identifiable as the language which provides such closed-class morphemes, and the embedded language is the one which does not. Asymmetry between the grammatical roles played by the matrix language and by the embedded language within these constraints is central to Joshi's theory, since "switching of a category of the matrix grammar to a category of the embedded grammar is permitted, but not vice versa" (1985:192).

Nevertheless, Joshi seems to assume that whilst a matrix language is identifiable at the clausal level, it will be uniform throughout a dataset, i.e. that every clause will have the same language as matrix language: cf. his statement that, in his Marathi-English data, "Marathi ... is the matrix language and English ... is the embedded language" (1985:191). It is possible that Joshi assumes that, in any given bilingual scenario, one particular language (Marathi in the case above) is the *de facto* matrix language for speakers of that language.

Joshi's notion of asymmetry between one language's participation in the bilingual frame with another was adopted, with modifications, by others. Most notably, Myers-Scotton (1993, 2002) developed Joshi's ideas of a matrix language into her theory of CS. This theory is formalised as the Matrix Language Frame (MLF) model, where the matrix language for any given clause is identified on the basis of two criteria: morpheme order and certain system morphemes. I discuss this

model and the various principles associated with it in the next section, before going on to describe my adaptation of it to identify language contact phenomena in bilingual speech.

3. Myers-Scotton and the MLF model

Central to Myers-Scotton's theory of code-switching (1993, 2002 etc.) is the Uniform Structure Principle (USP), which states that a constituent in any language has a uniform abstract structure and must be well-formed according to the structural requirements of that constituent type (2002:8). She bases this on empirical evidence, stating that "constituents in any language have a characteristic categorical structure" (2002:120)). Myers-Scotton's (2002:8) Matrix Language Principle states that, in bilingual speech,[49] only one language, called the Matrix Language (ML), acts as the source of the morphosyntactic structure of a bilingual projection of Complementizer (CP). The CP, Myers-Scotton's preferred unit of analysis, is the highest level of structure within a syntactic tree representing a clause, and can contain other structures, such as VPs, NPs and PPs (Myers-Scotton 2002:8). I interpret this to mean that a CP is approximately a clause,[50] following Myers-Scotton ("both independent and dependent clauses are CPs",

---

[49] Myers-Scotton and colleagues later (e.g. Jake, Myers-Scotton and Gross 2002:71) suggest that monolingual clauses also have a matrix language, though the embedded language in such clauses does not, of course, supply any morphemes.

[50] A CP is always a clause, of course, but I use the more generic term 'clause' to cover both the CP and e.g. co-ordinating conjunctions that appear outside the CP—such words occur frequently between clauses in speech, and it is necessary when analysing to associate them with one utterance or another. In this case, co-ordinating conjunctions were analysed as part of the CP they preceded, and this whole element is identified as a clause.

ibid.) and Deuchar (2006:1991). The other language participating in a bilingual CP or clause is called the Embedded Language (EL).

The ML can be the source of any kind of morphemes in a bilingual clause, but the EL is limited in the types of morphemes it can provide. Well-formedness in what Myers-Scotton calls classic code-switching[51] is maintained by this asymmetry between ML and EL, and stated formally in Myers-Scotton's Asymmetry Principle for Bilingual Frames (AP) wherein "only one of the participating languages [in a CP] is the source of the Matrix Language". In bilingual speech which adheres to classic CS, one finds "the morphosyntactic dominance of one variety in the frame" (2002:9). Following this, in a bilingual clause adhering to the AP, the EL can only provide certain types of morphemes. These are any content morphemes, early system morphemes and bridge outside system morphemes. A fourth type of morpheme, outside late system morphemes, can only be sourced from the ML. These are morphemes with thematic roles external to the head of constituent in which they occur. These four different types of morphemes are categorised and defined in Myers-Scotton's 4-M model, which is discussed further below.

Combining the tenets of the USP, AP, and the Matrix Language Principle, Myers-Scotton developed the framework which will be used to analyse the data for this thesis: the Matrix Language Frame (MLF) model. This model was designed to explain the kinds of "structural configuration" that occur in what Myers-Scotton calls "classic codeswitching" (Myers-Scotton 2002:10). The model states that the asymmetry between the ML and the EL predicted by the AP is maintained in a clause containing "classic" code-switching, and, using

---

[51] Myers-Scotton describes a second kind of code-switching, namely composite code-switching, which involves a composite morphosyntactic frame from both participating languages; this is discussed in more detail below.

the application of the two principles of the MLF model, it can be tested on bilingual speech data to identify which language is the ML of any given clause. Assuming that the clause contains classic code-switching, then one participating language will be identified as the ML and the other as the EL.

In the rest of this section I will discuss the specific parameters of the MLF model and demonstrate how it can be used to identify which language is the source of the ML in any given bilingual clause.

a. The principles of the MLF model

Two principles within the MLF model state that both a) outside late system morphemes and b) morpheme order within a bilingual clause showing classic code-switching must come from the ML. These principles are discussed in detail below.

i. System Morpheme Principle

The first principle is formulated as follows (Myers-Scotton 2002:59):

> In Matrix Language + Embedded Language constituents, all system morphemes which have grammatical relations external to their head constituent (i.e. which participate in the sentence's thematic role grid) will come from the Matrix Language.

The morphemes identified here are outside late system morphemes (often informally referred to as "late outsiders" by Myers-Scotton), so-

called because "such a morpheme must look outside its immediate maximal constituents for information about its form" (2002:88).[52] Maximal constituents here are what Myers-Scotton elsewhere terms maximal projections (2002:7), which is the expansion of a head constituent to the phrasal level, i.e. NPs, VPs etc. Late outsiders are system morphemes that agree with elements outside these phrases.

Examples of outsider late system morphemes include subject-verb agreement morphology, because the finite verb has morphology based on the properties (number, person etc.) of the subject constituent, which lies outside that VP, and determiners in languages like German, which are often inflected for case, number and gender, the information for which is contained in the remainder of the clause. I provide specific examples of outside late system morphemes below.

According to the SMP, all outside late morphemes in any given clause are provided by the same language, the ML. Outside late system morphemes belong to one type among four different types of morpheme that Myers-Scotton defines in her 4-M model (2002:75). I briefly discuss the 4-M model here.

Myers-Scotton describes four types of system morphemes present in a language: content morphemes and three types of system morpheme. The system morphemes are classified according to their point of activation in clause production. Early system morphemes are selected by the speaker immediately after content morphemes become activated at the lemma level. They have grammatical relations only within the same maximal projection as their heads (Myers-Scotton 2002:75), and serve to refine the meaning of those content morpheme

---

[52] They are called "late" because Myers-Scotton proposes that they are activated later in the sentence-building process than other types of morphemes; see the discussion of the 4-M model below.

heads (cf. Schmitt 2000:13, Myers-Scotton 2002:75). An example of an English early system morpheme would be the *at* in *look at*, which is selected after *look*, a content morpheme, is selected at the lemma level, and "refines" the meaning of *look* (Myers-Scotton 2002).

Bridge late system morphemes, meanwhile, act as 'bridges' between constituents (Myers-Scotton 2002:91), by "joining together phrases to form a larger constituent" (Myers-Scotton 2008:29). They, like early system morphemes, are dependent on grammatical information from within their maximal projection (Schmitt 2000:13), but differ from early system morphemes in that they lack conceptual meaning. Instead, they interconnect content morphemes without making reference to the grammatical properties of their head (Myers-Scotton 2002:75). That is, they are "invariant in form", having no overt marking for e.g. case (Myers-Scotton 2008:29). An example of a bridge late system morpheme is English *of* where it links two NPs, e.g. *a friend **of** the family* (Myers-Scotton ibid.).

Finally, outside late system morphemes are assumed to be activated last, and are used to 'tie together' the structures that have been built using the other three types of morpheme. They differ from bridge late system morphemes in that they have grammatical relations external to their head constituents, and build structure <u>across</u> phrases (Myers-Scotton 2008:23). Myers-Scotton's explanation of why outside late system morphemes specifically, rather than other types of system morphemes, must come from the ML in classic code-switching is that these morphemes are the most critical grammatical elements in building clause structure (Myers-Scotton 2008:25).

To show how identifying the outside late system morphemes in a clause help identify the ML of that clause, consider example (1) below,

which is taken from Myers-Scotton's Swahili-English corpus (English morphemes are underlined).

(1) Ile m-geni, hata si-ku-<u>comment</u>
DEM/CL9 Ch/S-visitor even 1S.NEG-PST.NEG-comment[53]
"That visitor, I did not even comment."
(from Myers-Scotton 2002:89)

Here all the morphemes are Swahili, apart from the verb stem *comment*, which is English, and it is also the clause's main verb. However, the 1st person negative past verbal morphology on *comment* is from Swahili. As subject/verb agreement and tense/aspect agreement are outside late system morphemes, and these come from Swahili, this identifies Swahili as the ML according to the SMP. English is identified as the EL. If the finite morphology in this clause had been English instead, then English would be identified as the source of the ML by the SMP.

Another example of an outside late system morpheme is found in (2) below, from Schmitt's Russian-English corpus (English morphemes are underlined).

(2) Yesli <u>Mermaid</u> spasy-ot <u>human</u>-a
if Mermaid.NOM.S save-3S.PERF human-ACC.Sm
"If Mermaid saves a human" (from Schmitt 2000:17)

Here the English morphemes *Mermaid*[54] and *human* are both content morphemes, and so may come from the EL or ML, but the Russian

[53] See Myers-Scotton (2002:89) for details on the glosses here.

[54] This assumes that *Mermaid* is not a proper noun. If it is, then I would consider it non language-specific.

singular masculine accusative case marking on *human* is an outside late system morpheme, and so must come from the ML if this clause shows classic CS. The ML of this clause is identified as Russian, and English as the EL (the EL supplies only content morphemes to this clause).

ii. Morpheme Order Principle

The second principle of the MLF model is set out as follows (Myers-Scotton 2002:59):

> In Matrix Language + Embedded Language constituents consisting of singly occurring Embedded Language lexemes and any number of Matrix Language morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the Matrix Language.

The first part of this definition defines the notion of a bilingual constituent, i.e. consisting of morphemes from both participating languages. The second part of the definition describes the hypothesis that only one language will provide the word order for that constituent; that language is identified as the source of the ML, and the other language is identified as the EL for that constituent, providing morphemes but not word order. A clause can also, of course, be a constituent, and so surface morpheme order in a clause will be sourced from the ML.

An example of using morpheme order to identify the ML can be seen in (3) below, a clause produced by a speaker of the mixed language[55] Tsotsitaal.

(3)	...want	ou	Tex	laat	ons	daa'i	group join
	because	old	Tex	made	1PL	DET	group join

"...because old Tex made us join the group."

(from Myers-Scotton 2002:63)

In (3) the English finite verb *join* is at the end of the clause. Neither English nor any of the other African languages which contribute to Tsotsitaal have this word order, but it is present in Afrikaans, indicating that a variety of Afrikaans is supplying morpheme order to this clause, and is thus identified as ML. Applying the SMP to this clause would also identify a variety of Afrikaans as the source of the ML, due to the verbal morphology of *laat* 'made'.

Example (4) is a bilingual clause by a Welsh-English speaker.

(4)	oedd	gynnon	ni	ystafell	yn
	be.3S.IMP	with.1PL	1PL	room	in
	Plas yn Dref,	ystafell	brilliant		
	Plas yn Dref	room	brilliant		

"We had a room in Plas yn Dref, a brilliant room."

(from Deuchar 2006:1994)

[55] A mixed language can be loosely defined as a language which arises from a situation of language contact and which combines both structure and lexicon from the two (or more) languages which are in contact, though there is little consistency in how this is manifested from mixed language to mixed language. See e.g. Matras (2000) for a discussion of the literature on mixed languages.

As noted by Deuchar (2006), the clause in (4) has VSO order, with the finite verb *oedd* preceding the subject *ystafell*, so the main clause word order indicates that Welsh is providing the ML. Furthermore, the English morpheme *brilliant* in the head/modifier NP *ystafell brilliant* 'brilliant room' occurs in the post-head position expected for Welsh word order, rather than the pre-head position expected for English, providing further evidence for Welsh as ML. Since both main clause VS word order and the position of the modifier following the head element in NPs show Welsh rather than English morpheme order, the MOP identifies Welsh as providing morpheme order in (4).

An exception to the principle that morpheme order within a bilingual clause is sourced from the same language is where such a clause contains a contiguous string of EL morphemes which are in word order from the EL and are well-formed according to the EL. These are called Embedded Language Islands by Myers-Scotton, formally defined as "full constituents consisting only of Embedded Language morphemes occurring in a bilingual CP that is otherwise framed by the Matrix Language" (2002:139). EL islands do not, arguably, contradict the MLF framework. An example of an EL island cited by Myers-Scotton is quoted below in (5); it is a clause produced by a Japanese-English bilingual.

(5) How do you say this <u>Nihongo</u> <u>de</u>?
Japanese in
"How do you say this in Japanese?"
(Azuma 1991:193, cited in Myers-Scotton 2002:139)

In (5) the word order of *how do you say this* has the finite verb preceding the object, which is typical of English (Japanese is primarily an SOV language). The clause-final position of the Japanese adposition

de, however, is not typically English. Instead it shows Japanese postpositional order. Because *Nihongo de* is a continuous string of morphemes from the same language, Japanese, and in the appropriate word order for that language, it may be identified as an island (a well-formed PP in this case) from Japanese. English is identified as the ML of (5) and Japanese as the EL.

b. Testing the MLF model on clauses containing classic code-switching

When testing the SMP and the MOP on a bilingual clause that shows classic CS, both principles are expected to indicate the same language as providing outside late system morphemes and morpheme order. That language is identified as the ML for that clause. I will now briefly demonstrate how the two principles of the MLF model can be applied to bilingual clauses from various corpora to identify the ML of those clauses.

The bilingual clause in (6) is taken from Kamwangamalu's study of siSwati-English data (the English morphemes are underlined).

(6) Tennis association l-discuss-ile

Tennis association NCL9-discuss-PERF.FV

le-problem ku meeting yabo

NCL9.DEM-problem LOC meeting NCL9.POSS

ye kugcina

NCL9.ASSOC last

"The tennis association discussed that problem at their last meeting."

(Kamwangamalu 1994:75, cited in Nicol 2001:90)

The clause in (6) is bilingual, because it contains both siSwati and English morphemes. First the SMP is tested on the clause. Even though there is an English verb stem, *discuss*, in the clause, the morphology on that verb is siSwati. The other 'grammatical words' found, e.g. *ku, yabo* and *ye*, and the inflection on the noun *problem*, also from siSwati, may be classified as outside late system morphemes because they look to information outside their constituents for their morphological information. siSwati is indicated to be the source of the ML according to the SMP. The MOP is also tested on the clause. Though the main clause word order is SVO, which is found in both English and siSwati, the word order of segments such as *meeting yabo ye kugcina* 'their last meeting', in which the head noun *meeting* precedes the possessive *yabo*, indicates that siSwati is the source of morpheme order. This points to siSwati as being the ML. Since both the SMP and the MOP point to siSwati as supplying structure, siSwati is identified as the ML of this clause.[56]

Another example is in (7) below, produced by a Chinese-English bilingual.

---

[56] The NP *Tennis association* is analysed as an EL island with word order and morphemes sourced from English.

(7) ni paper hai mei finish a?

you paper yet not finish PRT.AFFIRM

"You haven't finished your paper yet?"

(from a corpus by Wei 1998; cited in Myers-Scotton 2002:9)

Applying the SMP to this clause indicates that Chinese is the source of the function word *a*, which is arguably an outside late system morpheme. Second, the MOP is applied. The morpheme order is clearly from Chinese: e.g. the clause-final position of the verb *finish* indicates SOV order, which is more typical of Chinese than English. Since both principles point to Chinese as sourcing structure, the ML of the clause is identified as Chinese and English as the EL, which supplies the content morphemes *paper* and *finish*.

As can be seen, the MLF's predictions regarding the asymmetry between the participation of the two languages in the formation of bilingual clauses are upheld in the examples shown above, since one language is the source of the ML in each case.

Note that, at least in my view, no clause will ever have "no matrix language". It may frequently be the case that natural data produced by speakers will produce utterances that are in fragmentary form, i.e. in incomplete clauses, and that such utterances will not have enough interpretable material to be able to identify the ML providing structure to that clause. However, this does not mean that the clause has no ML, but simply that there is insufficient information in the clause for the analyst to be able to identify what the ML is. Presumably such clauses do have an ML assigned at some stage in the production

process, presumably at the lemma posited by Myers-Scotton in her language production model (Myers-Scotton 2002:24).[57]

In this study I use the term "unidentifiable ML" to describe these kinds of clauses, rather than e.g. "no ML", to represent the concept that all utterances, bilingual and monolingual, produced by speakers have a language providing structure—although sometimes it appears that this ML is sourced from more than one language, as discussed in detail in section 4. Before that, however, I propose to give a rationale for using the MLF model in this study.

c. Motivating the choice of the MLF model

To conclude this section it is appropriate to motivate why I have selected the MLF model as the basis for my analysis of bilingual data, rather than any other existing or original theory/framework of CS.

As noted in chapter 2 section 6, many alternative approaches have been presented on whether or not there are constraints on CS, and, if such constraints exist, what they are (see Clyne 1987 for a detailed discussion on the validity of the various proposed constraints, and note chapter 2 section 6 of this thesis where I comment that absolute constraints on CS probably do not exist). When analysing bilingual speech containing CS it is important to select the most appropriate theory of CS to support the analysis. 'Appropriate' signifies more than just 'most accurate' (etc.), and by choosing to use one theory

---

[57] The MLF model is itself is partly derived from the Speaking language production model proposed by Levelt (1989).

over another in this instance does not indicate that this theory is inherently 'better' than other proposed theories—only that it is the best and most appropriate theory in the context of the present study.

My primary stipulation for selecting a CS approach include that it satisfies the following criteria:

1. Allows for quantitative analysis of CS data;
2. Is applicable to production data, I.e. actual speech, given that corpus data will be what is analysed;
3. Has been demonstrated by existing research to account for a substantial body of CS data[58];
4. Is usable to identify and/or measure convergence synchronically[59] in the speech of bilinguals in some quantitative way.

I will refer to these as criterion 1, criterion 2, etc., in the discussion below.

As the reader will have gathered, I have chosen to use the MLF model for the present study, for reasons which I will outline below. First, however, I will briefly discuss a number of other existing approaches and explicate why I propose they do not satisfy all of the above criteria, whereas the MLF does.

Poplack's (e.g. 1980) proposals of specific syntactic constraints which CS is not expected to violate, viz. the equivalence constraint and free morpheme constraint (discussed in chapter 2 section 6), are not

---

[58] Following e.g. Clyne (1987:257): "for the implications for linguistic theory to be valid, the assumptions on code-switching must be correct and verifiable from corpuses from as many bilingual situations as possible".

[59] i.e. leaving aside diachronic methods involving real-time, which could presumably be adopted for most theoretical approaches of CS but are not useful for my situation given that I am analysing speech data collected from speakers at one given point in time.

implied to be universal (e.g. Poplack 1980:211), but are apt to describe the Puerto-Rican data described in Poplack (1980) as well as other Spanish-English data. Poplack (2000:207) states that the equivalence constraint "account[s] for virtually all quantitatively significant patterns of switching in most large corpora"—although she does point out evidence of exceptions to the rule (e.g. M'Barek and Sankoff 1988), which she nevertheless notes are rare and confined to certain constructions (Poplack ibid.). Poplack's framework has often been found to apply successfully to CS corpora (see chapter 2), but not universally—I have already noted above (chapter 2 section 6) that authors such as Bentahila and Davies (1983) and Berk-Seligson (1986) have found exceptions to Poplack's constraints in the bilingual corpora they analyse—so **criterion 3** is partially satisfied. **Criterion 2** is satisfied by Poplack's framework, since the constraints are originally based on data from a corpus of spontaneous speech collected from Spanish-English bilinguals from Puerto Rico (1980:215), and the same constraints have thereafter been tested on other bilingual speech corpora. **Criterion 1** is satisfied to the extent that data can be measured for the extent to which it adheres to the limits imposed by Poplack's constraints, by analysing switch types, percentage of spoken words from each language etc., as well as extensive sociolinguistic analysis based on such data (e.g. Poplack 1980). Note, however, that such analyses do not conceive of CS as being at the clause/sentence level, but rather at the morpheme level, by considering e.g. the switch points between words where languages are changed (e.g. Poplack 1980:212), or at the sociolinguistic level, by comparing CS behaviour according to extralinguistic variables (e.g. Poplack 1980) or across different communities (e.g. Poplack 1987). I have found no evidence of Poplack's framework being used to analyse

convergence,[60] and therefore **criterion 4** is not satisfied for Poplack's constraints. I propose that Poplack's constraints are not the appropriate framework for my proposed analysis, since it only satisfies 3 (at most) of the criteria specified above—specifically, its lack of obvious application to study of language change does not make it an ideal framework for my present purposes.

The typology of CS[61] outlined by Muysken (2000), aims to describe patterns of CS as a taxonomy as opposed to proposing CS constraints per se: Muysken denies that there exists a single model of CS (2000:3). Muysken proposes (2000:3; cf. Poplack and Walker 2003:679) that most existing constraint-based theories of CS are categorisable within a three-part typology, where CS takes the form of either insertion, alternation or congruent lexicalization, all of which occur in different structural conditions and vary according to "specific bilingual settings" (2000:3). Insertion (2000:3) in Muysken's typology consists of the use of one or more morphemes (including entire constituents) from one language in a sentence which contains words from another. Alternation (2000:4) consists of a switch from one language into another ("a true switch from one language to the other, involving both grammar and lexicon", 2000:5), whilst maintaining congruence with the structure of both languages at the switch point (cf. Poplack 1980). Alternation can occur either intrasententially (within clauses) or intersententially (between clauses) (Muysken 2000:4). Finally, congruent lexicalization covers the phenomenon where both participating languages share structure for a given construction or part of a clause, and that shared structure can be filled with lexemes from

---

[60] Poplack, Walker and Malcolmson (2006) does present a study of contact-induced language change (primarily lexical) in Quebec, but this does not make use of the CS constraints previously set forth by Poplack.

[61] Muysken refers to intrasenential CS as code-mixing (2000:1).

either language or both languages (2000:6); Clyne (1967) is among relatively few authors who have studied CS in the form of what Muysken calls congruent lexicalization. Muysken (2000:246–8) discusses all three types of CS in detail and in reference to several bilingual datasets, and proposes that different types of CS occur in different types of psycholinguistic and bilingual settings (see also Deuchar et al 2007).

Given that Muysken's CS typology is based on several existing corpora of bilingual speech production, as well as deriving from existing CS theories, it satisfies **criterion 2,** and furthermore has been shown to be successfully testable on Welsh-English speech data specifically (Deuchar et al 2007), although Muysken (2000:250) seems to suggest that overreliance on speech corpora as compared with experimental methods is undesirable. It has also been used (e.g. Muysken 2000, Deuchar et al 2007) to analyse CS data quantitatively (**criterion 1**), using a feature-based scoring system which allows the dominant CS pattern(s) for a given dataset to be identified (cf. Deuchar et al 2007:321 etc.). By its nature Muysken's typology is descriptive rather than constraint-based (that is, it is a typological framework not a model of CS per se), and though some data have been shown to be ambiguous as to which of the three typological types (if any) of CS they represent (e.g. Deuchar et al 2007:336, Poplack and Walker 2003:681; cf. Muysken 2000:231 who acknowledges this problem), it is classifiable as being a universally descriptive typology which has been tested successfully on a substantial and varied body of CS data, satisfying **criterion 3.** Lastly, though Muysken proposes (2000:268) that processes of language change such as those set out by Thomason and Kaufman (1988) can be viewed in terms of bilinguals gradually importing CS types (insertion &c.) from a language in contact that was previously not common or dominant in their speech, aside from this his

typology does not seem suited to identifying convergence in a synchronic way. His typology's capacity as a descriptive model means that it is not appropriate for quantitatively gauging when bilingual speech deviates from a perceived norm, as usually happens in convergence. Such anomalous structure would presumably simply be categorised as one of the three CS types under an application of Muysken's typology, rather than being flagged as being anomalous. The benefit of a constraint-based framework (e.g. Myers-Scotton) for my analysis would be that speech which does <u>not</u> adhere to such constraint norms would be flagged as possibly showing a form of language change. In addition, insertion and alternation are both concerned with lexical CS, and only congruent lexicalization is (partially) concerned with structural CS—and therefore all instances of anomalous word order in bilingual speech would presumably be labelled congruent lexicalization, even though that CS type does not necessitate nonstandard structure at all. Therefore Muysken's typology does not satisfy **constraint 4**, and so it is not appropriate as an analytical framework with regards to my aim in identifying convergence.

Lastly I discuss the viewpoint of the Minimalist Program (MP) on CS. Authors such as MacSwan (2005a etc.) propose that there is no need for a "third grammar" to account for CS, and that all instances of CS are accountable for using the grammar of monolingual speech (see Chomsky 1995)—that is, that lexically-encoded features are matched during the derivation of a sentence to produce a grammatical sentence (MacSwan 2005a:15). Proponents of the MP aim to construct a theory of CS which does not resort to CS-specific constraints (MacSwan 2005a:5). Its strength, therefore, is that it does not require the concept of a base or matrix language or of a CS-specific grammar, and means that bilingual speech can in theory be directly comparable with monolingual

speech and mainstream research on syntax. Whilst in principle the MP should therefore be universally applicable, some authors have found it lacking in its applicability to certain aspects of bilingual speech—e.g. Herring et al (in press) point out that the MP is often ineffective in explaining CS between a determiner and an NP in Welsh-English and Spanish-English data—and proponents of a constraint-based approach have suggested that it is generally inappropriate as a means of analysing CS (e.g. Jake et al 2002). Since the framework has not been successfully tested on a wide variety of CS corpora, I propose that it does not empirically satisfy **criterion 3.** Whilst some quantitative analysis of CS using the MP has been undertaken (e.g. Herring et al in press), most of the research is concerned with grammaticality judgements and elicited data (e.g. MacSwan 2000), so it may not be an appropriate framework to use when undertaking a corpus-based analysis, as I do. While studies like Herring et al demonstrate that it is possible to undertake a quantitative analysis of speech data using the MP, I suggest that the MP only partially satisfies **criterion 1.** The MP does not satisfy **criterion 2,** since it is, as with generative grammar, a framework that is concerned with speaker competence rather than production (Chomsky 1995). In the present study I am examining spoken language and its surface form rather than deriving the underlying structure of sentences, and given that the MP was not designed to be tested in this way on speech data, I suggest that it is more appropriate to use a framework which focuses on speech production for my analysis of convergence. Lastly, the MP has been used as a framework with which to analyse language change (e.g. Willis 2008, Biberauer, Newton and Sheehan in press, van Gelderen 2009), albeit not always in the context of speech containing CS; nevertheless, it is clearly possible to analyse language change via the MP, so **criterion**

4 is satisfied—in principle, it would be possible to use the MP to identify convergence in given sentences (although, as noted with respect to criterion 1, the MP might not be so suitable in quantifying that convergence).

In all, it is apparent that the MP is not an appropriate framework to use to for the purposes of the present study, for the reasons outlined above, since it only satisfies only one or two at most of the aforementioned criteria.

None of these three approaches, I suggest, are ideal frameworks with which to analyse the Welsh-English data for this study. Why, then, is the MLF model the most appropriate for the present analysis?

As studies such as Schmitt (2000), Deuchar (2006) and Smith (2006) have shown, the MLF model can be used to analyse CS data quantitatively (**criterion 1**), allowing for a clause-by-clause analysis of a dataset of bilingual speech: e.g. Deuchar (2006) demonstrated how it could be used to measure the frequency of clauses with a particular ML in Welsh-English data, while Smith (2006) used it to compare frequency of monolingual clauses versus clauses containing CS in Spanish-English data. Since it analyses bilingual speech in terms of identifying the matrix language of any given clause, it therefore seems the ideal framework for the purposes of identifying the extent to which one language or another is the predominant source of structure in a dataset. Although Myers-Scotton (2002:27) does suggest that the MLF model operates as both a production and a competence language model, most studies involving the application of the MLF model (e.g. Myers-Scotton 1993, 2002; Schmitt 2000, Smith 2006, Deuchar 2006, Deuchar and Davies 2009) analyse production data from speech corpora, rather than e.g. grammaticality judgement data, in order to identify the ways in which a bilingual's two languages interact at the 'surface' level, and perhaps its

worth is more as a production model rather than a competence model (**criterion 2**). Furthermore, it has been argued to be universally applicable in describing the constraints of CS and has been tested with success in a range of CS corpora, as described above (**criterion 3**), at least within the remit of "classic" CS (Myers-Scotton 2002:9). The second type of CS postulated by Myers-Scotton, namely "composite" CS (2002:105), aims to cover the type of CS which falls outside of the MLF model's proposed constraints and which represent some form of contact-induced change: this distinction between the two supposed types of CS is an issue which I will propose to resolve in this thesis, but for the present it is worthwhile noting that such a distinction between non-anomalous and anomalous language interaction is a very appropriate one in the present context of identifying convergence. That is, by distinguishing those clauses which are non-anomalous with those which are anomalous, clauses exhibiting language change can be flagged. This is the notion I capture with the concept of the dichotomous matrix language, described in chapter 3, and using this concept allows for convergence to be identified and measured quantitatively in a dataset (**criterion 4**).

The extent to which the four frameworks outlined above satisfy the four criteria I posited is illustrated in table 3 below. The MLF model satisfies all four criteria whereas the other frameworks do not.

	Muysken's CS typology	Poplack's CS constraints	Minimalist Program	Matrix Language Frame model
Criterion 1	✓	✓	(✓)	✓
Criterion 2	✓	✓	×	✓
Criterion 3	✓	(✓)	×	✓
Criterion 4	×	×	(✓)	✓

Table 3. Suitability of various approaches to CS for use in the present study.

While any of the CS frameworks I have outlined here could be argued to be appropriate for a certain research context, I argue that, for my aims with this analysis, the MLF model is the most appropriate, since it satisfies all four of the criteria to do with analysing CS and convergence I set out above. Having given my rationale for adopting the MLF model, I move on to discussing convergence from the perspective of this model in the next section.

4. Convergence from the MLF model's perspective

I have so far presented the MLF model as it applies to classic code-switching, where asymmetry between the participation of the two languages in a bilingual clause is maintained. Myers-Scotton (2002) has

also presented additional facets of this model which deal with constructions which do not adhere to this asymmetry. She proposes terms to describe three such phenomena: composite code-switching, a composite matrix language, and convergence (as she defines the term [2002:164]). I see all three phenomena as being very closely-connected, to the point where they appear to overlap in Myers-Scotton's descriptions of them. I will present the three below, and then discuss where they differ and where they are similar.

Myers-Scotton's (2002:105) description of the phenomenon that she calls composite CS is where (1) "the abstract morphosyntactic frame [is] derived from more than one source language" and (2) the morphemes come from the two participating languages (2002:8).[62] Thus the term describes the use of not only lexical material but also structural material from the EL, differing from classic CS which contains lexical but not structural EL material (apart from well-formed EL islands). Effect*ive*ly, the term composite CS describes those bilingual clauses which do not exhibit classic CS, which of course means that the MLF model (which is designed to describe and identify classic CS) cannot—at least in its present formulation by Myers-Scotton—describe occurrences of composite CS.

Second, Myers-Scotton describes composite CS as being present in a clause which has a composite ML (2002:22, 297). Such an ML does not source its material from just one language, but is rather a grammatical frame that uses structural material from more than one source (2002:22), so the structure of a clause which has a composite ML comes from both participating languages (2002:100). Note this

---

[62] Myers-Scotton (2002:105) also claims that composite code-switching also plays a role in long-term language attrition or shift, but this is not currently relevant, since I am primarily interested here in synchronic and not diachronic evidence of convergence in speech.

similarity to the definition of composite CS above: a clause which displays composite CS is thereby a clause with a composite ML. I interpret composite CS to be a phenomenon which mirrors classic CS, being a general concept describing CS activity in a given instance (e.g. within a dataset or within a speaker community), whereas a composite ML is the clause-level manifestation of composite CS. The clause-level manifestation of classic CS would be a ML sourced from a single language, as per the predictions of the MLF model.

The use of a composite ML by bilinguals, Myers-Scotton (2002:100) asserts, occurs when those speakers do not have full access to the morphosyntactic frame of the ML, such as in the case of imperfect second language learning. An example she gives of a clause with a composite ML is given in (8), cited from Schmitt's Russian-English data (the speaker is a bilingual child, supposedly with incomplete acquisition of Russian).

(8) odin byl pitcher
one be.3Sm.PAST pitcher
"One was a pitcher." (from Schmitt 2000:24)

The SMP would identify this clause as having a Russian ML, by looking at the inflection of the finite verb *byl*. However, if Russian were the ML, then the English morpheme *pitcher* should be assigned appropriate Russian instrumental case marking (*pitcher-om*), but in (8) it is a bare form. Myers-Scotton follows Schmitt in suggesting that this clause's ML is a composite of English and Russian, with Russian supplying some morphosyntax, but with English supplying the lexical structure of the inserted English noun *pitcher*, which is why it lacks overt case marking

(Myers-Scotton 2002:227). Note that this clause also displays composite CS.

Myers-Scotton (ibid.) also sometimes discusses a composite ML as being something beyond the individual clausal level. She interprets data by Bolonyai (1999) to show that "a composite Matrix Language is developing" in the speech of Hungarian-English bilingual children because of their mixture of English and Hungarian input during language acquisition. It appears that a composite ML here is being viewed as a generic ML choice for speakers who have multiple language inputs, rather than as a label for individual clauses.

Third, Myers-Scotton (2002:164), as I described in chapter 2 section 2, describes convergence as a process whereby morphemes from one language have structure from more than one language. Again, the resultant ML is composite. Though I use the word 'process', note that Myers-Scotton here interprets convergence as both a mechanism and an outcome. She equates the outcome of convergence with the outcome of attrition, noting that they are "perhaps identical" in this respect (ibid.).

The difference in definitions between this phenomenon and composite CS is that apparently convergence occurs in monolingual clauses which have a composite ML[63] while composite code-switching occurs in bilingual clauses which have a composite ML. Yet Myers-Scotton (2002:8, 105, 297; 2006:242) states repeatedly that composite code-switching is a form of convergence (e.g. 2002:105: "in effect, composite codeswitching necessarily entails convergence"). Conversely, Myers-Scotton states that convergence "often involve[s] codeswitching"

[63] cf. Myers-Scotton 2002:164 on the outcome of convergence: "all surface morphemes come from one language" and "the abstract lexical structure projecting these morphemes no longer comes from one language, but includes some abstract structure from another language."

(2002:165). Also, a clause which shows convergence therefore has a composite ML (ibid.): "the source of the grammatical frame ... is not a single language, but more than one."

An example Myers-Scotton gives of a clause which shows convergence is given below in (9). It is a monolingual clause.

(9) Era duro para yo meter-me
be.3S.IMP hard.SM for 1S deal-1S.INDOBJ
con esos niños ingleses
with DEM.PL.M boy-PL.M English-PL.M
"It was hard for me to deal with those English boys."
(from Morales 2000:48; cited in Myers-Scotton 2002:103)

The utterance in (9) was produced by a Spanish-English bilingual from Puerto Rico. The morphemes in the clause are all Spanish, but note how the word order of *yo meter-me* 'for me to deal' has been influenced by English pronoun positioning, and does not show the usual Spanish word order *meter-me yo*. This clause shows convergence towards English according to Myers-Scotton's definition of the term, in that the morphemes are from one language but clause structure is a composite of more than one language; it is not, however, apparently an example of composite CS, because the clause is (lexically) monolingual.

The difference, then, between composite CS, a composite ML and convergence is not clear-cut, at least not structurally,[64] since they are all terms which cover a very similar phenomenon, ostensibly from different perspectives. I interpret the terminology to indicate that a clause with a composite ML which is monolingual shows what Myers-

[64] It is possible that Myers-Scotton sees them as being different when viewed from e.g. a sociolinguistic perspective; cf. 2002:164-5.

Scotton calls convergence, while a clause with a composite ML which is bilingual shows composite CS.[65] However, to me "composite CS" and "convergence" are clearly the same process, that of structural material in a clause being sourced from more than one language. I view the signs of convergence as being visible in both monolingual and bilingual (not just, as Myers-Scotton would appear to have it, monolingual) clauses, which will have a composite ML, whilst I retain the term composite CS to describe a CS trend which is distinct from classic CS.

In fact, this dichotomy between classic and composite CS is crucial to the theoretical innovation I implement in this thesis. Myers-Scotton's formulation of the MLF model describes classic CS, where one language is the source of the morphosyntactic frame of a clause, i.e. there is asymmetry between the roles of the ML and the EL. If more than one language is the source of the morphosyntactic frame of a clause, then there is less asymmetry between the roles of the ML and EL in the production of that clause. Thus the MLF model cannot identify one language as the ML source, and so it identifies the clause as showing something which is not classic CS. To me, it is clear that a clause which does not show classic CS must *ergo* show composite CS, i.e. convergence. For a clause to not display classic CS, there must be disagreement between the language which is identified by the principles of the MLF model. In such instances where the two principles are not in agreement, then, this will be because that clause displays composite CS, yet this nuance is not made explicit in Myers-Scotton's incarnation of the MLF model.

---

[65] Though note, as discussed above, "composite ML" is also used by Myers-Scotton to describe phenomena beyond the individual clausal level (cf. Myers-Scotton 2002:227), and so I may be interpreting it in this paragraph in a more specific way than Myers-Scotton intends.

In the next section I present my innovation to the MLF model, namely the concept of a clause with a dichotomous ML, which links the MLF model with convergence, by making it able to identify and quantify data in which the ML/EL asymmetry predicted by the model does not apply.

5. The dichotomous matrix language

a. Adding to the MLF model

The principles of the MLF model, when tested on a bilingual clause which shows classic CS, are expected to indicate the same language as being the ML. In a clause where the principles do not indicate the same language as ML, logically this will be because that clause contains some structure from both participating languages, i.e. composite CS. Essentially, where the MLF model 'fails' to identify one language as being the ML source, it 'succeeds' in identifying that such a clause has a composite ML, perhaps due to convergence. Therefore, when the two principles of the MLF model do not 'agree' for a given clause, that clause can be identified as possibly showing convergence.

b. Defining a dichotomous ML

I suggest that the MLF model can identify that the ML of that clause is a composite of structures from both participating languages. When the MLF model is tested on a clause, if both principles point to language A,

then language A is the ML. If both principles point to language B, then language B is the ML. If both principles do not indicate the same language, then I propose that the clause has a **dichotomous matrix language.**[66] I define this concept formally below.

A clause with a dichotomous ML

In a monolingual or bilingual simple clause produced by a bilingual speaker, if, following the Matrix Language Frame model, the Morpheme Order Principle and the System Morpheme Principle do not indicate the same language as the sole source of the matrix language of that clause, that clause is identified as having a dichotomous matrix language.

The term 'dichotomous' is used because it indicates a dichotomy in the evidence presented by both principles for the ML of a clause: they point in two different directions. Not explicit in the definition, but implicit in its formulation, is that the two principles of the MLF model can disagree by either indicating entirely different languages as being the ML source, or where one or both principles have conflicting indicators for satisfying that principle (e.g. the SMP points to language A but the MOP points to both language A and language B).

The innovations I introduce with this concept are twofold:

1) A clause with a dichotomous ML can be monolingual or bilingual. This covers the notions of composite ML, Myers-Scotton's interpretation of convergence and composite ML in one phenomenon.

---

[66] I validate my reasons for not using Myers-Scotton's term 'composite ML' for this below.

2) A dichotomous ML clause is identifiable using the existing principles of the MLF model without affecting them or their ability to identify classic CS, and the model can now be used to explicitly identify clauses with an ML which is a composite. That is, the dichotomous ML concept links the MLF model with the idea of convergence in a way that was hitherto not explicitly present.[67]

It follows that a clause with a dichotomous ML may show convergence (in the sense that I have presented it in chapter 2 and elsewhere rather than in Myers-Scotton's narrower definition), since clauses showing convergence will have structure sourced from more than one language. Because of this the concept may be used to isolate dichotomous ML clauses in a dataset, which can then be analysed further for the presence of convergence.

I argue that the dichotomous ML is a more concrete and narrowly-defined notion than Myers-Scotton's term "composite ML". First, Myers-Scotton seems to pair speech which exhibits composite MLs with incomplete language acquisition, whereas a dichotomous ML has no such connotations. Second, in my opinion Myers-Scotton uses the term "composite ML" confusingly and in contradictory fashion, while a dichotomous ML is a very specific label which is clause-specific (rather than a term which, say, could be used to describe a whole discourse or

---

[67] In her application of the MLF model to Welsh-English data, Deuchar (2006) has an option for clause ML of 'neither', which she uses when there is a conflict between which language is identified as the source of clause word-order when testing the MOP. She posits that such a clause has an "unidentifiable" ML (2006:1999, 2001). This touches on the same ideas as the dichotomous ML, in that what I would classify as dichotomous ML clauses might be classified by Deuchar as unidentifiable using her method, but note that the dichotomous ML concept links the MLF model's principles with the idea of a composite ML, which Deuchar's 'neither' option does not (although she does suggest that convergence occurs in 'neither ML' clauses).

speech community). Third, Myers-Scotton does not associate the composite ML with the parameters of the MLF model, since the MLF describes classic CS but a composite ML is associated with composite CS. A dichotomous ML, meanwhile, is a natural extension to the MLF model, uses the existing parameters of that model (i.e. the two Principles) and makes a logical link between classic and composite CS. While ostensibly a clause with a dichotomous ML might be described as having a composite ML (by Myers-Scotton), the opposite does not necessarily follow. In light of this, I will not be using the term "composite ML" from this point on in this study.

c. Examples of clauses with a dichotomous matrix language

I will now look at the existing code-switching literature and discuss some clauses which other studies have highlighted as showing the structural result of language contact. I will apply the MLF model to these clauses and show how they can be reinterpreted as having a dichotomous ML. The interpretations below, unless noted otherwise, are mine.

The example in (10) below is from a Spanish-English bilingual, and is a monolingual clause.

(10)	yo	quiero		manzana	jugo
	1S.NOM	want.1S.PRES		apple	juice

"I want apple juice."

(taken from Myers-Scotton 2002:166)

The SMP identifies Spanish as the source of the outside late system morphemes in this clause, in the morphological inflection of *quiero*. The position of the modifier in relation to the head in the NP *manzana jugo*, however, is the English word order of the modifier preceding the head, rather than the Spanish order of the modifier following the head. The MOP identifies English as providing the word order for this NP (the rest of the word order in the clause could come from either English or Spanish). Since the SMP and the MOP point to different languages as being the source of the morphosyntax of this clause, no one language is identified as being the clause ML. The clause has a dichotomous ML.

The example in (11) below was produced by a Hungarian-English bilingual child.

(11) játsz-ok school-ot
play-1S.PRES.SUBJ.CONJ school-ACC
"I'm playing school." (taken from Bolonyai 1998:34)

This bilingual clause has Hungarian morphology on the verb *játsz-* and the English noun *school*, so the SMP identifies Hungarian as supplying outside late system morphemes. The word order of the finite verb *játszok* and the object *schoolot*, however, is English VO order rather than Hungarian, in which the object should be in (initial) Topic position (Bolonyai 1998:34). The MOP thus identifies English as the source of word order in this clause, and since this disagrees with the language identified by the SMP as supplying morphosyntax, I label the clause as having a dichotomous ML.[68]

[68] cf. Deuchar (2006:2008), who describes this clause as having "an 'in-between' or composite matrix language, which combines aspects of both English and Hungarian".

I now provide some examples of clauses produced by Welsh-English bilinguals which can be identified as having a dichotomous ML. Deuchar (2006) cites utterances which she suggests show word-order convergence, like (13) below.

(13) fi 'di bod i 'r bus lle
1S PRT.PAST be.NONFIN to DET bus place
"I've been to the bus place [? = bus stop]."

(from Deuchar 2006:1996)

This is a bilingual clause where all but one morpheme (English *bus*) is Welsh. The order found in the noun-noun NP *bus lle* 'bus stop' is English-predominant MH[69] order. The expected order for these morphemes if Welsh were providing clause structure would be *lle bus*, with the modifier *bus* following the head *lle* (HM), but instead the modifier is found preceding the head (MH). If a null auxiliary verb (e.g. *wyf*) is assumed to have been deleted from this clause,[70] then the clause can be interpreted as having Welsh outside late system morphemes, which would indicate, by applying the SMP, that Welsh is the ML. However, when the MOP is applied, the word order of the NP *bus lle* points to English being the ML. Since the two principles point to different languages as supplying structure, I consider this clause to have a dichotomous ML.[71]

---

[69] I use the following abbreviations: MH indicates an NP consisting of a head and modifier where the modifier precedes the head, whereas HM inidcates a head/modifier NP where the modifier follows the head. Note that when I use the term 'head/modifier order', we mean the relative order of head and modifier within an NP, without specifying what this order is.

[70] The undeleted form would presumably be ***wyf fi 'di bod i'r bus lle***.

[71] As I explain below, dichotomous ML clauses possibly show convergence. One argument I propose for the clause in (13) showing word-order convergence is that *bus lle* uses a word

Deuchar (2006) cites another example of a clause with unusual word order, given below as (14).

(14)	Ddaru	ni	gyfweld	[...]	am	ddeg	awr
	happen.PAST	1PL	interview.NONFIN		for	ten	hour
	<u>assistant</u>	i	*D.S.*				
	assistant	for	D.S.				

"We interviewed [...] for a ten-hour assistant for D.S."

(from Deuchar 2006:1994)

It is not possible to test the SMP on this clause, as subject-verb agreement does not apply: the verb *ddaru* is not inflected for person or number. When the MOP is tested, the VS order of the verb *ddaru* and the subject *ni* indicates that the word order is sourced from Welsh. However, the head/modifier NP *ddeg awr assistant* 'ten-hour assistant' appears to have English MH word order, as the modifying element *ddeg awr* precedes the head *assistant*. This order is not usual for Welsh. As the material within the NP is not all English, it cannot be labelled an EL island. The English word *assistant* would be expected to be positioned in a locus appropriate to the subject/verb order (Welsh), preceding the modifier, as in *assistant deg awr*. As we find that English seems to provide the word order in this NP, however, the result is that the verb/subject order and head/modifier order indicate different languages

---

order more prevalent in English than it is in Welsh, but present in both. An additional argument is, as Deuchar points out, that the morphemes in the whole clause appear to have a one-to-one morpheme correspondence with English surface word-order, where *fi 'di bod i* is equivalent to *I have been to*. If this is the case, then the word order for the whole clause, and not just the head/modifier NP, is converging from Welsh to English. See chapter 6 for further discussion of clauses with auxiliary deletion and the possibility of them showing convergence toward SV word order.

as being the ML. Since the MOP has conflicting information about which language sources morpheme order, I would to classify this clause as having a dichotomous ML.[72]

I have shown above how introducing the concept of a dichotomous ML to the MLF model allows the model to be applied to clauses produced by bilinguals to show that such clauses show signs of a contact-induced structural process. I next discuss to what extent dichotomous ML clauses show the signs of convergence.

d. The connection between a dichotomous matrix language and convergence

The concept of a dichotomous ML encompasses Myers-Scotton's concepts of convergence and composite CS. It is also compatible with the definition of convergence I provided in chapter 2 section 2, where I defined it as the use of morphemes from language A with structure which is available in both language A and language B, but is more

---

[72] It is worth discussing briefly whether or not the word order found in this example represents a system-preserving or system-altering process, since I have argued in chapter 2 that convergence is a system-preserving process. Here, the modifying element *ddeg awr* precedes *assistant* in word order which is unexpected for Welsh, but would be expected in English if the morphemes were English. However, the word order of a modifying element preceding its head does exist in Welsh, in compounds and with certain adjectives, as already described. Therefore, the word order in *ddeg awr assistant* does not represent a 'new' word order in Welsh, but rather the extension (= redistribution), due to the influence of this MH word order in English, of an existing Welsh word order to morphemes which would not be expected to have this word order in conventional Welsh grammar. The process here is system-preserving, then, not system-altering, because it is a redistribution of existing Welsh patterns rather than the introduction of an entirely innovative, and thus system-altering, construction from English.

prominent in language B. I suggest that the MLF model can be used to identify clauses which have undergone that process; I explain how here.

For a clause to have a dichotomous ML, it must have some morphemes from language A with structure from language B; cf. the Welsh morphemes *ddeg awr* occurring in English-predominant MH word order in example (14) above. When a construction like word order is said to be 'from language B', this indicates word order that is not expected to be used for the morphemes found in that particular structure. The language origin of that word order must be ascertained on a clause-by-clause basis, because it might be word order which is also available in language A, or it might be word order which is only usually available in language B. If the construction under question is available in language B and is also available in language A, but not in this context, then convergence may be postulated as an explanation of why that word-order is found. The prominence of that construction in language B has influenced the speaker to extend its use for morphemes from language A.

However, if the construction under question is available in language B but not usually in language A, then I argue that convergence should not be the interpretation. Rather, this is more likely to be the transfer into language A of a construction from language B which is not ordinarily present in language A. Despite such a phenomenon not being convergence, it can still occur in a clause which has a dichotomous ML. The structure found in a dichotomous ML clause can either show convergence or not, depending on individual circumstances.

In order to identify the results of which process is found in any given dichotomous ML clause, the analyst must ask whether the construction in question is available in both language A and B, and if so, whether that construction is more prominent in this context in language

B than A. If the answer to both these questions is affirmative, then the analyst can identify that clause as showing convergence. If it is not, then a different account of the construction is necessary. Whilst the concept of a dichotomous ML clause can identify clauses which contain some structural anomaly, and which probably show convergence, they do not <u>necessarily</u> show convergence.

In this section I have explained my innovative concept of a dichotomous ML. This concept can be used to identify clauses which may show convergence. In chapter 5 I will demonstrate how I applied this concept to a set of Welsh-English data in order to measure the extent of word-order convergence therein. First, in section 6 below, I will explain how I apply the model to data.

6. Methodology of applying the model to data

In this section I discuss various methodological aspects of my data analysis; methodology of data collection will be outlined in chapter 4, along with the extralinguistic aspects of the participants, while a description of the data themselves will be given in chapter 5. Below, I first explain my system of clause division (since my analysis will be done clause by clause), before explaining the aspects of Welsh and English which are being used as a means of identifying the ML in clauses produced by bilinguals in those languages. I also outline the word order differences between Welsh and English, because, as noted in chapter 2, it is expected that convergence is most likely to occur in parts of the two grammars which are similar. Finally I describe the practical application

of the MLF model to clauses, and the individual 'options' for the ML of any given clause.

a. The clause as the unit of analysis

As noted above in section 3, Myers-Scotton uses the CP (Projection of Complementizer) as her unit of analysis for the MLF model. I interpret this to mean a clause, and therefore, for the purposes of applying the model to my data, those data must be separated into discrete units to be analysed.

Generally speaking, data will be subdivided into separate clauses. Simple clauses such as (15) below are unproblematic, and constitute one unit of analysis. In these invented examples I label the clausal units of analysis with square brackets.

(15)	[wnaeth	Jac	daro	'r	bêl	griced]
	do.3S.PAST	Jac	hit.NONFIN	DET	ball	cricket
	"Jac hit the cricket ball."					

In (15) the verb *wnaeth* is a finite auxiliary in a construction which has the nonfinite verbal noun *taro*[73] 'hit' as the main verb.

Complex clauses need to be subdivided into their constituent clauses, with each individual clause within the complex clause counting as one unit of analysis. In (16), the subordinate clause *taro peli criced* is one unit of analysis and the residual main clause *oedd Jac yn hoffi* is another.

[73] *Taro* in the example has been changed to *daro* through soft mutation.

(16)	[oedd	Jac	yn	hoffi	[taro	peli
	be.3S.IMP	Jac	PRT	like.NONFIN	hit.NONFIN	balls
	criced]]					
	cricket					

"Jac liked hitting cricket balls."

Such embedded clauses can be nonfinite, as in (16), or finite, as in (17), where *tarodd Jac y bêl* is one unit of analysis and *y taflodd Jên ato* is another (a relative clause modifying the object *bêl*).

(17)	[tarodd	Jac	y	bêl	[y	taflodd	Jên
	hit.3S.IMP	Jac	DET	ball	PRT	throw.3S.IMP	Jên
	ato]]						
	to.3Sm						

"Jac hit the ball that Jên threw at him."

Complex clauses can consist, hypothetically, of any number of embedded clauses, both finite and nonfinite. Such clauses are subdivided into individual embedded clauses where possible, with the residual main clause also being a unit of analysis. In cases where that residual main clause itself is a complex clause with embedded clauses, as in (18), maximum subdivision is again aimed for.

(18) [oedd Jac a Jên [oedd yn ffrind
be.3S.PAST Jac and Jên be.3S.IMP PRT friend
penna iddo] yn chwarae criced [pan
best to.3Sm PRT play.NONFIN cricket when
oedden nhw 'n tyfu fyny]]
be.3PL.IMP 3PL PRT grow.NONFIN up
"Jac and Jên, who was his best friend, played cricket when they were growing up."

In (18) there are three units of analysis: (1) *oedd Jac a Jên ... yn chwarae criced*,[74] (2) *oedd yn ffrind penna iddo*, and (3) *pan oedden nhw'n tyfu fyny*.

With regard to the MLF model, it is expected, following Myers-Scotton, that each individual clause which is a unit of analysis will have its own ML (even if that ML cannot be identified by the overt morphemes), and it is possible that the ML can 'change' within a complex clause, in that any embedded clauses can have a different language as the source of the ML from that of their main clause. In (19) below the embedded clause *y taflodd Jên ato* has Welsh ML, while the main clause *Jac hit the ball* has English ML. Since both clauses are discrete units of analysis, the same language is not expected to be the source of the ML for both, although there would also be no reason why both a main clause and its embedded clause(s) should not have the same language identified as their respective ML.

[74] As indicated, clausal units for analysis can sometimes be discontinuous.

(19) [Jac hit the ball [y taflodd Jên ato]]

Jac hit the ball PRT throw.3S.PAST Jên to.3Sm

"Jac hit the ball that Jên threw to him."

In chapter 2 section 7 I discussed linguality, or the classification of clauses as monolingual or bilingual, and the relationship between language mode and CS and language change. As far as ascertaining the linguality of complex clauses is concerned I only count the linguality of the individual clause that is the unit of analysis, and interpret it as being bilingual if it contains morphemes from both languages. I do not consider it relevant, for the purposes of analysis in terms of linguality, if a complex clause has a main clause which is monolingual in one language but an embedded clause that is monolingual in another language, because I am counting individual clauses rather than complex clauses. The complex clause in (20), then, would be subdivided into two clauses for analysis: (1) *tarodd Jac y ball* is one bilingual clause and (2) *y taflodd Jên ato* is one monolingual clause.

(20) [tarodd Jac y ball [y taflodd Jên ato]]

hit.3S.IMP Jac DET ball PRT throw.3S.IMP Jên to.3Sm

"Jac hit the ball that Jên threw at him."

In (21) the residual main clause is monolingual while the embedded clause is bilingual, and would be analysed as two separate units.

(21)	[tarodd	Jac	y	bêl	[y	taflodd	Jên
	hit.3S.IMP	Jac	DET	ball	PRT	throw.3S.IMP	Jên
	to ___ him]]						
	to	him					

"Jac hit the ball that Jên threw to him."

Meanwhile, (22) would be divided into two bilingual clauses (both with Welsh ML).

(22)	[tarodd	Jac	y	ball	[y	taflodd	Jên
	hit.3S.IMP	Jac	DET	ball	PRT	throw.3S.IMP	Jên
	to ___ him]]						
	to	him					

"Jac hit the ball that Jên threw to him."

Since the data I analyse is natural data, they contain many clauses which are incomplete fragments, and also feature instances of repetition or reformulation of clause material, both of which render clause division more complex. Fragments, which I define as clauses which do not have a complete surface structure (e.g. a fragment may have a finite verb but no subject/object, or it may have no finite verb), are considered incomplete clauses, but they are nevertheless counted as individual clause units. Clauses where content is repeated with reformulation, meanwhile, are treated by ignoring the fragmental parts of the utterance insofar as the reformulated parts of the utterance themselves form coherent clauses. So, (23) below would be analysed as

two clauses, because *dw i'n mynd* is not a fragment, whereas the first *dw* in (24) is only a finite verb on its own if separated from the rest of the utterance, and so is a fragment. In such cases as (24), the redundant sections are ignored, so (24) only consists of one clause, *dw i'n mynd i'r dre*.

(23)	[dw	i	'n	mynd]	[///]	[mae	o
	be.1S.PRES	1S	PRT	go.NONFIN		be.3S.PRES	3Sm
	'n	mynd	i	'r	dre]		
	PRT	go.NONFIN	to	DET	town		
	"I'm going... He's going to the town."						

(24)	dw	[/]	[dw	i	'n	mynd	i
	be.1S.PRES		be.1S.PRES	1S	PRT	go.NONFIN	to
	'r	dre]					
	DET	town					
	"[I]'m... I'm going to the town."						

To reiterate, utterances in the data are subdivided, for the purposes of analysis, into 'simple' clauses as far as this is possible.

b. Identifying outside late system morphemes in Welsh and English

In order to apply the MLF model to Welsh-English data, I must first identify the outside late system morphemes which are present in both languages. As stated above in section 3.a.i, outside late system

morphemes have grammatical relations external to their head constituent. It appears that the only such morphemes in both Welsh and English are those of finite verbs, where, in a simple main clause, for example, they are expected to match the features of the subject constituent for person and number.[75] See the clause in (25) below.

(25)	oedd	y	dyn	yn	cario	flower basket
	be.3S.IMP	DET	man	PRT	carry.NONFIN	flower basket

"The man was carrying a flower basket."

Here the subject *dyn* is a singular masculine noun, and the finite verb *oedd* matches this for person. This verb is from Welsh, and so Welsh is the source of the outside late system morpheme in this clause.

Note that Welsh does have inflected prepositions, such as *iddi* 'to (her/it)' in (26) below, the base form of which preposition is *i* 'to'. Such prepositions can only occur with object pronouns, and have inflection which show agreement with the number, person and gender of the pronominal object, which is usually overt in spoken Welsh (Borsley, Tallerman and Willis 2007:199–200). In (26) *iddi* shows 3rd person singular feminine agreement with the object pronoun *hi*.

(26)	rhoddodd	y	dyn	flodau	iddi	hi
	give.3S.IMP	DET	man	flowers	to.3Sf	3Sf

"The man gave flowers to her"

Since such a preposition depends on the properties of the complement NP to determine its inflectional morphology, it might be considered that

[75] Though note, for noun subjects only, that Welsh has anti-agreement with number, so a plural subject would require a singular and not plural finite verb (see King 1996:137).

such prepositions are outside late system morphemes. However, they appear to only require to 'look within' their PP (which in this case contains the complement NP) to obtain grammatical information, and do not have to 'look outside' this projection at morphemes in, for example, the main VP which contains the PP. Since inflected prepositions appear in the same surface level maximal projection as their complements (within the PP), I suggest that they are in fact early system morphemes (cf. Myers-Scotton 2002:75). As early system morphemes they are not useful as clear indicators of clause ML.

I thus look at finite verb morphology to help identify the ML for a Welsh-English clause from the viewpoint of the SMP, on the assumption that this morphology (outside late system morphemes) must come from the ML if the clause contains classic CS.

c. Identifying word-order similarities between English and Welsh

Unlike outside late system morphemes, which (at least in Welsh and English) point definitively to being sourced from one language only, word order is sometimes similar in English and Welsh. Indeed, as I noted in chapter 2, convergence occurs in parts of the grammar which are similar in both languages in contact. Following Thomason's (2001:93) suggestion that an essential step in identifying contact-induced structural change (not only convergence) in a language is to identify those structural features shared by both languages, I now describe the word order similarities found in Welsh and English.

The specific word order under analysis is the relative positions of the subject and the finite verb in a clause and the position of the head and the modifier within an NP.[76] I discuss below both word orders in turn.

i. Subject and finite verb

English is an SV-order language in both main and subordinate clauses, with the verb occupying post-subject position,[77] as shown in (27).

(27)	The cat	caught	the mouse
	S	**V**	**O**

Some English constructions, however, are not SV. 'Yes-No' interrogative constructions are usually formed with an inflected auxiliary, as shown in (28), with the exception of constructions using the main verb 'to be' where the inflected verb itself is in clause-initial position, as shown in (29). Both constructions result in resulting in VS word order (the finite verb precedes the subject).

(28)	Did	the cat	catch	the mouse?
	Aux	**S**	**V**	**O**

[76] It should be noted that when referring to head/modifier NPs I mean only noun-plus-adjective or noun-plus-noun constructions. I do not include in my description, for example, relative clause constructions like *the boy bored with learning*, as, in this instance, the relative clause always appears in the same position—following the head NP—in both English and Welsh, and is therefore not a useful way of differentiating between Welsh and English ML.

[77] See Quirk et al (1989:724).

(29) Is the mouse here?
V S

WH-questions using the main verb 'to be' are again formed with a verb preceding the subject, as in (30), again resulting in VS word order.

(30) Where is the mouse?
WH V S

Thus English has SV word order in certain constructions and VS in others.

Welsh, conversely, is primarily a VS language (the finite verb occupies the first main position[78]). VPs can either be formed using a synthetic main verb, as in (31), or a periphrastic construction using an auxiliary verb, shown in (32), with the meaning the same for both constructions. Note that both constructions have an initial finite verb and are both declarative.

(31) Daliodd y gath y llygoden
Caught the cat the mouse
V S O
"The cat caught the mouse"

(32) Wnaeth y gath ddal y llygoden
Did the cat catch the mouse
Aux S V O
"The cat caught the mouse"

[78] See King (1996:21).

Questions in Welsh have rising final intonation and, occasionally, a pre-verbal particle *a* which causes soft mutation of the initial consonant of the finite verb. The position of the subject and verb are not changed, however, so Welsh questions, like Welsh declaratives, have VS order. Examples are shown in (33) and (34) below, and note again that both synthetic and periphrastic constructions are verb-initial constructions, like (31) and (32) above.

(33) Ddaliodd y gath y llygoden?
Caught the cat the mouse
V S O
"Did the cat catch the mouse?"

(34) Wnaeth y gath ddal y llygoden?
Did the cat catch the mouse
Aux S V O
"Did the cat catch the mouse?"

WH-questions, however, where the subject is replaced by a WH morpheme and appears in the first position, have SV word order; see (35) below.

(35) Pwy ddaliodd y llygoden?
Who caught the mouse
WH V O
"Who caught the mouse?"[79]

[79] This form could also mean "Whom did the mouse catch?"

Furthermore, Welsh usually marks clauses for emphasis by having the emphasised constituent occurring in the initial position; if this is the subject, the resultant clause has SV order. This is again the case whether the construction is synthetic, as in (36), or periphrastic, as in (37).

(36)	Y gath	ddaliodd	y llygoden
	The cat	*caught*	*the mouse*
	S	**V**	**O**

"It was the cat which caught the mouse"[80]

(37)	Y gath	wnaeth	ddal	y llygoden
	The cat	*did*	*catch*	*the mouse*
	S	**Aux**	**V**	**O**

"It was the cat which caught the mouse"

Note that WH-question word order (as in [35] above) could be considered another example of a clause-initial subject being used for emphasis. The most prevalent word order in both declaratives and in questions in Welsh is VS, with SV word order sometimes used.

Thus, whilst the normal declarative word order of Welsh is different from English (VS versus SV), Welsh has SV word order in some constructions, while English has VS word order in some constructions. Given, then, that both word orders are available in both languages, convergence from SV to VS in English or VS to SV in Welsh would be a possibility.

---

[80] Note that (38) could also mean "It was the cat whom the mouse caught" (i.e. OVS order).

ii. Head and modifier in NPs

I now briefly outline the main differences and similarities in the word order of the head and the modifier in head/modifier NPs in English and Welsh. NPs of this kind include adjective/noun and noun/noun collocations, and can contain any number of morphemes as long as there is a head/modifier relationship.

In English head/modifier constructions, the order is for the modifier to precede the head (MH) in most cases, as seen in (38), an adjective-noun construction, and (39), a noun-noun construction.

(38) red wine
Adj N
Mod Head

(39) bus driver
N N
Mod Head

There are a small number of English collocations where the order is HM, e.g. *court martial*, *queen consort*, *Attorney General*, and *sergeant-major*, but such word order is very rare.[81] English predominantly has the

[81] Moreover, some of these are borrowings from Romance languages (e.g. *sergeant-major* < French *sergent-major*, first attested in English in 1574 according to the Oxford English Dictionary), and others are anomalous, such as *court martial*, which is first-attested as *martial court* (normal English MH order) but apparently obtained the new HM order in the mid-17th century according to the OED, perhaps due to conflation with the concurrent term *court marshal*. Note, however, that occasionally modern usage produces phrases such as *Tate Britain* (the museum name) or, in the 2008 Olympics, *Team GB* (Great Britain). Perhaps this is done deliberately for effect, to emphasise the head rather than the modifier. Nevertheless, it is still rare and probably not productive in actual speech.

same order—MH—in its head/modifier constructions, with very few exceptions.

In Welsh the most common order in such head/modifier constructions is HM, as seen in (40), a noun-adjective construction, and (41), a noun-noun construction.

(40) gwin coch

wine *red*

N **Adj**

Head **Mod**

"red wine"

(41) gyrrwr bws

driver *bus*

N **N**

Head **Mod**

"bus driver"

Some modifying adjectives in Welsh, however, precede the head, e.g. *hen* 'old' and *prif* 'main, prime'; examples are given in (42) and (43), where the order is MH.[82]

[82] Note that adjectives such as *hen*, however, can also be found following the head without an apparent change in meaning (e.g. *dyn hen* 'old man'). For example, in the corpus from which the data for this thesis is taken, the following clause is found:

Mae	rai	hen	gymaint	gwell	na	rai	newydd
be.3S.PRES	ones	old	so-much	better	than	ones	new

"The old ones are so much better than the new ones." [DAVIES13-MEI1027]

Note how here *hen* 'old' follows the head *rai* 'ones', so this NP has the HM word order *rai hen*, rather than having the MH order *hen rai* (which would also nevertheless be possible).

(42) hen ddyn

old *man*

Adj **N**

Mod **Head**

"old man"

(43) prif fynedfa

main *entrance*

Adj **N**

Mod **Head**

"main entrance"

With certain exceptions (such as *cyntaf* 'first', *olaf* 'last'), both ordinal and cardinal numerals in Welsh also precede the head they modify, resulting in MH order, as shown in (44) below.

(44) Ail ddiwrnod

Second *day*

Num **N**

Mod **Head**

"Second day"

Furthermore, any Welsh modifier can presumably precede its head, causing soft mutation of the initial consonant of the head. Whilst this is primarily a literary style, it also occurs (albeit much less frequently) in colloquial use, and often results in the construction being interpreted as a compound. In fact, compounds in Welsh are usually MH order constructions where the initial consonant of the head morpheme is

mutated.[83] Some examples of compounds are given in (45) and (46) below.

(45) cryno - ddisg

(cryno + disg)

compact *disc*

Adj **N**

Mod **Head**

"compact disc"

(46) heddwas

(hedd + gwas)

peace *servant*

N **N**

Mod **Head**

"policeman"

Note that the first consonant of the head nouns *disg* 'disk' and *gwas* 'servant' undergo soft mutation of [d] > [ð] and [g] to [ø], because these morphemes are preceded by the modifying element (*cryno* and *hedd* respectively). Such constructions as these, however, are not likely to be productive, in the sense that the modifiers used, *cryno* and *hedd,* probably do not usually occur in the pre-head position outside these specific contexts.

---

[83] Mutation is a morphophonemic process common in the Celtic languages whereby, in Welsh, initial consonants are changed depending on the linguistic context. Perhaps a difference between NPs with MH word order which are compounds and those which are not compounds is based on productivity: e.g. compound elements are not productive whereas other morphemes are.

Welsh, therefore, has more than one option for word order within head-modifier constructions, primarily HM but also MH, whilst English, with a few exceptions, has only one option, MH. On the basis of the definition of convergence given in chapter 2 section 2, convergence from Welsh HM order to MH would be predictable, under the influence of English MH order and based on an extension of the existing but less prominent Welsh MH order. On the other hand, convergence in English of MH to HM order would not be predictable.

One consideration that must be addressed is the usefulness of the word order within a head/modifier NP as an indicator of ML. I will discuss these types of head/modifier NP which can be used to identify an ML in more detail in chapter 5, section 4.a, but I will address them briefly here.

There are some types of head/modifier NPs whose word order can be used as a criterion for identifying the ML. First, if the word order within a monolingual NP is not that expected for those morphemes, then it is likely that it is the other language providing word order for that construction. Consider the phrase in (47) below.

(47) <u>teacher</u> <u>brilliant</u>

H M

'brilliant teacher'

Here both morphemes are English, but the word order is not that expected for English. Rather, it is HM, which is that expected for Welsh. Regardless of whatever other morphological and grammatical material exist in the clause in which this NP might occur, then, it is identifiable from this NP alone that Welsh is providing its structure, and that Welsh

is the ML of this 'clause'. Thus, such head/modifier NPs as this one will be used to identify the ML in clauses in the data.

Second, if a head/modifier NP is bilingual (i.e. has morphemes from both Welsh and English), then it cannot be interpreted as being an EL island, and thus it will be clear from looking only at this NP, without resorting to material in the rest of the clause for information, what the language providing the structure for those morphemes is. Consider (48) below.

(48) athro brilliant
teacher brilliant
H M
"brilliant teacher"

This is not a monolingual NP and so it cannot be interpreted as being an well-formed island of morphemes from the EL; clearly, there are morphemes from both the ML and the EL in this NP. In any given clause, word order outside EL islands must be provided by the ML, and so since *athro brilliant* is not an EL island, its word order must be being sourced from the ML. The word order in (48) is HM, with *brilliant* following and modifying *athro*, which is predominantly a Welsh word order. This means I identify Welsh as the ML of this construction. Bilingual head/modifier NPs will also be used as indicator of the ML in clausal units in these data.

There are other types of head/modifier NP, however, whose word order is not useful as an indicator of ML. As described above, Embedded Language islands are multi-word strings from the EL which maintain the word order and grammar of the EL. One example of an EL island is a monolingual head/modifier NP where the word order of the morphemes

is appropriate for the language of those morphemes. Naturally, a head/modifier NP with morphemes and word order sourced from the ML, rather than the EL, would also have word order appropriate for its morphemes. So, any given monolingual head/modifier NP with word order appropriate for the language of the morphemes therein could be identified as sourced from the ML or the EL, depending on the rest of the material in that clause. An example is given in (49) below.

(49) brilliant teacher

M H

Here an English adjective and an English noun are in MH word order, which is appropriate for English. This NP could occur in a clause which has English ML, in which case the NP word order is sourced from the ML, or it could occur in a clause which has another language as ML, in which case the NP word order is sourced from the EL, and this NP is treated as an EL island. I infer from this that head/modifier NPs such as this are not particularly useful as an indicator of which language is the ML for clauses that contain them, because there is insufficient information in the NP itself to indicate whether its word order is sourced from the ML or EL.

I have shown here how Welsh and English largely differ in terms of predominant word order, but that, crucially, there are similarities between the two systems. I propose that where such similar constructions, but with dissimilar distribution, are found, convergence may occur.

d. The options for identifying the matrix language of a clause

In section 3 of this chapter I described the MLF model and its governing principles. Here, and to conclude this section, I show how the model allows one to identify the ML of a given clause in Welsh-English data, assuming the structural properties of Welsh and English as noted in section 6.c above.

Note that the following classifications are based on my interpretation of the MLF model (albeit also based in part on the interpretation found in Deuchar 2006[84]). For the analysis, in any given clause one of the following classifications must be noted as the ML: Welsh, English, dichotomous or unidentifiable. No clause has 'no ML', although many have an ML which cannot be identified for various reasons, as explained below.

i. Welsh ML

If the MOP and the SMP both indicate Welsh, then the ML is Welsh. That is, the following preconditions apply:

- Morpheme order of the clause (relative position of subject and finite verb and relative position of head and modifier in NPs) are

[84] Deuchar (2006) used similar criteria to satisfy the principles of the MLF when testing her Welsh-English data, i.e. subject-verb agreement and the word order of subject/verb and head/modifier in an NP. Her options for ML were Welsh, English, Either (where word order was compatiable with both languages) or Neither (where word order is not compatible with either language).

compliant with Welsh and represent the word order which is expected in Welsh. If the morpheme order is available either in Welsh or in English for the construction in question, the MOP registers an ambiguous word order ('either'), and the ML is decided conclusively (if possible) from the indication of the SMP.

- The outside late system morphemes (the language providing the morphology of the finite verb, which is expected to agree with the subject in terms of grammatical features, though not necessarily with the language of the subject morpheme) come from Welsh.

1. Examples[85]

Example (50) below is of a monolingual clause consisting entirely of Welsh morphemes.

(50)	Dw	i	'n	darllen	llyfr	hir
	be.1S.PRES	1S	PRT	read.NONFIN	book	long
	"I'm reading a long book."					

Testing the MOP on this clause indicates that the finite verb *dw* precedes the subject *i*, and also that, in the head/modifier NP *llyfr hir* 'long book', the modifying adjective *hir* 'long' comes after the head noun *llyfr* 'book'. These word orders are both expected in Welsh. Thus the MOP identifies Welsh as the ML of this clause. The outside late morphemes in the clause are identified here as the morphological agreement of the finite

[85] The examples I give in this section do not come from actual data, but rather are invented for illustrative purposes.

auxiliary *dw*. This verbal morphology is Welsh, and so the SMP identifies Welsh as being the ML of this clause. Since both principles point to Welsh, this conclusively identifies Welsh as the ML. Consider the bilingual clause example in (51).

(51) Dw i 'n darllen long book
be.1S.PRES 1S PRT read.NONFIN long book
""I'm reading a long book."

The SMP identifies Welsh as supplying the outside late system morphemes of *dw* again. Also, here the word order of the finite verb *dw* and the subject *i* are the same as in (50), again identifying Welsh. But the position of the English morphemes in the NP *long book* are not in a word order compliant with Welsh: the modifier *long* precedes the head *book*, which is an English-predominant word order. But since both the morphemes in the NP are English, this constituent is analysable as an EL island, where switched morphemes from one language can occur in structure from that language in a clause which otherwise has structure from the other language. In (51), then, the ML is Welsh, as identified by the subject/verb order and verbal morphology, with an EL island NP *long book*, which does not affect the identification of the ML as Welsh. In the bilingual clause in (52) the ML is also Welsh.

(52) Dw i 'n darllen book long
be.1S.PRES 1S PRT read.NONFIN book long
""I'm reading a long book."

Here in (52), as in (51), the finite verb *dw* precedes the subject *i* and indicates Welsh as ML, as does the morphology of the verb *oedd*. The

difference between (52) and (51) is that in (52) the English morphemes *book long* are in a Welsh-predominant word-order rather than English-predominant word-order, i.e. the modifier *long* follows the head *book*. *Book long* is therefore not identified as an EL island but as two individual English morphemes positioned in Welsh word-order. This reinforces the identification of the clause ML as Welsh. Welsh is the source of word order in this clause, for both Welsh and English morphemes.

ii. English ML

If the MOP and the SMP both indicate English, then the ML is English. That is, the following preconditions apply:

- Morpheme order of the clause (relative position of subject and finite verb and relative position of head and modifier in NPs) is compliant with English and represent word order which is expected in English. If the morpheme order is available either in English or in Welsh for the construction in question, the MOP registers an ambiguous word order ('either'), and the ML is decided conclusively from the indication of the SMP.
- The outside late system morphemes (the language providing the morphology of the finite verb, which is expected to agree with the subject in terms of grammatical features, though not necessarily with the language of the subject morpheme) come from English.

1. Examples

Example (53) below is of a monolingual clause consisting of only English morphemes.

(53) I am reading a long book
1S be.1S.PRES reading DET long book

The subject *I* precedes the finite verb *am*, giving SV word order, which is expected of English. Likewise, the NP *long book* has the modifier *long* preceding the head *book*, which is also the word order expected of English. Thus the MOP points to English being the ML for this clause. The outside late system morpheme in the clause is the verbal agreement morphology of *am*. This is English, agreeing with the subject, and so the SMP also points to English as being ML. Since both principles indicate that English is providing structure to the clause, English is identifies as the clause ML. Consider the example in (54), which is a bilingual clause.

(54) I am reading a llyfr hir
1S be.1S.PRES reading DET long book
"I am reading a long book."

Here the SMP identifies English as supplying the outside late system morpheme in the verbal morphology of *am*. The relative position of the finite verb *am* and the subject pronoun *I* is also typically English, SV. The Welsh morphemes in the NP *llyfr hir* 'long book' are in a word order compliant with Welsh rather than English, and these morphemes are identified as constituting an EL island, where the ML is English and the

EL is Welsh. Similarly to example (51) above, example (55) below has EL morphemes in word order from the ML.

(55)	I	am	reading	a	hir	llyfr
	1S	be.1S.PRES	reading	DET	long	book

"I am reading a long book."

The word order for the head and modifier in an NP in English is MH, which is what is found with the NP *hir llyfr*, where the modifier *hir* precedes the head noun *llyfr*. Since English is providing all word order and system morphemes in this bilingual clause, it too is identified as having an English ML.

iii. Dichotomous ML

When the MOP and the SMP do not identify the same language as the source of the ML for a clause, or where there is conflicting internal information for the purposes of satisfying one principle (or both principles), the clause has a dichotomous ML. That is, one of the following holds:

1. The MOP indicates that Welsh is supplying word order to the clause, while the SMP indicates that English is supplying outside late system morphemes.
2. The MOP indicates that English is supplying word order to the clause, while the SMP indicates that Welsh is supplying outside late system morphemes.

3. The SMP indicates one language as supplying outside late system morphemes to the clause, but there is some word order in the clause that is sourced from Welsh and some word order which is sourced from English, and so the MOP does not identify one language as the sole source of the ML.
4. The MOP indicates one language as supplying word order to the clause, but there are some outside late system morphemes in the clause that are sourced from Welsh and some outside late system morphemes which are sourced from English, and so the SMP does not identify one language as the sole source of the ML. However, since the only outside late system morphemes in Welsh and English are finite verb morphology, and it is difficult to conceive of a simple clause which contains multiple finite verbs, this type of scenario will not apply to Welsh-English data.

1. Examples

Example (56) below is a dichotomous ML clause of the first scenario type outlined above. It, along with the other examples below, is bilingual. Again, these examples are hypothetical and are used to illustrate the application of the model.

(56) Read I a llyfr. long

read I DET book long

"I read a long book"

In (56) the finite verb *read* is English, and so the SMP identifies English as the ML. However, the word order of the clause is not typically English. The verb *read* precedes the subject *I* (VS), and the modifier *long* follows the head noun *llyfr* (HM) in the NP *llyfr long* 'long book'. Both VS and HM are predominantly Welsh word orders, and so the MOP indicates that Welsh is supplying all the morpheme order in this clause. Since the SMP and the MOP point to different languages, this clause would have a dichotomous ML.

(57) Fi wnaeth orffen hir book
1S do.3S.PAST finish.NONFIN long book
"I finished a long book."

In example (57) above, the second scenario type, the finite verb *wnes* is Welsh, so the SMP identifies Welsh as the source of the ML. However, the morpheme order in the clause is not typically Welsh: the SV order of *fi* and *wnaeth*, as well as the MH word order of *hir book* 'long book', indicate that English is supplying clause word order. The MOP points to English while the SMP points to Welsh. Since both principles point to different languages as the source of the ML, (57) would have a dichotomous ML.

(58) Dw i 'n darllen long llyfr
be.1S.PRES 1S PRT read.NONFIN long book
"I'm reading a long book."

As in (57) above, the clause in (58), which is of the third scenario type, has a Welsh late outside system morpheme (*dw*), so the SMP indicates Welsh. The word order of the finite verb and the subject is also

typically Welsh: *I* follows *dw* (VS). However, the word order in the head/modifier NP *long llyfr* 'long book' is MH, which is predominantly an English word order. The MOP has conflicting information about which language morpheme order is sourced from: the clause would have a dichotomous ML. The clause in (59) below is another example of the third scenario type.

(59) I 'm reading a llyfr long
I am reading a book long
"I'm reading a long book."

The finite verb *am* in (59) is English, so English is providing the outside late system morpheme in this clause. The relative positions of *am* and the subject *I* is SV, also indicating that it is sourced from English. However, the word order in the head/modifier NP *llyfr long* is HM, which is a predominantly Welsh order. The MOP points to both Welsh and English as being sources of morpheme order: the clause would have a dichotomous ML.

iv. Unidentifiable ML

If any of the following holds, then the clause has an unidentifiable ML:

- The clause has no finite verb (or no overt finite verb), and so the inflectional morphology of this morpheme cannot be analysed to satisfy the requirements of the SMP. For this reason, nonfinite clauses and verbless clauses (examples of which I give below) often have an unidentifiable ML. However, if a finite verb is not

overt but the subject is overt, the finite verb is presumed to have been ellipted, and so the clause has an identifiable ML based on the presumed language of the ellipted verb; this could be identified e.g. as Welsh by the common presence of the participle *yn* in clauses where the deleted verb is a form of *bod* 'be' (see chapter 6 for more discussion of this type of clause).[86]

- If there is no finite verb or head/modifier NP present in the clause, then the MOP cannot be tested.

1. examples

Examples (60) and (61) below are (bilingual) fragments without enough information to identify an ML. The clause in (60) is nonfinite, and lacks a head/modifier NP, so neither principle can be applied to it. The clause in (61) is verbless, and the same applies.

(60)	...	taflu	'r	<u>ball</u>	
		throw.NONFIN	DET	ball	
		"...throw the ball."			

(61)	...	bron	mor	<u>brilliant</u>	â	Huw
		almost	so	brilliant	as	Huw

[86] This does not mean, however, that such participles (e.g. *wedi*, *yn*) are required to be interpreted as being outside late system morphemes, as they have no grammatical relations external to their maximal projection. Such morphemes are probably best viewed as being bridge system morphemes since they serve to link a finite verb to its complement phrase. By themselves they are therefore not sufficient to identify a clause's ML as being Welsh according to the parameters of the MLF model, since a hypothetical example could be created where a clause has an English outside late system morphemes (i.e. finite verb morphology) and word order but has a Welsh participle such as *yn*. Such a clause would have an English ML with a Welsh insertion.

"...almost as brilliant as Huw."

As the data I will be analysing are natural data, monomorphemic utterances such as *yeah*, *na* 'no', *ella* 'maybe' etc. are numerous. It is not possible to identify an ML in such utterances, and so these fragments are labelled as having an unidentifiable ML.

7. Conclusion

In this section I have discussed the concept of a matrix language in bilingual speech, focusing on Myers-Scotton's Matrix Language Frame model. I have described how it can be used to analyse the participation of languages in classic code-switching, but also how it can be used to identify composite code-switching and, crucially, convergence, by introducing the concept of a dichotomous ML. I have then described the differences and similarities between the grammars of Welsh and English, and suggested loci where convergence might occur. Finally, I have shown how the MLF model can be applied specifically to Welsh-English data, where the options for identifiable ML for a clause are Welsh, English or dichotomous. This last type may show signs of convergence.

In the next chapter I describe the methodology of data collection and coding used for the data analysed in this thesis.

Chapter 4 — Methodology of data collection and transcription

1. Introduction

In the previous chapter I explained the methodology of my data analysis for this thesis, centred on the Matrix Language Frame model.

In this chapter I outline the methodology of collecting, transcribing and codifying my Welsh-English data. The analysis of this data will be presented and discussed in chapters 5 and 6.

As the data used derives from a larger corpus of Welsh-English speech data, called Siarad, this chapter will focus on the methodology used in building that corpus.

2. The source of the data

The research conducted for this thesis uses data collected as part of a 5-year project (2005—10) funded by the Arts and Humanities Research Council (AHRC) entitled "Code-switching and convergence in Welsh: a universal versus a typological approach". This project was staffed by a full-time research assistant, myself, a second doctoral student,[87] and a director (the award holder Prof. Margaret Deuchar), as well as

[87] Both myself and the other student held Ph.D. studentships attached to the project.

administrative and technical staff. All members of the project team contributed to data collection and transcription. The aims of the project were both practical and theoretical, and are outlined below.

- Theoretical aim: to test the theoretical models of code-switching proposed by Myers-Scotton (1993, 2002), viz. the Matrix Language Frame model, and Muysken (2000), viz. the Typological Model, on Welsh-English code-switching data to ascertain to what extent the principles of both models apply to Welsh-English speech.
- Practical aim: to construct a corpus of Welsh-English code-switching data by collecting approximately 40 hours of recorded and transcribed spoken data from Welsh-English bilinguals, to be made publicly available on the Talkbank website (http://www.talkbank.org) for the benefit of future researchers.

The corpus consists of 40 hours of recorded and edited data, comprising 69 recordings with 151 speakers overall. Most recordings were between two people and lasted approximately 30 minutes[88], but this varied from recording to recording.

The data were collected and transcribed in the manner I describe below. A subset of the transcribed data was used for the research related to the current thesis.

3. Data collection

[88] i.e. after editing.

In this section I describe the approaches used in data collection, which include finding and contacting participants, recording their discourse, and obtaining information on extralinguistic variables.

a. Designing the sample

In order to ensure diversity within the data, a stratified random sample method was adopted, in order for the project analysts to "have the possibility of making inferences about the population based on the sample" (Sankoff 1988:900). For this to be achieved, each independent variable within a certain set of pre-determined variables had to be reflected as fully as possible in the final data set. Within those categories of variables, however, the participants were chosen partly at random by adopting an advertising approach and partly via a social network/word-of-mouth approach (see the next section for details).

Independent extralinguistic variables are those variables which can be manipulated to see which ones correlate to dependent linguistic variables. The five main independent variables considered relevant to the project's research aims were:

1. **Gender**
    - Both males and females were sought.
2. **Age**
    - A wide age range, covering children from around 10 years old upwards through to the older generations.
3. **Regional background**
    - The area in which a speaker was born, grew up in and subsequently lived in as an adult. These areas were to

reflect the breadth of Wales (and other Welsh-speaking areas), particularly reflecting the traditional north-south regional divide of Wales.

4. **Socio-economic background**
    - 'Class' variation, identified via a person's employment and their highest level of formal education.
5. **Linguistic background**
    - The language spoken to a speaker by their parents when growing up, as well as the medium through which their school education was conducted, contribute to the level of a speaker's bilingualism in terms of balance and proficiency. This allowed for analysis of e.g. speakers who were more dominant in Welsh than English or English than Welsh, and/or who had acquired one language before the other.

b. Methods of obtaining participants

To find speakers distributed among these variables, two main methods were used:

- Participants known to members of the project team were contacted by word of mouth, using the 'friend of a friend' approach (cf. Milroy's [1980] social network approach), ensuring at the same time that potential participants did not know much about the project's specific theoretical aims. Letters, e-mails and telephone calls were used to contact potentially willing participants (see Appendix A for the letter used to contact

potential participants), and arrangements were agreed on for making a recording.

- For other speakers, a set of posters and advertisements was used to target potential participants from different social groups. To target university students, informally worded monolingual Welsh posters, using the 2nd person singular Welsh pronominal forms as a means of address,[89] were placed in university buildings, on university computer network bulletin boards, and some circulated via e-mail. To target people from outside the university, it was decided to take a more formal approach, and so posters with similar wording but using the 2nd person plural pronominal forms[90] were placed in shops, libraries, schools, community centres etc. It was reasoned that Welsh-only posters would aid with the aim of attracting only participants with reasonable fluency in Welsh. (See Appendix A for a facsimile of a poster we used.) These types of advertising were done at a local level and also in other parts of Wales where the project team had contacts, in order to widen the search as much as possible. People interested in participating in the project's research were invited to contact the project team by e-mail or telephone. A list of potential speakers was created from interested potential participants who also fitted into the stratified sample. These potential participants were then contacted directly to arrange a recording.

---

[89] This is the more informal mode of address in Welsh. We considered that using an informal register would more likely to attract younger participants, in this case students.

[90] This is the more formal mode of address in Welsh, which we considered to be more appropriate when dealing with the general public than with students within the University.

c. Recording the participants

To facilitate the process of recording all the participants, a spreadsheet was created to keep details about the speakers and their status with regard to recording, which was colour-coded according to their willingness and ability to be recorded. Keeping this document up to date was one of my primary responsibilities on the team. After a recording had been completed, details about it were transferred to a second spreadsheet to keep details of the progress of collecting the corpus data.

i. Arranging recordings

Having been recruited, speakers were instructed to find a speaking partner to be recorded with. This person was to be someone with whom they would normally speak casually and comfortably, ideally a close friend, colleague or family member. Whilst this approach did not allow the project team to guarantee that all speakers would represent the variables of the stratified sample, it was considered that allowing participants to choose their own co-participants would increase the naturalness of the data sample.

Speakers were given their choice of a location to conduct the recording, also to maximise naturalness.

ii. Recording process

A date, time and location for the recording was arranged with participants, and they were attended to by a researcher (occasionally more than one), who always initiated the meeting through the medium of Welsh (on the premise that speakers would be more inclined to speak at least some Welsh in their recording if they were greeted in that language).

Participants were told that they were participating in research on bilingual communication in Wales, but did not have the research topic explained to them in detail. They were invited to speak in whatever language they wished (Welsh or English). They were told that they could speak on any topic, but in case they ran out of inspiration they were provided a couple of non-specific topics, such as "holidays" or "what I'll do after University" by the researcher. These were rarely required, however.

The researcher was not present in the same room as the speakers (or, when recorded in open surroundings, not in the immediate vicinity) during the recording. This was in order to address the Observer's Paradox (Labov 1966), as the aim of obtaining naturalistic data is affected by the present of a researcher: it was considered that if the researcher were not present with the participants that they would feel less inhibited and so less likely to modify their normal speech patterns (and e.g. adopt a more formal register). The researcher retired to a separate room after instructing the speakers to begin their conversation.

Recording was done using one of the two following sets of equipment:

- Two radio microphones were given to the speakers. Such microphones are small and unobtrusive and run on batteries.

They broadcasted sound to a monitoring and recording device in a different room, which the researcher(s) observed, listening in on the conversation to make sure the microphones were transmitting or that the speakers had not fallen silent. The sound data were saved onto a Marantz hard disk recorder. The twin microphones allowed speakers to be recorded on two separate audio tracks in stereo. When there were more than two speakers, either only one microphone was used, placed between the speakers, or one speaker was given a personal radio microphone and the other radio microphone was placed near the remaining speakers (the latter method allowing for at least some stereo distribution of participants). The advantage of this equipment was that it allowed for recording of sound files in stereo, which facilitates transcription, as individual audio tracks can be isolated to make the speech of the speaker on that track clearer. A disadvantage to this piece of equipment was its physical size and weight, which made it more practical to use in the project offices and laboratory than in external recording venues.

- A portable Sony minidisk recorder. This operated with a standalone microphone placed between the speakers, which recorded their conversation directly onto the minidisk. This machine recorded in mono, rather than stereo, and thus all speakers were on the same audio track. The advantage of this equipment was its portability, making it the more practical choice for recordings conducted in external locations, but its disadvantage was its incapacity to record on multiple audio tracks, making transcription of data recorded on the minidisk recorder potentially more difficult than that recorded using the

Marantz recorder, in terms of the intelligibility of unclear sections of the recording or overlapping speech.

After approximately 35 minutes[91] had elapsed, the researcher returned to the participants and stopped the recording. At this point the participants were asked to fill out a questionnaire (this questionnaire outlined in section 3.d below).

One of the most salient contributions I made to the corpus was collecting data in the form of recordings. 17 of the 69 recordings (24.6%), those titled DaviesX, were collected by me (or me and another researcher in a small number of instances), compared to 32 (46%) being the most recordings collected by a single person attached to the project, who was a researcher essentially working full-time on data collection. I also helped other researchers with undertaking other recordings, e.g. Stammers1 and Stammers2. Since most data collection was undertaken during the first two years of the project period (2005-6), I was therefore highly active during this time in the task of data collection alongside my other work associated with the project/PhD study.

The recordings were afterwards transferred onto computer disk. Every recording was subsequently edited using Adobe Audition software, subject to the following rules:

- The first 5 minutes (approximately) of every recording was deleted. It was considered that potential stiltedness of speech by participants in response to the recording

[91] In some circumstances the recording period was shorter by necessity, e.g. Davies9, which was recorded at an outdoor Eisteddfod at a busy public tent, and the participants had limited time.

equipment would only last a few minutes before participants adjusted to the situation. This was a move to ensure more naturalistic output from the speakers.

- Participants were asked afterwards whether there were any portions of their conversation which they would prefer not to be used as data (such as private information). If there were any such requests these portions were deleted from the recording (this only happened with a small number of recordings).
- Sections where the researcher took part in the discourse (such as when re-entering at the end of the session or when attending to difficulties with the recording equipment) were deleted, though the researchers' speech was not used for analysis.

Recordings were saved as waveform (.wav) sound files under filenames denoting the name of the researcher who collected the data and the number of the recording (Davies1, Fusser2, Stammers3, etc.).

d. Speaker questionnaires

A questionnaire was administered to every participant after each recording was conducted. I was a core part of the small team that finalised what questions should be includes on the questionnaire and how it should be set out. The primary aim of the questionnaire was to provide information about independent variables which would be used to examine variation in the data.

The questionnaire was designed as a collaborative work by all members of the project team. The questionnaire had to be designed during the first few months of the project period (and of my first year studying for a PhD), since it was essential to have it ready before any participants were recruited or recordings made. At this time the primary researchers involved with creating the questionnaire were myself, the second PhD student attached to the project, one Research Asssistant and the project director. Discussions were held during project team meetings in order to decide what types of questions would be posed in the questionnaire, and then I and the second PhD student worked together between meetings to develop the precise formulation of these questions. During subsequent meetings the formulation was refined until a final version was created. The final stages of formatting and design were undertaken primarily by the second PhD student, during which time I focused on recruiting participants to record.

The questionnaire was in Welsh on one side and English on the other; participants were instructed to answer the questions in whichever language they preferred.

The questionnaire also included a consent form, which the participants signed in order to show that they acquiesced to their speech being analysed as data and to be uploaded onto the Talkbank website, with the understanding that their real names would not be used. Participants under 18 were issued a consent form for a parent or guardian to sign to signify that they agreed for the children's data to be used. A copy of this consent form can be seen at the end of the questionnaire in Appendix B.

The researcher was present with the participants as they filled in the questionnaires, to attend to any queries. In some isolated cases the participants had insufficient time to complete the questionnaires during

the recording session, and so took them home to complete them and thereafter returned them to the project team by post.

The completed questionnaires were transferred into a MS Excel spreadsheet for the purposes of analysis.

i. Questionnaire variables

I set out here the categories of variables which were represented as questions in the speaker questionnaire. These categories were not explicitly stated on the questionnaire. In each category I note the questions (Q) which were posed to the participants in order to collect information relevant to that category. See Appendix B for a copy of the questionnaire.

1. Gender

Participants were asked if they were male or female (Q1). This allowed for later extralinguistic analysis by gender, to discern if patterns in the data varied between men and women.

2. Age

Participants were asked for their date of birth (Q2), to identify the age of the participants at the time of recording. This allowed for analysis of a speaker's age as an independent variable. particularly with regards to

the apparent time paradigm (Chambers 1995; also see e.g. Tagliamonte 1998); see chapter 6 for more discussion of this concept.

In addition to determining the age of the speaker at the time of recording, the date of a speaker's birth allowed the possibility of calculating other factors, e.g. the age of the speaker when they lived in a particular region.

3. Occupation

Participants were asked for their present occupation, or most recent occupation if they were retired (Q3). A blank space was left for the participant to fill in the details. Results for this question allowed for variation analysis according to a participant's (approximate) social position ('class').

4. Regional origin

Participants were asked to give all the places in which they had lived for any substantial amount of time over the course of their life (Q4). This was not expected to include locations where the speaker only spent a small amount of time, or had not become integrated into that community, though did include e.g. time they had spent at University. There was no limit to the number of locations a speaker was allowed to cite. Exact time frames were not required, only the years between which the speaker lived at that location.

This question allowed the use of regional origin as an independent variable, to identify whether or not a participant who had

lived in e.g. a predominantly Welsh-speaking area showed different linguistic variation compared with a speaker who had lived in a predominantly English-speaking area, or whether or not there was significant linguistic variation between participants who had lived mostly e.g. in north Wales compared to those who had lived mostly in south Wales (reflecting the traditional north/south dialectal split in Welsh).

5. Education

Participants were asked for their highest level of formal education (Q5). This was to ascertain the educational level of a participant, to allow for variation analysis according to this independent variable.

The various stages of a participant's educational experience was set as being at: the end of secondary school (year 11 in the UK) or before; the end of sixth form/college; the end of University (i.e. whether graduated); and postgraduate qualifications (no distinction was drawn between Master's and Doctorate degrees).

In addition to standard school and University qualifications, NVQs and GNVQs were included in the selection, along with National Diplomas. though specific occupational qualifications were not. Any speaker possessing an academic qualification not listed on the questionnaire was invited to ask the researcher present under which subset their qualification should fit.

6. Age of acquisition

A set of questions was devised to elicit information about a speaker's language acquisition of both Welsh and English. Q6 and Q7 asked about the time when the speaker learnt each language, whilst Q10, Q11 and Q12 asked for information about the languages the speaker heard whilst growing up at home.[92] Q13 and Q14 concerned the language through which the speaker was taught at school. Most children in schools in Wales are not taught exclusively through Welsh in secondary school (English literature/language classes are usually taught in English, for example), so the speaker was asked to select the language through which they were taught the majority of the time. The options were 'Welsh', 'English' or 'Welsh & English'; this third option was intended to be selected if the speaker had been taught significantly through the medium of both languages.

Using the age at which a participant acquired their languages as an independent variable allowed for analysis of linguistic variables across bilinguals whose first language was acquired significantly earlier than their second language, as well as more balanced bilinguals, i.e. those who had acquired both Welsh and English at around the same time.

Q8 and Q9 sought a self-report from the participants of how well they thought they spoke Welsh and English. This allowed analysis of the correlation of the participants' own report of their linguistic ability with their actual output.

---

[92] Q12 asked for the language spoken to the child by 'any other guardian or caregiver'. This was intended to cater for those speakers who had grown up with a significant guardian who was not a parent, but some speakers, when filling in the questionnaire, interpreted this as the language spoken by a third party who was significant in their upbringing, such as a grandparent.

7. Social network

Participants were asked (Q15) to list the five people with whom they spoke most frequently, naming them if desired, or simply stating their relationship ('mother', 'best friend' etc.), and specifying the language (Welsh or English, or both, or another language) they usually used when conversing with that person.

This question aimed to elicit the linguistic social network of the speaker, and thereby to measure which language was dominant for that speaker (I.e. the languages they used most often when speaking to other people), possibly to be taken into consideration when measuring whether or not a speaker is a balanced bilingual.

8. Attitudes to Welsh and English

Participants were asked to give their personal views on various aspects of Welsh and English (Q16—Q20).

For Q16 and Q17 participants were asked to rate Welsh and English on a ratio of 1 to 5 for six pairs of adjectives (1 represented a negative view of that language and 5 a positive view). Three pairs of adjectives were concerned with a general feeling about the language ('modern'/'old-fashioned', 'friendly'/'unfriendly', 'beautiful'/'ugly') and the other three pairs were concerned with the practicality of the language ('influential'/'uninfluential', 'inspirational'/'uninspirational', 'useful'/'useless').

Q18 was concerned with the participants' self-report on national identity, which asked them whether they felt they identified themselves

as being Welsh, English, British, European or something else. This allowed for the analysis of extralinguistic variables focusing on whether e.g. those who identify as being Welsh exhibit variation to those who do not.

Q19 and Q20 asked for the participants' opinions about code-switching, the main focus of the project. Q19 was concerned with the participants' self-reported individual use of CS. Q20 enquired after the participants' opinion on code-switching (from Welsh to English) in general. The answers elicited from these questions permitted later comparison between the participants' reported switching activity and their actual switching activity, to discern, for example, to what extent Welsh-English bilinguals are 'aware' that they are indulging in code-switching.

4. Transcription and coding of data

Each recording (excluding the edited or deleted sections outlined above) was transcribed using the CHAT system (see below). The transcript of each recording was divided into clauses, with each clause glossed and with a paraphrased English translation supplied where necessary (i.e. where there was at least one non-English word in the clause). In addition, every word in each clause was tagged with a language marker to denote the language from which that word was sourced.

a. The transcription process

The transcription was completed, as with the data collection, by the doctoral students and research assistants. Each individual recording was first transcribed by one researcher, who also completed all the glossing and translation. Transcription conventions were agreed upon by the project team whilst the transcription phase took place.

On a subset of the completed transcripts (22%) an independent transcription was undertaken, in order to check inter-transcriber reliability. This was done as follows. After the initial transcriber completed a transcription, a randomly-selected 1 minute section of the recording was chosen by a second transcriber (who had not previously read the original transcript), and then transcribed as part of a reliability checking system. The relevant transcribed portions of the recording were then compared by the two transcribers to establish that both transcribers agreed on the interpretation of the speech and also on transcription conventions. Any significant anomalies were attended to by the original transcriber in the original transcript. The two transcript portions were then sent through the Turnitin plagiarism software program to provide a statistic on how similar they were: this was called the reliability score. The average reliability score for the transcripts that were independently transcribed was 75%, indicating that all transcribers were transcribing according to the same conventions—this method gave scores of less than 100% due to transcribers typically interpreting pauses, repetitions and/or dialectal orthographic systems differently, rather than being radically dissimilar in the interpretation of any given section of a recording.

Next, for any transcript completed before the end of March 2007,[93] a third transcriber proofread the entire original transcript to check for any errors. The third and first transcriber would then meet to discuss any issues, and any errors or discrepancies were attended to. After this a transcript was considered definitive.

i. Encountering audio difficulties

Whilst the sound quality of the audio recordings was high, participants often spoke at speed or unclearly, or otherwise interference to the equipment (such as their clothing ruffling against the radio microphones) made their speech unclear to a certain extent. These sections were discussed between the project team members to discern if there was any agreement on what the utterance might be, and if there was, then this was inserted into the transcript, usually with a marker in the transcription (see below) to indicate that there was uncertainty. If a section was indeed not discernible, then it was, either in part or in their entirety, marked as being incomprehensible (see below).

ii. Language allocation of words

In order to facilitate the analysis of code-switching in the data (a major aspect of the project's theoretical aims), it was decided to code each

---

[93] This date was chosen because we surmised that transcriber error would be more likely to occur during the first year and a half of the project period, while transcribers were still learning the system.

word produced by speakers according to language origin, either from Welsh, from English, from both, or potentially from either.

The criterion used was whether or not a word occurred in certain dictionaries, with phonology and prosody used as a basis for this. Welsh dictionaries (*Geiriadur Prifysgol Cymru*, the computer-based dictionary *Cysgeir*, as well as King 1996) were used as a basis for identifying whether or not a word was in the Welsh lexicon. Welsh Words which did not appear in these Welsh dictionaries but did appear in the *Oxford English Dictionary* were marked as being English. Some words (e.g. *lot, shop*) are homophones and homonyms in Welsh and English, and so such words were tagged differently, indicating that these words could potentially be sourced from either Welsh or English.

The language marker tags used are discussed in section 4.a.iii.2 below.

iii. The CHAT system, LIDES, and the Talkbank website

The system used for the transcription is CHAT, and the computer program used is CLAN. Both the system and program are outlined below.

A sample of a transcription I contributed to the corpus is given in Appendix E.

One incentive to use the CHAT system to encode the transcription for this project was that it is the preferred system used by LIDES (Language Interaction Data Exchange System; see MacWhinney 2000). LIDES is an internet-based database to which researchers around

the world have provided transcriptions of bilingual data in many languages, and the CHAT system derived from this project. LIDES data is now stored on the Talkbank website (http://www.talkbank.org). The aim of LIDES is to provide a large corpus of data for researchers to access and add to themselves, and to further their own research.

After the whole Siarad corpus was fully-transcribed, it was uploaded onto the Talkbank website as a downloadable database, both transcripts and sound files.

This present thesis also makes use of LIDES data for languages other than Welsh and English,[94] and the CLAN programme allowed comparison between datasets from different language corpora.

1. The tiers used in CHAT

Each CHAT transcription was prefixed by a header which provided information about the participants in the recording, though confidential details (including their real names and their places of residence) were omitted. The age and gender of participants (a mandatory part of the CHAT transcript header) was noted, and in these transcriptions the general regions where each participant had lived during their lives, the relationship of all the participants with one another (e.g. siblings, friends, colleagues), details about the recording itself (e.g. date, situation and location), and the identity of the researcher and the first transcriber were all noted.[95]

---

[94] As I show in chapter 5, I use some of Eppler's (2004) Austrian German-English data, located on Talkbank.

[95] These are not obligatory according to the CHAT system, but it was considered to be information that would be of use to anyone making use the project's transcriptions. The only obligatory tier for the purposes of CLAN and CHAT is the main tier.

Every participant in the recordings was given a unique pseudonym to protect their privacy. Each pseudonym was further given a unique three-letter abbreviation for identification within the transcription as a main tier prefix (see the next section), according to the requirements of the CHAT system. When speakers referred to one another by name in the recording, this was replaced by a pseudonym in the transcript.

CHAT utilises a system of tiers to represent the output of the speakers. The main tier is mandatory whilst others are optional according to the CHAT system. For the transcriptions of this project, two other tiers were used, namely a gloss tier and a translation tier. The contents and purposes of these three tiers are outlined below.

2. The main tier

The main tier was the tier representing actual speech. The conversation was divided into complex clauses where possible, and each main tier only represented speech by one speaker at a time. The main tier was labelled by the three-letter abbreviation of the pseudonym of the speaker, e.g. a main tier of Amranwen's speech would be labelled *AMR.

The transcription in the main tier was an orthographic representation of the output. Elided phonemes were, where appropriate, indicated by round brackets; e.g. the pre-verbal particle *yn* [ən] often has the vowel deleted after a vowel phoneme, and was thus transcribed as (*y*)*n* in such contexts.

Initial mutation was transcribed when it occurred. This included mutation occurring on English-source morphemes, with Welsh

orthography used for the mutated grapheme but English orthography used for the rest of the morpheme; e.g. when *mileage* underwent soft mutation and was pronounced [vailidʒ], it was transcribed *fileage*, because [v] is represented by *f* in Welsh orthography.

In the event that a word consisted of morphemes from both Welsh and English bound together (e.g. *love-io* 'love', where the stem was English but the affix Welsh), the morphemes were divided by an underscore (*love_io*). An underscore was also used in lieu of a hyphen in hyphenated phrases (e.g. *mini_bus*), as well as some multi-lexeme constructions which could not usefully be subdivided into their component parts (e.g. *dim byd* 'nothing (lit. not world)' was transcribed *dim_byd).*

It was also in the main tier that language marker tags were attached to each word. These tags denoted the language source of the word, as noted above in section 4.a.ii above. If the word was sourced from Welsh it was tagged *@1* (e.g. *cartref@1*). If the word was sourced from English it was tagged *@2* (e.g. *home@2*). Words containing a morpheme from Welsh and a morpheme from English were tagged *@21* if the first morpheme was Welsh and the second morpheme English (e.g. *gazump_io@21*), and *@12* if vice versa. Morphemes which were identified as being homophones and homonyms in Welsh or English, such as established lexical borrowings like Welsh *siop* 'shop', were marked *@0* (e.g. *shop@0*); English orthography was used to transcribe such words.

Some Welsh words were borrowings but had a substantially different pronunciation in Welsh compared to the Welsh English pronunciation. We were concerned with marking when words were homophones and when they were not. Therefore these morphemes were

tagged *@0* if the pronunciation was not typically Welsh English and *@2* when it was, so e.g. *right* when pronounced [reit] would be identified as either a Welsh or English pronunciation, so was tagged *right@0*. When the word was pronounced [rait], however, it was identified as not being a typically Welsh pronunciation, and it was tagged *right@2*.

Words which were from languages other than Welsh or English were marked as @3, @4 etc, where each number represented a different language; e.g. @3 represented a Spanish word.[96]

The intention of using the language derivation markers was to disambiguate where possible, and stress patterns were taken as a useful additional indicator that the word pertained to one language rather than another, especially that pronunciation of segments was not otherwise an adequately disambiguating indicator. For example, a word such as *convention* would have a similar pronunciation in standard English and Welsh English, and, moreover, would display a similar stress pattern ([kɒn'vɛnʃən], with the stress on the penultimate syllable). This word would be marked as *@0*. However, if the word was *conventional*, the stress would occur on a different syllable in the word, depending on whether the stress pattern produced was Welsh or English. So, if the output was [kɒn'vɛnʃən,əl], this was judged to be an English stress pattern, and the word would be transcribed *conventional@2*. However, if the output was [kon,ven'ʃən,ol], this appeared to be a Welsh stress pattern (with the stress on the penult), so the word would be transcribed *confensiynol@1*.

Incomprehensible words or strings of words (sometimes entire turns), as noted above, were transcribed with 'xx' or 'xxx'.[97] Words that

---

[96] Marking of words from languages other than Welsh and English was done regardless of whether or not the speaker in question had any fluency in those other languages.

could be tentatively identified, but, due to the poor quality of sound or unclear pronunciation by the speaker, were not identifiable with more certainty, were marked with a question mark in square brackets, ('[?]'). Phonological fragments which were comprehensible but not identifiable as a word were transcribed phonetically with an ampersand ('&') and a broad IPA transcription of the utterance, sometimes with a guess by the transcriber of the intended word in square brackets following it; e.g. *&danʃin [?: dancing]*.

Quoting of direct speech was marked with triangular brackets and a ["] marker. Pauses or hesitation were shown using the hash: '#' for a brief pause, or '##' for a lengthier pause, with '###' for a very long pause. Repetition, self-correction and reformulation of parts of utterance by speakers were represented by [/], [//] and [///] respectively.

Non-linguistic sounds (such as laughter, coughing, sighing, etc.) were represented by a set of markers beginning with an ampersand, e.g. &=laugh, or [!=laugh] if occurring simultaneously with speech.

The above transcription conventions (including language tagging system) were designed and developed by the project team in collaboration, based on (a) some existing conventions within CHAT and (b) some conventions used for a pilot project (also directed by Margaret Deuchar) for which a smaller Welsh-English corpus was collected. Innovative conventions developed for our team include the tagging of all language items (rather than e.g. just those from one language), the language tagging of individual morphemes within words, a detailed glossing of verbs in the gloss tier etc. I was highly active in the lengthy

---

[97] If the incomprehensible sounds were interpreted as probably being just one word, they were transcribed as xx. If they were interpreted to be multiple words, they were transcribed as xxx.

discussion process during which these conventions were developed and finalised.

3. The gloss tier

The second tier used for each utterance was the gloss tier, headed by '%gls', giving a word for word gloss of the morphemes from the main tier. That is, each item in the main tier was represented by one item in the gloss tier. Non-linguistic sounds and phonetic fragments marked with an ampersand, along with any material enclosed in square brackets, were excluded from this rule.

English items (marked *@2* in the main tier) and items which could be from either Welsh or English (marked *@0*) were repeated in the gloss tier with the same form as in the main tier (without further morphological marking). Welsh items (marked *@1*) were given an equivalent English translation in the gloss tier, and, depending on their word class, were glossed in various ways, as outlined below.

- Nouns were not glossed for inflection (such as plurals).
- Finite verbs were glossed for person, number and gender where applicable, separated by periods; e.g. *rhedodd* is glossed *run.3S.PAST* as it is 3rd person singular and in a past tense; *wyt* is glossed *be.2S.PRES* as it is 2nd person singular and in the present tense. Nonfinite verbs (verbal nouns) were glossed as such (e.g. *rhedeg* is glossed *run.NONFIN*).
- Adjectives were not glossed for gender, but numerals were when they had both a masculine and feminine option; e.g. *dau* 'two',

the masculine form, and *dwy* 'two', the feminine form, were glossed *two.M* and *two.F* respectively.

- Inflected Welsh prepositions, e.g. *iddo* 'to him', *danom* 'underneath us', were marked for person and number; e.g. *iddo* was glossed *to.3Sm* as it is 3rd person singular masculine; *danom* was glossed *under.1PL* as it was 1st person plural.
- Other parts of the main tier were glossed in various ways, e.g.
    - The definite article (*y,* yr, 'r) was glossed *DET*.
    - Most particles (e.g. pre-verbal *a*, participial *yn*, relative *mai*) were glossed *PRT*, though the perfective particle *wedi* was glossed *PRT.PAST*.

4. The translation tier and other tiers

The third tier used for each utterance was the translation tier, headed by %eng. This provided a free English translation of the text of the main tier. It was not necessary for the word-for-word correlation found between the main tier and the glossed tier to be maintained here. The aim of the translation tier was to aid those reading the transcripts who were not familiar with the Welsh language.

When further comment about a feature of the main tier was required, such as clarifying a speaker's meaning with a given sentence, a comment tier headed by %com was used. When more general information concerning a given utterance appeared useful, e.g. background noises required to be noted, or the highlighting of a particular linguistic point seemed relevant, a comment tier headed by

@Comment was added following the other tiers to aid users of the transcription.

b. CLAN

The program CLAN was used to transcribe the data. In CLAN it is possible to run the sound file of the recorded conversation parallel to the transcript, and, by entering 'bullets' at the end of each main tier that correspond to contiguous sections of the sound file, CLAN can execute a running audio presentation of the transcript. It also allows any utterance to be heard individually by clicking on the sound bullet adjacent to the relevant main tier.

CLAN was also useful for the analysis of transcriptions, as it had a number of functions which searched for certain items within a given transcription (or set of transcriptions). These were executed by commands in the Command box, and included the following commands:

- FREQ: this command checked the frequency of items within the transcription, and provided a list of each word (in alphabetical order) with the number of times that item was found, along with the item's language marker. FREQ could also make a language distribution check, which showed how many words from each language appeared in a given transcription.
- KWAL: this command isolated every occurrence of a given word, or set of words, or part of words, and printed each main tier that included it. This can be used to e.g. analyse all tokens of a certain word type in the data.

- COMBO: this command identified a particular string of words and returned every tier which matched. It can be used to identify e.g. CS within the same utterance using the language tags.

5. Conclusion

In this chapter I have explained the methodology that was used to choose speakers for the data, how those data were collected, recorded, edited, and then transcribed using the CHAT transcription system. I have also explained the type of independent variables for which data was gathered via the speaker questionnaires.

In the next chapter I present the results of applying the MLF model, as described in chapter 3, to these data.

Chapter 5 — Using the MLF model to find word-order convergence

1. Introduction

In chapter 4 I described the data and the method by which it was collected. In this chapter I will show the results of applying the Matrix Language Frame (MLF) model (as described in chapter 3) to a subset of these data.

As I noted in chapter 1, my intention in this thesis is to see the extent to which English has influenced the morphosyntactic structure of Welsh. To that end, there are two aims in applying the MLF model to the Welsh-English data:

1. to ascertain the distribution of the ML in clauses produced by speakers, to identify which language (if any) is the primary provider of morphosyntax, and
2. to determine how useful the MLF model is with regards to identifying convergence in a set of Welsh-English data.

I will present my analysis using the MLF model in this chapter, responding to both aims. In addition to an analysis of Welsh-English data, I finish this chapter by presenting a shorter analysis of data from Austrian German-English bilinguals, in order to demonstrate how my methodology also works on data from another language pair.

2. Describing the dataset

In chapter 4 I described the Siarad corpus of Welsh-English bilingual data collected for the AHRC-funded project, which this thesis uses as a data source. For the purposes of an analysis involving the application of the MLF model, I selected a subset of these data to be analysed.

This dataset includes data from six speakers, two each from three recordings. The total output is 88 minutes 24 seconds of recorded conversation, or 3065 clauses. I refer to the three recordings in the dataset as Davies6, Fusser6 and Fusser27[98]. I briefly describe the participants of these three recordings below.

Davies6 is a transcript of a recording of a conversation between two male friends, Hector[99] and Daniel. At the time of recording, Hector was aged 23 and a primary school teacher. Daniel was aged 25 and a full-time student. Both had lived in the same part of north-west Wales since birth. For both, Welsh was the language spoken to them by their parents whilst growing up and the medium of their schooling, and both reported learning English at or just before primary school age. The recording was made at the project's research laboratory.

Fusser6 is a transcript of a recording of a conversation between two female work colleagues, Antonia and Amranwen. At the time of recording, Antonia was 52 and a lecturer. She was born in south Wales but moved to live permanently in north-west Wales when she was 21.

---

[98] Recordings are named for the surname of the researcher who made the recording, e.g. (Peredur) Davies, and the sequential number of the recording made by that researcher, e.g. the sixth (6), = Davies6. Examples from the data cited in this thesis use a shorthand form of these recording names, e.g. DAV6 for Davies6.

[99] As noted in chapter 4, pseudonyms have been provided for the participants in order to hide their identity.

Her father spoke both Welsh and English to her but her mother was a monolingual English speaker. Her schooling was conducted through the medium of both Welsh and English. Amranwen was aged 36 at the time of recording and worked as a secretary. She had always lived in north Wales apart from a four-year period spent living in the south of England in her late teens. Welsh was the language spoken to her by both her parents and teachers at school. Both Antonia and Amranwen report having acquired both Welsh and English simultaneously. The recording was made at the speakers' work place.

Fusser27 is a transcript of a recording of a conversation between two female friends, Mabli and Lisa. Both were full-time undergraduate students at University. Mabli was aged 19 when the recording took place. She had lived in north-west Wales since birth. Though she had received all her schooling through the medium of Welsh, she came from a home where her father spoke only English and her mother both Welsh and English. Lisa was aged 20 at the time she was recorded. She had lived in south-west Wales from birth until she moved to north-west Wales to attend University. Her father spoke English to her, whilst her mother used Welsh. Welsh was the language through which she received her schooling. Like the speakers in FUS6, Mabli and Lisa both reported that they acquired Welsh and English at the same time. The recording was made at the speakers' University building, in the project's research laboratory.

The speakers could be grouped as follows: 4 are female, 2 are male; 4 are from north Wales, 2 are from south Wales; all received Welsh-medium education to some level, and all heard at least some Welsh at home when growing up, though three speakers had only one Welsh-speaking parent; 4 of the speakers are in the <30 year old age group, whilst two are over 30, and one, Antonia, is over 50.

In the following sections, when I refer to this dataset as a whole I take all the results of all six speakers together. When I refer to individual speakers by pseudonym, I discuss only the utterances provided by that speaker.

3. Results

In this section I present and discuss the results of the application of the MLF model to the dataset. First I discuss the results of all six speakers taken together, before proceeding to discussing the difference in results from speaker to speaker.

The analysis was undertaken using a Microsoft Excel spreadsheet to record the data. A sample of this spreadsheet is given in Appendix C.

a. Overall results

i. ML distribution in all clauses

The MLF model was applied to all clauses transcribed in the dataset, both monolingual and bilingual.[100] I will discuss monolingual and bilingual clauses separately below when I separate clauses according to finiteness, but for the moment I discuss monolingual and bilingual clauses together without differentiating between them, and then compare the results in monolingual versus bilingual clauses.[101]

---

[100] See chapter 3 section 6a for a description of the clause-division system used for this analysis.

[101] Recall that when I use the term 'linguality' in this discussion, I refer to whether or not a given clause is monolingual or bilingual.

Figure 1 below illustrates the results of applying the MLF model to all clauses.

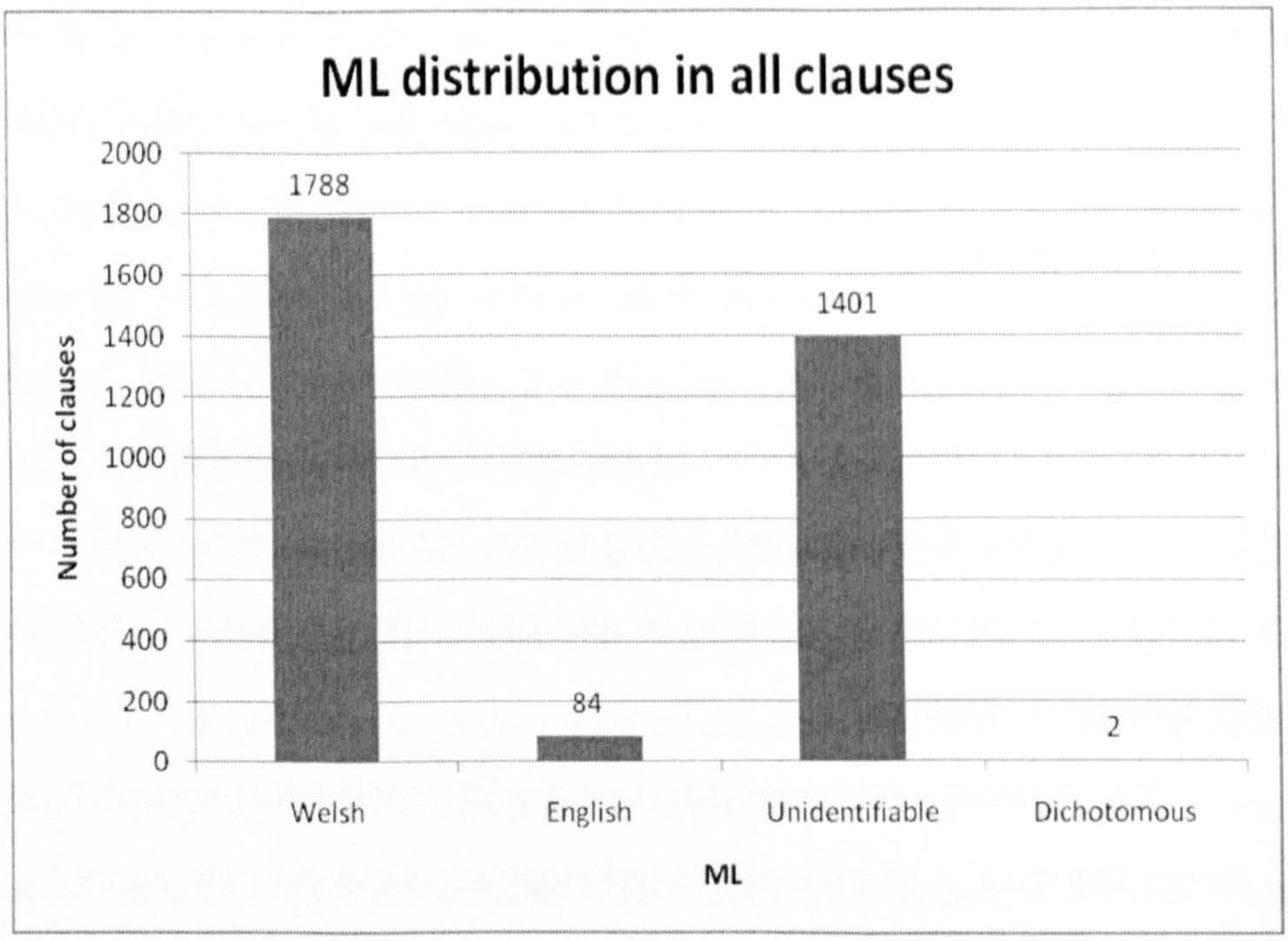


Figure 1. The Matrix Language distribution across all the clauses in the dataset.

Of the 3275 clauses uttered by all six speakers, 1874 or 57.22% had an identifiable ML, by which I mean were either identified as having a Welsh, English or Dichotomous ML. Of these clauses in which an ML was identifiable, 1788 (95.41%) had a Welsh ML, 84 (4.48%) had an English ML and two (0.11%) had a dichotomous ML. The remaining 1401 clauses (42.78% of all clauses) analysed could not have an ML identified, as there was insufficient information within those clauses on which to test the MOP and SMP.

In clauses where an ML is identifiable, Welsh is the most frequent provider of morphosyntax, being the ML in 95.41% of such clauses in these data. Note that there are slightly more clauses in the data which have an identifiable ML than those which do not have an identifiable ML.

I now look at the ML distribution across all monolingual clauses. These results are shown in figure 2 below.

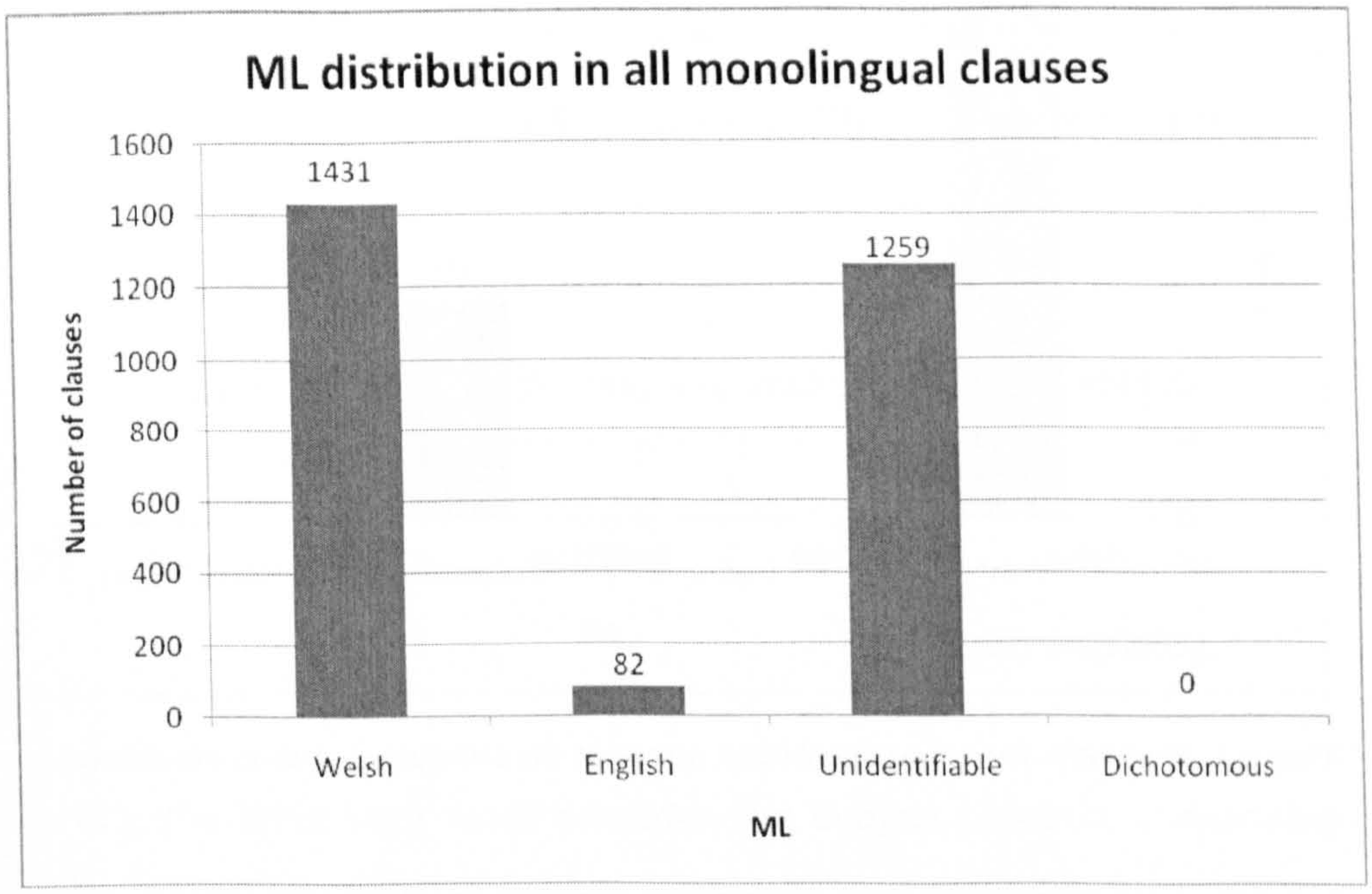


Figure 2. The Matrix Language distribution across all the monolingual (Welsh, English or either) clauses in the dataset.

There are 2772 monolingual clauses in the dataset. As figure 2 shows, an ML is identifiable in 1513 clauses (54.58% of monolingual clauses). Of these, 1431 (94.58%) monolingual clauses have a Welsh ML, whilst 82 (5.42%) have an English ML. Welsh is again the primary source of ML in monolingual clauses. 1259 monolingual clauses (45.42%) do not have an identifiable ML.

I now turn to look at the ML distribution across all bilingual clauses. These results are shown in figure 3 below. There are 503 bilingual clauses in the dataset.

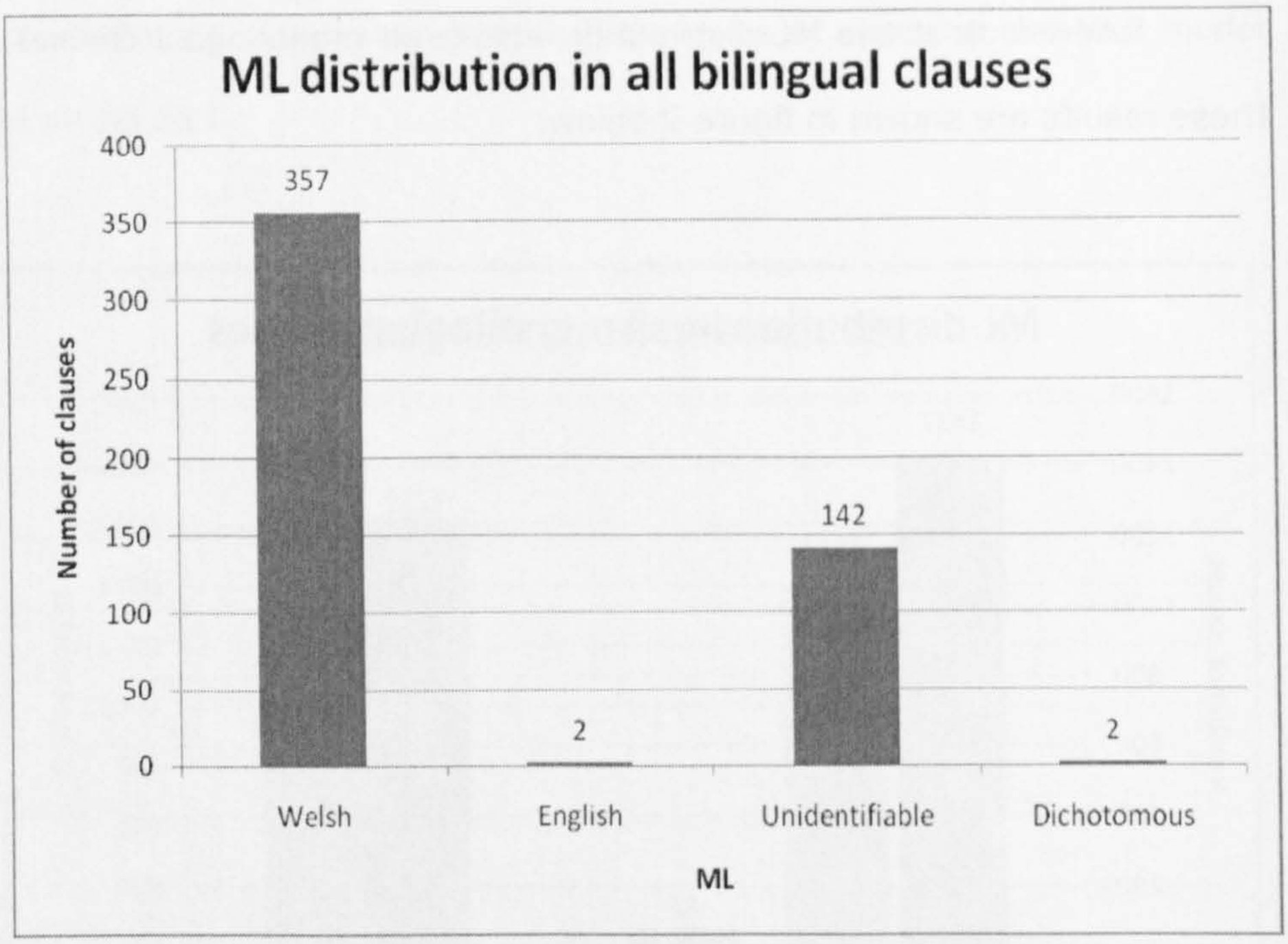


Figure 3. The Matrix Language distribution across all the bilingual clauses in the dataset.

An ML is identifiable in 361 bilingual clauses (71.77%). Of these bilingual clauses in which an ML was identifiable, 357 (98.89%) have a Welsh ML, two (0.55%) have an English ML and two (0.55%) have a dichotomous ML. Welsh is almost always the ML source in bilingual clauses in these data.

An ML was identifiable more frequently in bilingual clauses than in monolingual clauses in these data (71.77% to 54.58%), but the overall ML distribution pattern is the same in clauses in which an ML is identifiable, regardless of linguality. In both bilingual and monolingual clauses, Welsh is the most prominent source of clausal morphosyntax.

ii. ML distribution in finite clauses

I now examine the ML distribution in finite clauses (disregarding clause linguality for the moment). Example (1) below shows a Welsh

monolingual finite clause with a Welsh ML and example (2) shows an English monolingual finite clause with an English ML.

(1)	wnes	i	'm	cael	hi	tro	cynta
	do.1S.PAST	1S	NEG	get.NONFIN	3Sf	time	first
	de						
	TAG						

"I didn't get it the first time, you know." [DAV6-DAN430]

(2) *oh* there 's nothing wrong with you

[FUS6-ANT283]

In (1), the finite verb *wnes* precedes the subject pronoun *i*, indicating Welsh as the provider of morpheme order; this indication is supported by the order in the head/modifier NP *tro cynta* 'first time', where the adjective *cynta* follows the head noun *tro*, showing Welsh-predominant MH word order. The inflection on *wnes* matches the person of the subject and thus indicates that Welsh also provides the outside late system morphemes in the clause. As both MOP and SMP thus point to Welsh, that language is identified as the ML. In (2), the subject *there* precedes the verb *is*, indicating English-predominant SV order. The morphology of the finite verb (*is*) is also English. Thus the MOP and the SMP both point to English as the source of the ML.

There are 1862 finite clauses in the dataset, representing 56.85% of all the clauses in the dataset. An ML was identifiable in all of those clauses. 1778 (95.49%) finite clauses had a Welsh ML. This accounts for 54.92% of the clauses in the dataset overall. English was the ML in 82 (4.40%) finite clauses, accounting for only 2.50% of all clauses in the dataset. There were two finite dichotomous ML clauses

(the only two dichotomous ML clauses in the data), accounting for 0.11% of finite clauses and 0.06% of all clauses in the dataset. The ML distribution in finite clauses (irrespective of linguality) is illustrated in figure 4 below.

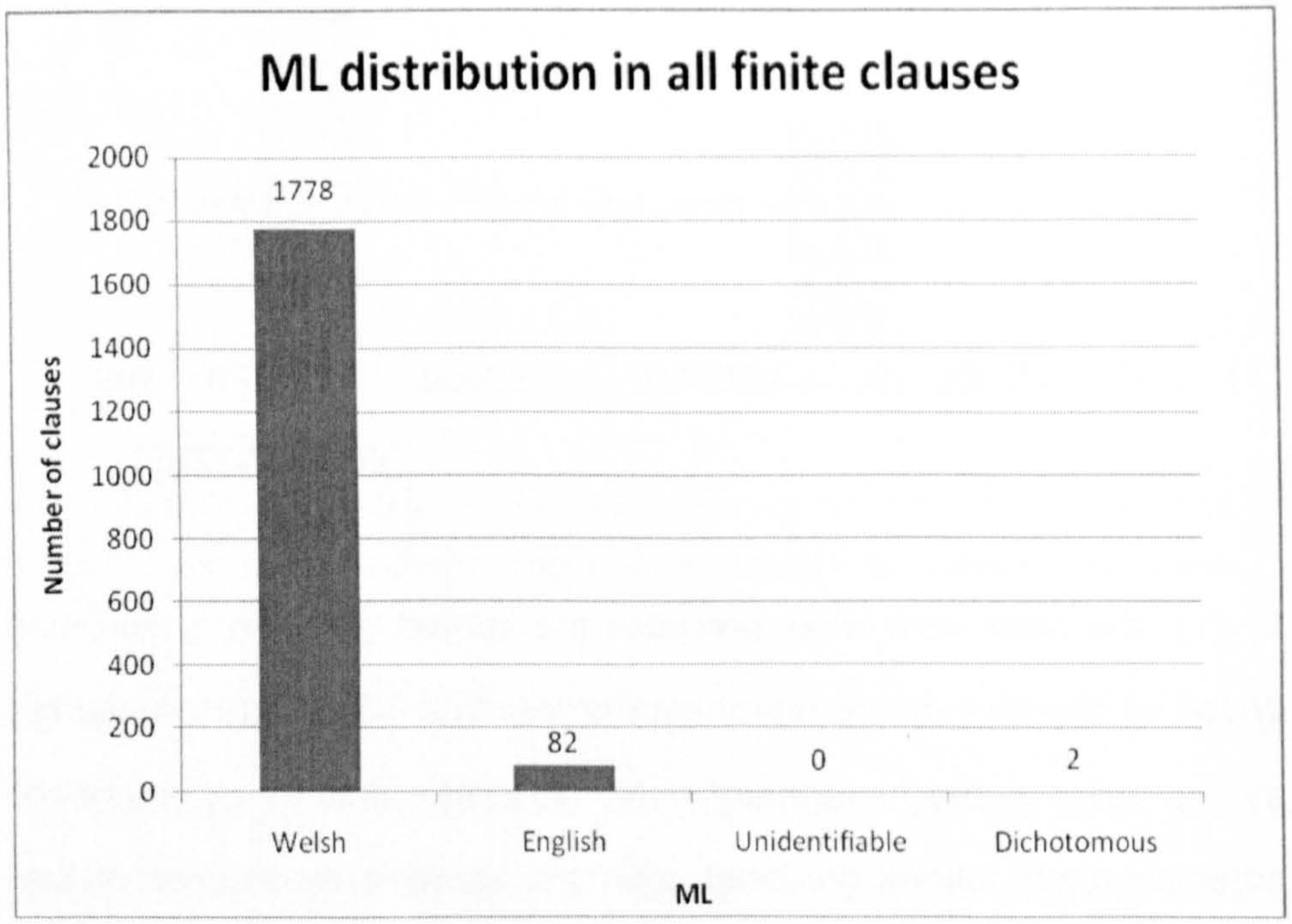


Figure 4. The Matrix Language distribution across all the finite clauses in the dataset (monolingual and bilingual combined).

I now discuss further the ML of finite clauses by subcategorising them into monolingual clauses and bilingual clauses. The results for the ML distribution in finite clauses, comparing monolingual and bilingual clauses, are shown in figure 5 below.

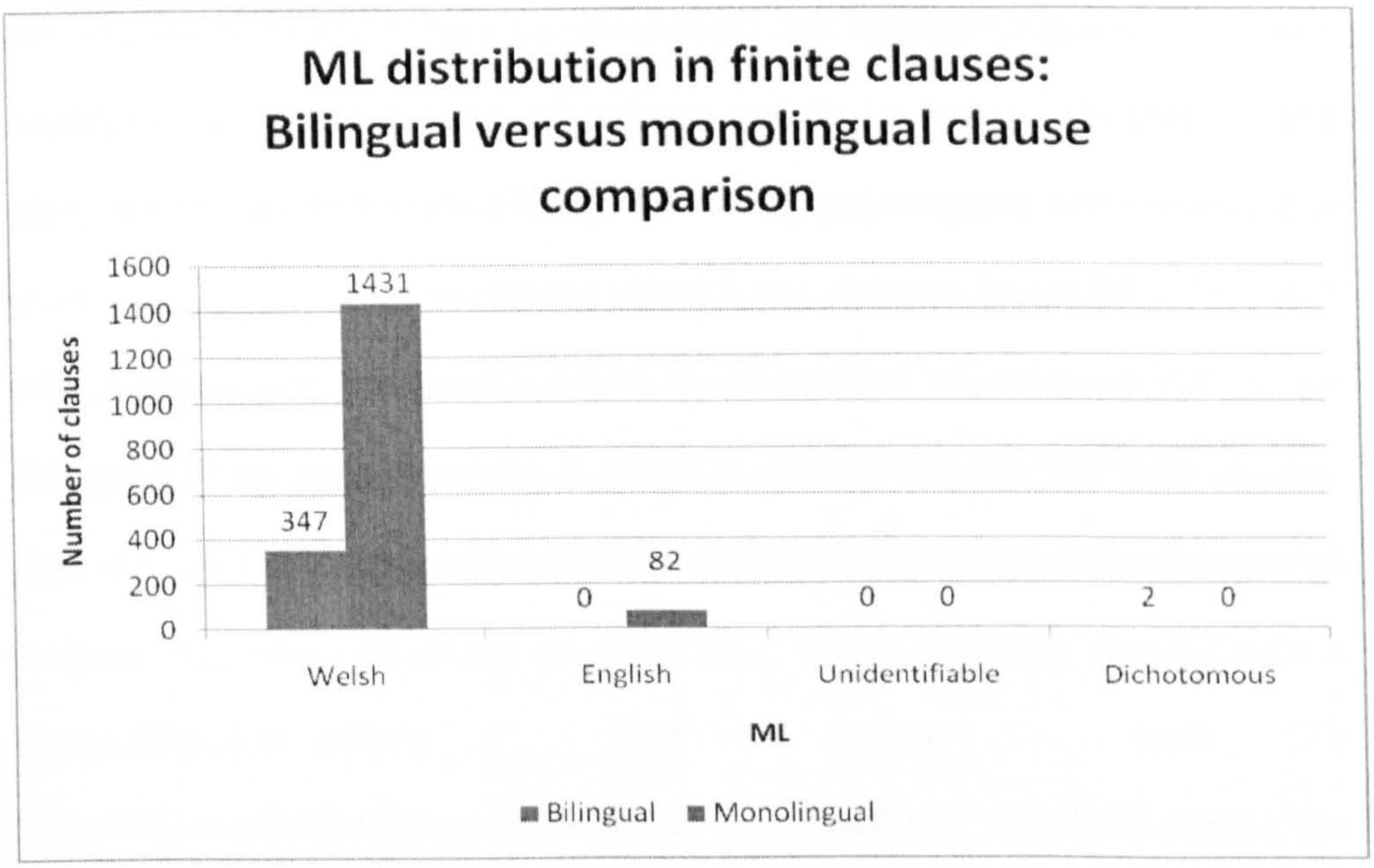


Figure 5. The Matrix Language distribution across all finite clauses in the dataset, comparing the distribution in bilingual clauses to the distribution in monolingual clauses.

First I discuss monolingual finite clauses. Of the 1862 finite clauses in the data, 1513 (81.26%) are monolingual whilst 349 (18.74%) are bilingual. The monolingual clauses are distributed in terms of ML as follows. 1431 (94.58% of monolingual clauses) clauses had a Welsh ML while 82 (5.42%) had an English ML. The great majority of monolingual finite clauses in the data have a Welsh ML.

These results show a number of things. First, in monolingual clauses the language which is the source of the morphemes in those clauses is always also the language which is the source of the ML. Second, there are more monolingual Welsh ML clauses than there are monolingual English ML clauses, so Welsh is apparently the language generally selected for the discourse frame. There is a uniformity to these data that belies the sociolinguistic differences between the six speakers as noted in the previous section. This could be ascribed to the fact that all six speakers have roughly equal ability (that is to say, fluency) in both Welsh and English, and so producing monolingual clauses which

adhere to the grammar of the respective languages is unproblematic for these speakers. However, these results do not indicate what language will provide the grammatical frame for CS. To identify this, clauses with CS, that is bilingual clauses, need to be examined.

I now turn to the bilingual finite clauses in the dataset, the results of which are illustrated in figure 5 above. Examples (3) and (4) below show two bilingual finite clauses with a Welsh ML.

(3)	mae	raid	bod	gynno	fo	
	be.3S.PRES	necessity	be.NONFIN	with.3Sm	3Sm	
	ryw	fath	o	<u>attitude</u>	<u>problem</u>	de
	some	kind	of	attitude	problem	TAG

"He must have some kind of attitude problem, eh." [DAV6-HEC261]

(4)	mae	<u>Americans</u>	yn	mwy	##	<u>commercial</u>
	be.3S.PRES	Americans	PRT	more		commercial

"Americans are more commercial." [FUS27-LIS437]

In (3) the finite verb *mae* precedes the subject *raid* 'necessity' in Welsh-predominant VS morpheme order. The verb morphology of *mae* is also Welsh. Welsh is thus identified as the source of the ML. The head/modifier NP *attitude problem*, which has English-predominant MH order and contains only English morphemes, is an EL island. In (4) the subject *Americans* follows the finite verb *mae*, again indicating Welsh VS order; *mae* has Welsh outside late system morphemes, the SMP also points to the ML being sourced from Welsh.[102] In addition, *mae* is a

[102] Note that the plural *–s* suffix on *Americans* is found in the Welsh dictionary and does not therefore violate the predictions of the SMP.

singular verb whilst *Americans* is a plural subject. This shows anti-agreement, which is an obligatory aspect of Welsh (e.g. Roberts 2005:44) but does not occur in English, again reinforcing the identification of Welsh as the ML of (4).

There were 349 bilingual finite clauses in the dataset. Of all the bilingual finite clauses, the vast majority had Welsh ML: 347 clauses (99.43% of the total bilingual finite clauses). No bilingual clause had an English ML. The remaining two (0.57%) bilingual finite clauses had a dichotomous ML.

I infer two things from these results. First, Welsh is clearly the predominant source of the ML in bilingual finite clauses produced by these speakers. Second, English is never the (sole) source of the ML in the bilingual finite clauses of these speakers as recorded in these data. Also note that the two dichotomous ML clauses found in the dataset are bilingual.

It is not surprising that no English ML finite clauses in the data contain code-switching. Such clauses are also very rare in previous studies involving application of the MLF model to Welsh-English data: see e.g. Deuchar (2006), where only 4 out of 163 (2.45%) bilingual clauses had an English ML, compared to the 141 clauses out of 163 (86.50%) which had a Welsh ML. Previous studies involving other language pairs, as discussed in chapter 2, show that bilinguals commonly have an asymmetry in their output, in that they produce bilingual clauses with one language as the source of structure but not with the other language as the source of structure (e.g. Smith 2006). Why should it apparently be the case that one of a bilingual's languages is more likely to be the morphosyntactic frame in bilingual clauses than the other?

The reason may be related to Grosjean's (1998 etc.) notion of language mode, as discussed in chapter 2 section 7. The speakers I have studied are Welsh-English bilinguals, who can speak Welsh or English with other Welsh-English bilinguals, but only English with English monolinguals. No Welsh adult monolinguals presumably exist, as I discussed in chapter 1. Thus, Welsh-English bilinguals talking to English monolinguals will have English highly-activated and Welsh less activated. The linguality of such clauses would have to be monolingual English, and, as shown in these data, monolingual clauses from Language A also always source the ML from Language A. Welsh-English bilinguals talking to English monolinguals must use English ML clauses to be understood.

However, when talking to other Welsh-English bilinguals, these speakers' Welsh will become more activated and their English will probably become less activated than in the English monolingual mode. If Welsh becomes more activated, it is more likely to be used as a source of ML. For these speakers, use of English cues monolingual mode and English morphosyntax, while use of Welsh cues bilingual mode and Welsh morphosyntax.

A glance at transcripts of data from L1 English speakers from elsewhere in the Siarad corpus suggests that bilingual English ML clauses are also very rare in their speech, suggesting that this trend is uniform across Welsh speakers in general, not just those who have Welsh as L1. A more in-depth and quantitative study of clauses produced by L1 English Welsh-English bilinguals would verify this one way or the other, but this is beyond the scope of this thesis, and is an area for future research.

iii. ML distribution in nonfinite clauses

I have shown that an ML is identifiable in all finite clauses in the dataset. The MLF model can be successfully tested on finite clauses (assuming that my data sample is representative[103]). I will next look at clauses which do not have a finite verb, to see how frequently an ML is identifiable in such clauses.

First I will focus on nonfinite clauses.[104] Examples (5) and (6) show nonfinite clauses from the dataset (both are bilingual). (5) has a Welsh ML and (6) has an English ML.

(5)	fod	yna	bethau	Egyptian	yno
	be.NONFIN	there	things	Egyptian	there

"That there are Egyptian things there."

[FUS27-MAB248]

[103] Earlier in this chapter I suggested that my sample of six speakers is reasonably representative of speakers in the data as a whole. There was a variety of male/female, northern/southern dialects, young/middle aged, etc., though potentially, since the speakers are all L1 Welsh, they are not representative of L2 Welsh bilinguals. I do not expect, however, that the MLF model would have any less success when tested on clauses produced on such speakers.

[104] It will be recalled that it was stated in chapter 3 that a clause with a visible nonfinite verb and a visible subject were usually considered to have a null finite auxiliary verb for the purposes of this analysis. Such clauses were therefore considered finite clauses. Clauses of this type will be examined in greater detail in the next chapter.

(6) [Ti angen] wneud fath â #

2S need do.NONFIN kind with

rhedeg exercise a ballu beth bynnag

run.NONFIN exercise and suchlike whatever

wedyn

after

"[You need to] do, like, running exercise and so on anyway afterwards." [FUS27-MAB938]

In (5) the nonfinite verb is *bod*. The NP *pethau Egyptian* 'Egyptian things' is bilingual, and follows the Welsh-predominant order of modifier following head. Thus this identifies Welsh as supplying structure for the clause, and Welsh is the ML. The clause in (6), meanwhile (excluding the material enclosed in square brackets, which is part of the larger complex clause in which [6] appears but I have included here for information), has the nonfinite verb *wneud*, which does not have any outside late system morphemes, so the SMP cannot be applied. The NP *rhedeg exercise* 'running exercise', where the head noun *exercise* follows its modifier, the verb-noun *rhedeg*, is bilingual: here the word order is head following modifier, which is an English-predominant word-order. Note that here an interpretation of *rhedeg* as a nonfinite verb, with *exercise* as its object (a word order that is common to both Welsh and English), is probably not valid in this instance, since *rhedeg exercise* is clearly the object of the nonfinite verb *wneud* according to both clause structure[105]

[105] Borsley et al (2007:70) distinguish between the use of Welsh verb-nouns as nonfinite verbs and as nouns. *Rhedeg* here is clearly a noun because it, as a phrase with *exercise* can be preceded by the definite article *y* (*y rhedeg exercise*), but that phrase cannot be focused to the front of the main clause of this complex clause without the insertion of a clitic *ei*—a property of nonfinite verbs—which is ungrammatical to a native speaker (*Rhedeg exercise ti angen [ei] wneud) (2007:73).

and the prosody of the sentence as uttered by the speaker[106]. I interpret both *rhedeg* and *exercise* here as being nominal rather than verbal. The ML of the clause in (6) is therefore identified as English.

There are 450 nonfinite clauses in the data, which is 13.74% of all the clauses in the data. Of these, four (0.89%) have an identifiable ML. Figure 6 below illustrates the ML distribution in nonfinite clauses.

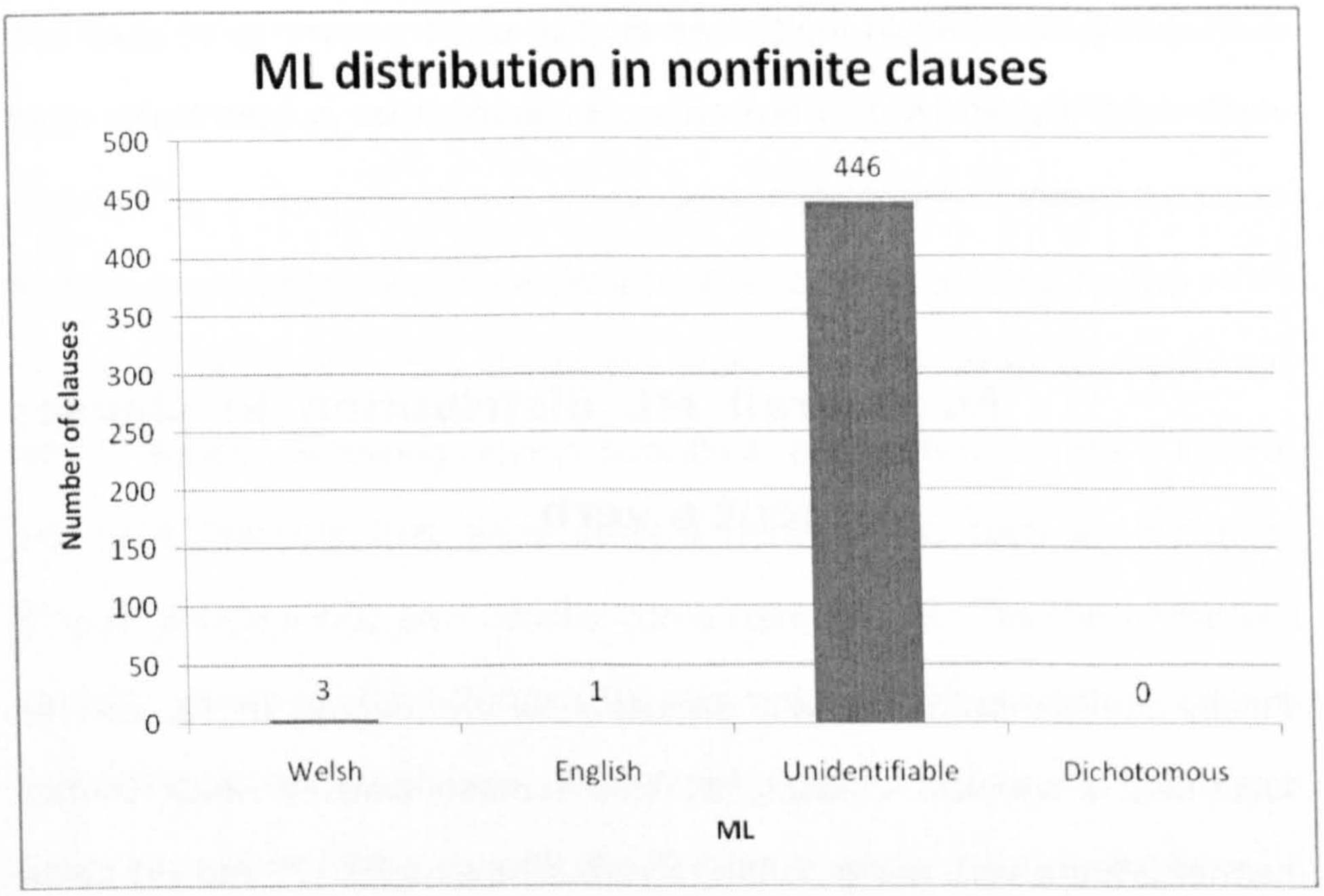


Figure 6. The Matrix Language distribution across all the nonfinite clauses in the dataset (monolingual and bilingual combined).

As the figure shows, of the nonfinite clauses in which an ML is identifiable, three clauses (75%) have a Welsh ML and one (25%) has an English ML.[107] 446 out of 450 nonfinite clauses (99.11% of all

---

[106] In addition I have consulted the original speaker of this clause (personal communication), and she has confirmed, upon listening to the recorded utterance again, that she produced *rhedeg exercise* as a Noun Phrase rather than a nonfinte verb + object construction.

[107] Note that all four of the nonfinite clauses for which an ML can be identified are bilingual, and that the evidence used within each clause for identifying the ML was a bilingual

nonfinite clauses; 13.62% of all clauses in the data) do not have an identifiable ML.

The reason for the difficulty in identifying a ML in such clauses is, as described in chapter 3, first, because of the absence of a visible finite verb, which means the SMP cannot be applied, and, second, because the only testable word-order criterion is NP head/modifier position of certain types. This is because the word-order locus of the nonfinite verb is not sufficiently dissimilar in Welsh and English constructions to be usable as distinctive identifiers of word-order, meaning that subject/verb order cannot be tested on nonfinite clauses.

iv. Overall ML distribution in clauses without a verb

Finally in this section I will discuss how frequently an ML can be identified in verbless clauses. By this I mean clauses which contain neither a finite verb nor a nonfinite verb.[108] Examples (7) and (8) below are of a Welsh ML verbless clause and an English ML verbless clause from the data respectively.

(7) y *boy* P.R. famous yna
DET boy P.R. famous there
"That famous P.R. bloke." [DAV6-HEC188]

---

head/modifier NP. I will return to the issue of the various types of NP usable as evidence for clause ML in section 4 below.

[108] Recall that my analysis includes clause fragments.

(8) fath â <u>farmhouse</u> peth

kind with farmhouse thing

"Like, a farmhouse thing." [FUS27-MAB905]

In both these examples the ML is identifiable because of the presence of a head/modifier NP. In (7) the NP *y boy P.R. famous* 'the famous P.R. boy' contains the English morphemes *P.R.* and *famous* following *boy*. The word order in the NP is HM, since *P.R. famous* modifies *boy* (*boy* here means 'guy, bloke' and is a lexical transfer into Welsh from English[109]), a typically Welsh word order—the same morphemes in an English word order would be *famous P.R. boy*. According to the MOP, Welsh is identified as the clause ML. Note that the determiner *y* in this NP is Welsh, following Myers-Scotton's predictions in the Uniform Structure Principle that early system morphemes, such as Welsh or English determiners, also usually come from the ML ("as the unmarked choice", Myers-Scotton 2002:120). Example (8), meanwhile, contains the bilingual NP *farmhouse peth* 'farmhouse thing', where the Welsh head noun *peth* 'thing' follows the English modifier noun *farmhouse*, in English-predominant MH word order. Applying the MOP to this clause, then, indicates that English provides clause word order, and is identified as the ML source.

There were 963 verbless clauses in the dataset, representing 29.40% of all clauses in the dataset. Only 8 verbless clauses (0.82%)

---

[109] In the transcription this word *boy* has been assigned the language tag @0, signifying a word which could either come from Welsh or English (that is, the word is found in both Welsh and English dictionaries). The consensus decision of the project team was to transcribe all @0 words with English orthography, for consistency (rather than making the arbitrary choice that e.g. some transferred words from English are 'more Welsh' than others and thus should be transcribed using Welsh orthography). The Welsh orthography of this established borrowing would be *bol*.

have an identifiable ML. Figure 7 below illustrates the ML distribution in verbless clauses.

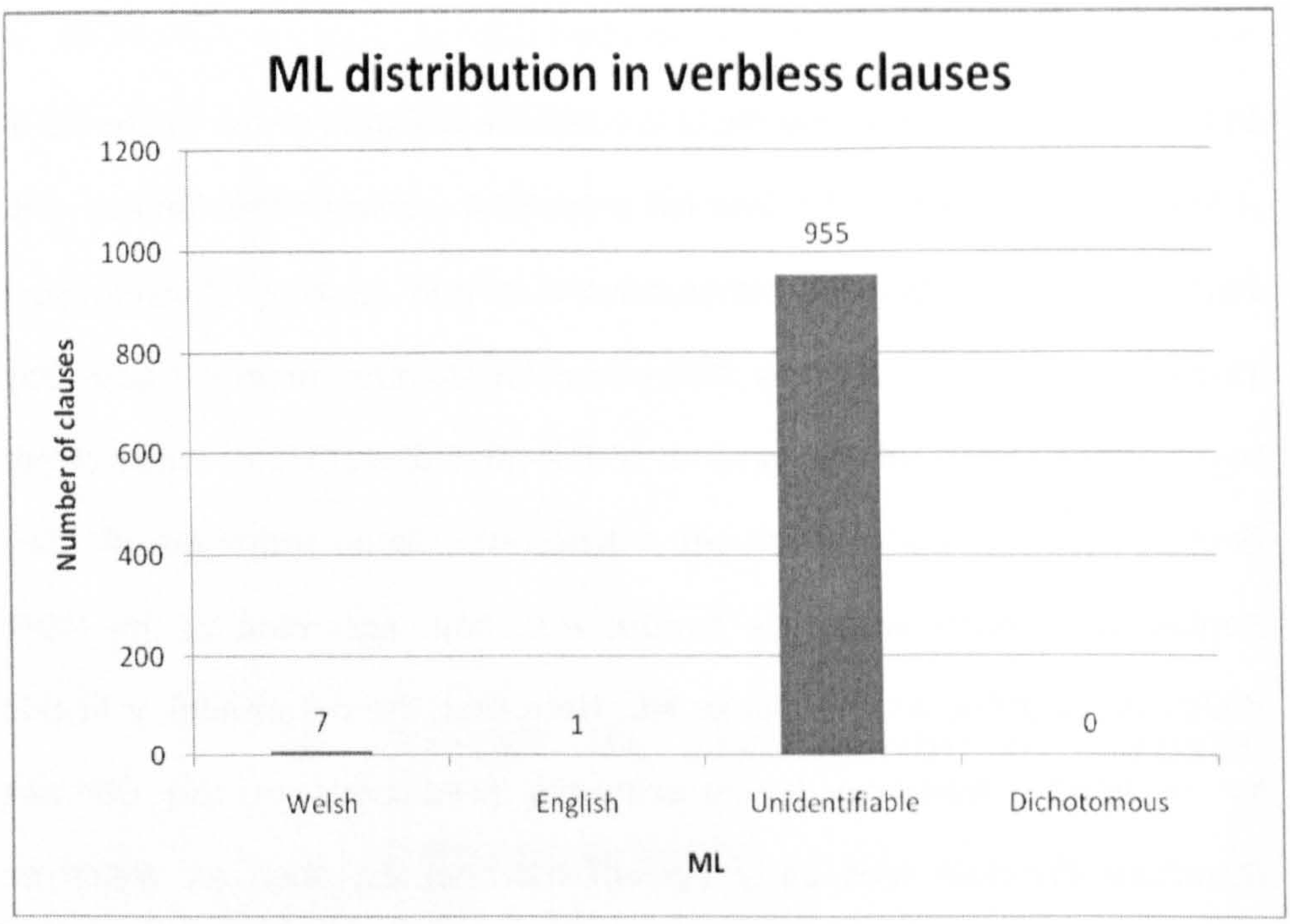


Figure 7. The Matrix Language distribution across all the verbless clauses in the dataset (monolingual and bilingual combined).

Of the verbless clauses with an identifiable ML, 7 (87.50%) had a Welsh ML and one (12.50%) had an English ML. The remaining 955 verbless clauses (99.17% of verbless clauses) in these data do not have an identifiable ML.

As in nonfinite clauses, then, an ML is frequently not identifiable in verbless clauses: less than one percent verbless clauses have an identifiable ML. Compare this to my application of the model to finite verbs, where all clauses in this dataset had an identifiable ML. What this indicates is that if a given clause in Welsh-English data lacks a finite verb, it is quite unlikely that an ML will be identifiable in that clause. I will return to this issue in the next chapter, when I discuss the success of applying the MLF model to these data.

In this section I have given the results of the analysis when considering the dataset as a whole, as well as for finite, nonfinite and verbless clauses. An ML is always identifiable in Welsh-English finite clauses but less frequently identifiable in Welsh-English clauses lacking a finite verb.

In the next section I will compare and contrast the ML frequency distribution in individual speakers within the dataset.

*b. **ML distribution according to speaker***

In this section I analyse the ML distribution in the output of individual speakers, in order to compare them to the results of the dataset taken as a whole.

Figure 8 below shows these results, counting both monolingual and bilingual clauses together. Given the difficulty of identifying an ML in clauses which are not finite, I only examine the distribution in finite clauses here.

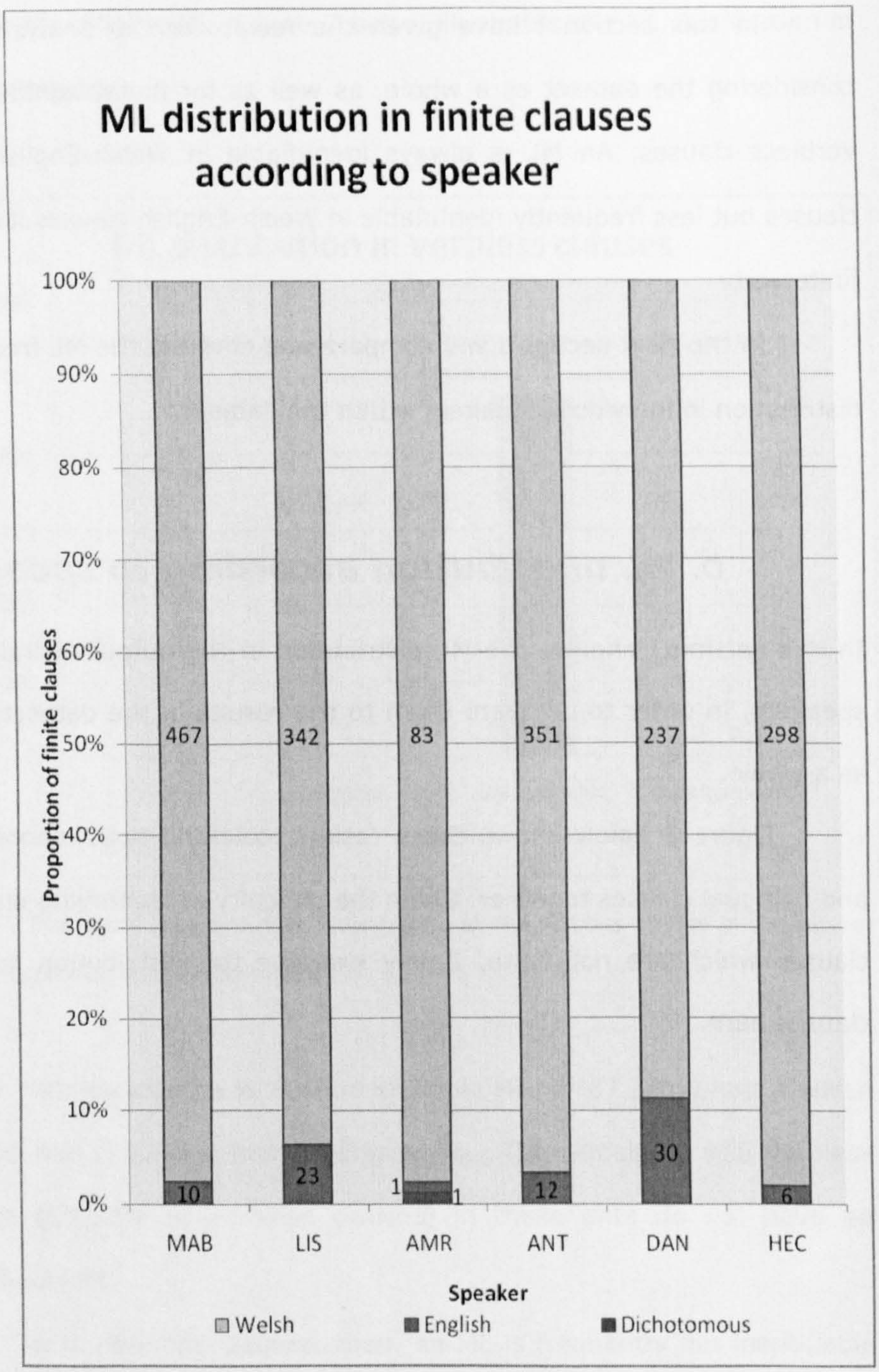


Figure 8. The Matrix Language distribution across speakers in all finite clauses (monolingual and bilingual) in the dataset.[110]

[110] The numbers on the bars themselves represent the number of clauses produced by each speaker with either a Welsh, English or dichotomous ML. The relative size of the bars represent the proportion of each speaker's finite clause output which had that language as ML.

The general pattern found when the dataset was examined as a whole is repeated when the data are examined for each speaker separately. Welsh is the ML in the great majority of clauses, and English only in a small minority of clauses.

Hector produced the greatest proportion of Welsh ML clauses in his total output of finite clauses: 298 out of 304 (98.03%) were Welsh ML compared to 6 clauses with English ML (1.97%). Daniel produced the lowest proportion of Welsh ML clauses, with 237 out of 267 clauses (88.76%), also a high proportion. All speakers produced a high proportion of Welsh ML clauses, and the difference between them is very small: a Chi-square test for independence shows $p>0.05$ ($p=46.08$, $df=5$), showing that there is not a significant relationship between the frequency of Welsh ML and English ML clauses produced by the 6 speakers[111].

The speaker who produced the largest proportion of English ML clauses in the dataset was Daniel, with 30 out of 267 (11.24%) English ML clauses. Amranwen produced the lowest proportion of English ML clauses: one out of 85 (1.18%). The next highest proportion Of English ML to Daniel was Lisa, with 23 out of 365 clauses (6.30%),[112] and the other speakers produced very few English ML clauses (Antonia 3.31%; Mabli 2.09%; Hector 1.97%). All speakers produced few English ML clauses, and again there is no significant relationship between the speakers (see Chi-square results in the previous paragraph).

---

[111] The two dichotomous ML clauses (produced by Amranwen and Mabli) were discounted from this test. Note that since Amranwen only produced 1 English ML clause, this is not ideal for a Chi-square test (all cells are expected to have a value of at least 5). Nevertheless, even if Amranwen is discounted from the analysis altogether, the relationship between the remaining five speakers is still not significant ($p>0.05$; $p=42.52$, $df=4$).

[112] Lisa is also the speaker with the second lowest proportion of Welsh ML clauses (342 out of 365, 93.70%) in the dataset, although it is still a high proportion.

Note that, when the monolingual English ML clauses of each speaker are specifically examined, it can be seen that the majority of them are instances of direct quotation, such as when a speaker is referencing what another person has said to them in English, or when they are quoting from a text. Daniel and Hector both brought tabloid newspapers into the recording environment, which they proceeded to quote from periodically during the recording. It may be that, for these speakers, monolingual clause production is governed by discourse context rather than by any linguistic characteristic (e.g. language 'dominance') of a given speaker.

I now move to discussing the bilingual clauses these speakers produce. As I noted above when I discussed the results taken from the dataset as a whole (see section 3.a.i), there are no bilingual finite clauses in the data with English ML. All six speakers here, when they construct finite bilingual clauses, with the exception of the two dichotomous ML clauses, source their clause morphosyntax from Welsh.[113] This shows that Welsh is clearly the discourse-dominant language for all the speakers analysed, and I predict that such a trend would prove to be the case for most of the other speakers in the overall data if they were analysed in the same manner. What is of interest is that all six speakers fit the observed pattern of Welsh as the (almost) universal provider of structure in bilingual clauses. Whether they are younger or older speakers, male or female, or born in the south or the north of Wales, they all use Welsh as ML in bilingual clauses.

---

[113] Bilingual clauses do not form a large proportion of the finite clause output of any speaker. Lisa has the highest proportion, with 111 out of 365 (30.41%) of her finite clauses being bilingual. Daniel has the lowest proportion, with 28 out of 267 (10.49%) of his finite clauses being bilingual. These speakers therefore, as described when discussing the results from the dataset taken as a whole, produce monolingual clauses more frequently than bilingual clauses, which could be taken to indicate a preference for monolingual Welsh in clause production.

There is a distinct difference between ML distribution in monolingual finite clauses compared to bilingual finite clauses. In monolingual finite clauses, there is a distinct bias towards Welsh as ML, though English is used as ML in a small proportion of clauses, usually as direct speech. In bilingual finite clauses, however, Welsh is always the ML (except for the two with dichotomous ML).

In this section I have discussed the ML distribution in individual speakers, showing that it does not differ greatly from the distribution pattern found in the dataset as a whole. Different speakers, then, do not show much variation.[114] Clearly Welsh is the predominant provider of morphosyntax for clauses, monolingual or bilingual, produced by these speakers, and dichotomous ML clauses are extremely rare.

In the next section I will examine clauses in the dataset which contain head/modifier NPs, before proceeding to an analysis of the only dichotomous ML clause found in these data.

4. Head/modifier Noun Phrases in dichotomous ML clauses

In chapter 3 I hypothesised that dichotomous ML clauses in the data would show convergence. Earlier in this chapter (section 3.a.i), I stated that one clause in the dataset was identified as having a dichotomous ML. As I will show below, this clause has a dichotomous ML because of a disagreement between the ML as indicated by the word order of the

---

[114] This appears to be the case for the Welsh-English data, at least. It would be of interest, in future research, to compare the production of Welsh-English bilinguals in Wales to Welsh-Spanish bilinguals in Patagonia, to see if, in a different context, those speakers show variation when extralinguistic variables are taken into consideration.

subject and finite verb and the word order of the head and modifier in an NP in the clause.

As I described in chapter 3 (section 6.c.ii), there are certain configurations of head/modifier NP which can be used to identify the ML of a Welsh-English clause as being sourced from Welsh or English. In this section I will discuss the various configurations of head/modifier NP possible in Welsh and English, but focus those that identify a clause in which they appear as having Welsh, English or dichotomous as ML, providing examples of the different configurations found in the analysed dataset.

I will then proceed to discuss how the head/modifier NPs found in the one dichotomous ML clause in these data is different to other NPs found in the data, and evaluate whether or not this clause shows word-order convergence.

a. Head/modifier NPs that can be used to identify clausal ML

As described in chapter 3 (section 6.c.ii) I described how Welsh has two options for head/modifier NP word-order, HM but also MH, whereas English has only one option, MH. If Welsh-English data were to feature NPs which show a deviation to this pattern, such as MH order NPs with Welsh morphemes, or HM order NPs with English morphemes, this could be interpreted as evidence of the convergence process. In this section I will first outline the different configurations of head/modifier NPs which can be used to help identify the ML of the clause in which they occur, both the monolingual and the bilingual types. I will then give examples of each type, taken from the analysed dataset.

Assuming a monolingual head/modifier NP construction has one head morpheme and one modifying morpheme, and that both morphemes come from the same language, either Welsh or English, there are logically four configurations which can occur:

A. Welsh head + Welsh modifier, e.g. **gwin coch**

B. English modifier + English head, e.g. **red wine**

C. English head + English modifier, e.g. **wine red**

D. Welsh modifier + Welsh head, e.g. **coch gwin**

Monolingual NPs of the types A and B here are in the usual word order for that language, whilst types C and D show an alternative word order which is not usual for that language. Types A and B could be interpretable in any given clause, without looking at any other information in that clause, as either a NP with word order from the ML, or as an 'island' of morphemes from the Embedded Language, with word order also from the EL. They are not by themselves useful indicators of clause ML.

However, because an NP of type C or D has morphemes from one language but (apparently) word order more common to the other language, it cannot be identified as an EL island, and so must be analysed as having word order sourced from the ML. As such, types C and D monolingual NPs <u>can</u> be used to identify the ML of a clause in which they occur.

Turning to bilingual NPs, there are again four configurations which could logically occur:

E. Welsh head + English modifier, e.g. **gwin red**

F. English head + Welsh modifier, e.g. **wine coch**

G. Welsh modifier + English head, e.g. **coch wine**

H. English modifier + Welsh head, e.g. **red gwin**

In each of the four types, one morpheme in the NP will come from the ML and one will come from the EL. Types E and F have the same word order, modifier following head (HM), which I identified above as being the Welsh word order for head/modifier NPs. In NPs of these types, Welsh can conclusively be identified as the language supplying word order, and so the NP can be used as evidence to identify the clausal ML. In types G and H, meanwhile, the word order found is modifier preceding head (MH), which is an English order (though also an alternative Welsh word order), so can also be used to identify the ML of the clause.

Note, nevertheless, that whilst it is possible to identify which language provides word order for these four types of bilingual NP, that does not mean the ML of the clause in which they occur will be same the language as that indicated by the NP word order. If the material in the rest of the clause, i.e. subject/verb order and agreement, do not indicate the same language, then the clause will be identified as having a dichotomous ML.

I will now give examples of first the monolingual and then the bilingual NPs which can be used as evidence to identify the ML of the clause in which they appear.

Types A and B monolingual NPs source morphemes and word order from the same language. Type A NPs have Welsh morphemes in the order appropriate for Welsh, and type B NPs have English morphemes in the order appropriate for English. Below are two examples, (9), which is a type A NP, and (10), which is a type B NP. Note that the head/modifier NPs in these examples is in bold.

(9) ond um fel **peth cymunedol**
but IM like thing communal
"But, um, like a communal thing." [FUS6-ANT67]

(10) *just* **<u>wishful</u>** **<u>thinking</u>** yn y diwedd
just wishful thinking in DET end
"Just wishful thinking in the end." [FUS27-LIS119]

In (9) the NP *peth cymunedol* "communal [or community] thing" has Welsh morphemes in Welsh HM word order, whilst in (10) the NP *wishful thinking* has English morphemes in English MH word order. Wherever they appear, they cannot positively identify the ML as being specifically Welsh or English, because there is insufficient information present to indicate whether or not the NP word-order is being provided by the ML or the EL. Type A and B NPs occur quite frequently in the dataset, but they are of limited interest here, since they are not useful as indicators of ML, so I will not discuss them further.

Type C monolingual NPs have morphemes from English but word order from Welsh. An example is given in (11) below.

(11) a oedden nhw 'di gael rywun #
and be.3PL.IMP 3PL PRT.PAST get someone
'di sgwennu enw fi efo ***symbols*** ##
PRT.PAST write name 1S with symbols
Egyptian [xx][115]
Egyptian
"And they'd had someone who'd written my name with Egyptian symbols." [FUS27-MAB616]

Applying the SMP to this clause identifies Welsh as ML on the basis of the finite verb *oedden*. The MOP identified Welsh as providing subject/verb word order (VS, where *oedden* precedes the subject pronoun *nhw*). The MOP can also be tested on the head/modifier NP *symbols Egyptian* 'Egyptian symbols' in this clause: the NP has English morphemes in Welsh-predominant word order, with the modifier *Egyptian* following the head noun *symbols* (which could be an English or Welsh lexeme). *Symbols Egyptian* is not an EL island, because the inserted morphemes do not have word order appropriate for the language of those morphemes. Welsh is clearly providing the morphosyntax of this clause, so I identify it as the ML. I interpret the English morphemes *symbols* and *Egyptian* to be separate switches with meanings similar to the Welsh equivalents *symbolau* 'symbols' and *Eifftaidd* 'Egyptian', which would normally be expected to occur in the HM word order found in (11).

Similar type C NPs are found in (12), which is a verbless clause, and in (13), which is nonfinite but which I interpret to have a null finite verb.

---

[115] As noted in chapter 4, the code xx as used in the CHAT system indicates uninterpretable speech. Here it is probably a single morpheme (or less).

(12) a fel <u>like</u> y **<u>list</u>** **<u>massive</u>** 'ma

and like like DET list massive here

"And, like, this massive list." [FUS27-LIS1036]

(13) hwnnw 'n **steal**-o **<u>ideas</u>** **<u>Britain</u>** i gyd

that one PRT steal.NONFIN ideas Britain all

"That one steals/stole all of Britain's ideas." [FUS27-LIS415]

The ML of (12) cannot be identified on the basis of the SMP, because there is no finite verb, and the MOP cannot be tested on subject/verb order either. But the NP *list massive* 'massive list' is testable. This NP has the English morphemes *list* and *massive* in Welsh HM order and not English MH order, which would be *massive list*. Thus Welsh is identified as the ML, because it clearly provides word order to the clause. The clause in (13) has a finite verb form of *bod* 'to be' as a null element; this is identified because of the presence of a subject (*hwnnw*), an aspectual particle (*yn*) and a nonfinite verb (*steal-o*). Thus the SMP is identifiable as Welsh. The NP *ideas Britain* 'Britain's ideas' can also be used to identify the ML, since the modifier *Britain* follows the head noun *ideas*, identifying Welsh as providing the word order for those morphemes.

Moreover, note that it can be established that the structure of the NP in this clause is two individual English insertional switches, rather than a word-for-word translation of an equivalent English head/modifier construction. *Ideas Britain* constructed using the equivalent Welsh morphemes would be *syniadau Prydain* (syniadau 'ideas' + *Prydain* 'Britain'), with Welsh HM word order. But the English morphemes *ideas* and *Britain* in English order (MH) would be **Britain ideas*, which is not grammatical in English; *Britain's ideas* or *ideas of Britain* would be. As

ideas Britain is not likely to be a direct translation of *Britain ideas*, the likely interpretation is that *ideas* and *Britain* are each inserted separately.

Type D NPs are monolingual Welsh NPs that can be used to identify the ML of the clause in which they occur. This is where the word order of the Welsh morphemes is MH, a predominantly English word order. Only one clause in the dataset contains such an NP: it is the NP is *drws-nesa pobl* 'next-door people',[116] shown as (14) below.

(14)	roedd	**drws**	**nesa**	#	**pobl**	yn
	be.3S.IMP	door	next		people	PRT
	gwneud		**sloe**	***gin***		
	make.NONFIN		sloe	gin		

"the next-door people made sloe gin." [FUS6-AMR371]

Here the modifying element *drws nesa* precedes the head *pobl*, giving English MH order rather than the expected Welsh HM order (which would be *pobl drws nesa*).[117] This NP indicates that English is the clausal ML, but this indication is at odds with the language identified as ML by the rest of the clause, which is Welsh (cf. the VS order of the finite verb *roedd* and subject *drws nesa pobl, and the* verbal morphology of *roedd*). Since there is conflicting information in the clause about which language

---

[116] This assumes *drws nesa pobl* is analysed as an NP. I discuss other interpretations of this clause structure in the next subsection.

[117] In fact, there is an embedded NP in the NP *drws nesa pobl*, namely *drws nesa*, where *nesa* modifies *drws,* in Welsh HM word order. This implies that there is a Welsh word-order NP within an English word-order NP. However, due to the likelihood that *drws nesa* is a common lexical phrase in Welsh, I suggest it is, much like *next door* in English, a lexical 'chunk' (cf. Backus 2005), i.e. a single entry in the mental lexicon which is constructed out of more than one morpheme. For this reason I interpret *drws nesa* to be a single modifier modifying *pobl*.

is ML, this clause has a dichotomous ML. I discuss it in greater detail in the next section.

I now turn to the bilingual head/modifier NPs which can be used as evidence for clausal ML. Types E and F NPs noted above are constructions where the word order is HM, which is Welsh, and in which one element is English and the other element is Welsh. Type E NPs have a Welsh head preceding an English modifier. Examples are shown in (15) and (16) below.

(15)	na	m—	mae	Prydain	yn	well	am
	no		be.3S.PRES	Britain	PRT	better	about
	pethau		**<u>independent</u>**				
	things		independent				

"No, Britain is better at independent things [= films]."

[FUS27-MAB434]

(16)	dw	i	'n [/]	dw		i	yn
	be.1S.PRES	1S	PRT	be.1S.PRES		1S	PRT
	hoffi	rhai	o	'r #	um	**llyfrau**	#
	like.NONFIN	some	of	DET	IM	books	
	<u>American</u>						
	American						

"I, I like some of the, um, American books." [FUS27-MAB469]

Here the two NPs *pethau independent* 'independent things' and *llyfrau American* 'American books' have a Welsh head morpheme (*pethau*, *llyfrau*) preceding an English modifier morpheme (*independent*, *American*). The word order of the NP is Welsh HM. The information in the rest of both clauses also point to Welsh as providing structure. In

(15) the VS order of *mae* and *Prydain* and the verbal morphology of *mae* satisfy both the MOP and the SMP and identify Welsh as being the ML. In (16) the VS order of *dw* and *i* and the morphology of *dw* also indicate Welsh as ML. In both cases, the morpheme order within the bilingual NP matches the morphosyntax of the rest of the clause, and satisfies the MLF model in identifying Welsh as the clausal ML.

Type F NPs have an English head followed by a Welsh modifier, which is again Welsh HM order. Examples are given in (17) and (18) below.

(17)	gorodd		o	gael	**alternator**	**newydd**
	need.3S.PAST		3Sm	have.NONFIN	alternator	new
	dechrau	'r	wythnos			
	start	DET	week			

"He had to get a new alternator [at] the start of the week."

[FUS6-AMR446]

(18)	mae	Nistelrooy	'di	cael	**season**
	be.3S.PRES	Nistelrooy	PRT	have.NONFIN	season
	gwael				
	poor				

"Nistelrooy has had a poor season [= of soccer]."

[DAV6-DAN654]

In (17) the NP *alternator newydd* 'new alternator' has the Welsh modifier *newydd* positioned after the English head noun *alternator*, and in (18) the NP *season gwael* has the Welsh modifier *gwael* positioned after the English head noun *season*. Both bilingual NPs have Welsh HM word order, as with type E NPs. The rest of the material in both clauses

also point to Welsh as providing structure: the VS order (*gorodd o* in (17) and *mae Nistelrooy* in (18)) and the subject/verb agreement (*gorodd* and *mae*) are from Welsh. Welsh again provides the structure for the NP as well as for the rest of the clause.

Thus, NP types E and F can both be used to reinforce the evidence for ML found in the rest of the clause in which they occur, in that the head/modifier word order within such NPs represents a third source of information to positively identify which language is providing structure, alongside subject/verb order and subject/verb agreement.

Type G and H bilingual head/modifier NPs have MH word order, which is predominantly English. Type G NPs have a Welsh modifier preceding an English head noun. An example of a type G NP is given in (19) (already shown in (6) above). The NP *rhedeg exercise* 'running exercise' is a noun-noun construction. Note that (19) contains two embedded co-ordinating nonfinite clauses; here I am looking at the first embedded clause, *wneud fath â rhedeg exercise* 'do, like, running exercise'.

(19)	[Tl	angen]	wneud	fath	â	#
	2S	need	do.NONFIN	kind	with	
	rhedeg	**<u>exercise</u>**	a	ballu		beth bynnag
	run.NONFIN	exercise	and	suchlike		whatever
	wedyn					
	after					

"[You need to] do, like, running exercise and so on anyway afterwards." [FUS27-MAB938]

In the NP *rhedeg exercise,* the Welsh modifier *rhedeg* precedes the English head noun, *exercise*, giving English MH word order.[118] Since it is a bilingual head/modifier NP, it can be used as evidence to identify the clausal ML, which identifies this nonfinite embedded clause as having an English ML. This particular clause is the only one in the dataset which has a type G bilingual NP in it.[119]

Type H bilingual NPs have an English modifier preceding a Welsh head. There is one type H bilingual NP in the dataset. It is given below as (20), where the speaker is referring to a ballgown.

(20)	and	sydd	ddim	yn	***Cinderella***	**type**	#
	and	be.REL	NEG	PRT	Cinderella	type	
	peth						
	thing						

"And which isn't a Cinderella type thing." [FUS27-MAB612]

The NP *Cinderella type peth* consists of two English modifiers (*Cinderella* and *type*) preceding the Welsh head (*peth*). The NP, like type G above, has MH word order, a predominantly-English order. It is used to identify the ML of this clause as being dichotomous, as I discuss in more detail in section 4.b below.

In this section I have identified the different types of head/modifier NP which can be used as evidence for clausal ML in Welsh/English data, and given examples of each from the data. Only two types, D and H, occur in clauses with a dichotomous ML in the analysed

---

[118] Whilst, ostensibly, this clause could be interpreted as meaning "you need to run an exercise", which would mean that *rhedeg exercise* is not actually a head/modifier NP, it is clear from discourse context that this is not the speaker's intended meaning. The two speakers at this point are talking about training for a running competition by exercising.

[119] A consequence of English not normally being the ML of bilingual clauses in these data.

data, and they are NPs in which Welsh morphemes occur in a predominantly-English word order (MH). I will now proceed to discuss these dichotomous ML clauses in detail, and analyse them for the presence of convergence.

b. Word order in NPs with a dichotomous Matrix Language

I now examine in detail the two clauses in this dataset which have a dichotomous ML. The most noteworthy point to be made is that there are practically no dichotomous ML clauses in the dataset. In the three transcripts which form this dataset, there are only two such clauses, a proportion of 0.06% of the 3275 clauses in those transcripts. If this percentage is presumed to be representative of the number of dichotomous ML clauses in the whole Welsh-English corpus, then this would suggest that a dichotomous ML clause would be produced fewer than once per conversation on average, which is a negligible amount. Nevertheless, I believe these two clauses do show word-order convergence to English, as I will explain below.

The first clause has a monolingual head/modifier NP which is significant to the clause being identified as having a dichotomous ML; it was cited above as (14) and I repeat it below as (21).

(21)	roedd	**drws**	**nesa**	#	**pobl**	yn
	be.3S.IMP	door	next		people	PRT
	gwneud		**sloe**	***gin***		
	make.NONFIN		sloe	gin		

"the next-door people made sloe gin." [FUS6-AMR371]

As noted above, the NP *drws nesa pobl* 'next-door people' has a Welsh modifying element *drws nesa* preceding an English head noun *pobl*, resulting in MH word order, which indicates English is the source of morpheme order for this head/modifier NP. However, the VS subject/verb word order of *roedd* and *drws nesa pobl*, along with the outside late system morphemes of the morphology of *roedd* indicate that Welsh is providing some structure to the clause. The SMP indicates Welsh but the MOP indicates both English and Welsh. Since there is conflicting information for ML this clause has a dichotomous ML, on the basis that there is a discrepancy between the language providing word order in a head/modifier NP within the clause and the language providing subject/word order for the clause.

The second clause which is identified as having a dichotomous ML was given above as (20); I repeat it below as (22).

(22)	and	sydd	ddim	yn	***Cinderella***	**type**	#
	and	be.REL	NEG	PRT	Cinderella	type	
	peth						
	thing						

"And which isn't a Cinderella type thing." [FUS27-MAB612]

This is a bilingual clause (the words *and* and *type*[120] are from English), and the finite verb *sydd* is Welsh, so the SMP indicates Welsh as being the ML source. There is no overt subject in the clause, so the verb/subject word-order criterion cannot be tested. The word-order within the NP *Cinderella type peth* 'Cinderella type thing' is MH, since

[120] Note that *type* in this utterance is pronounced [taip] by Mabli, i.e. essentially the English pronunciation, rather than the more ambiguous form [teip].

the two modifiers *Cinderella* and *type* precede the head noun *peth* 'thing'. This word order is identified as being English according to the MOP. Since the SMP points to Welsh but the MOP points to English as supplying clause morphosyntax, I identify the clause as having a dichotomous ML.

Since there are only two dichotomous ML clauses identified in the dataset, this means that the speakers in these data almost always construct head/modifier NPs (and the clauses in which they occur) in line with the predictions of classic code-switching, and only in the dichotomous ML clauses discussed above does any speaker deviate from this norm.

I now proceed to the question of whether these dichotomous ML clauses actually show convergence. The anomalous word order found is presumably the result of some process, but is that process the mechanism of convergence?

In the NP *drws nesa pobl* in (21), what is found are Welsh morphemes in English-predominant MH word order. Since the material in the rest of the clausal structure in which this NP occurs has Welsh not English structure, it is not expected for this NP to have anything other than Welsh word order. This NP shows clear structural discrepancy in relation to the rest of the clause in which it occurs, arguably signs of the result of a contact-induced structural effect.

In the NP *Cinderella type peth* in (22), morphemes from both Welsh and English are found in, again, an English-predominant word-order, MH. The verb *sydd* is taken to indicate that Welsh is the source of some clause structure, so Welsh might be expected to also supply word order to the clause. Instead, English is the source of the word order of the head and modifiers in an NP in the clause. It is clear that both Welsh

and English are being accessed to source morphosyntactic structure to (22), and so again I identify that a contact-induced effect has occurred.

Recall the definition of a construction which has signs of word-order convergence, set out in chapter 2 section 2, which has L_A morphemes with a word order which is more common in L_B than in L_A. In the NP *drws nesa pobl* in (21), all the morphemes are Welsh, but the word order (MH) is more common in English than in Welsh—but it is available in Welsh, as a less frequent alternative.

Furthermore, recall the description of the signs of convergence as set out by Myers-Scotton (2002:164), where "all the surface morphemes [came] from one language" but have "some abstract structure from another language". *Drws nesa pobl* consists of "surface" morphemes which are all Welsh, but the word order is more like English word order than it is like Welsh order. This NP fits the definition of convergence I have provided, so I argue that word-order convergence from Welsh to English has occurred in the NP construction shown in (21). Here, Welsh morphemes are positioned in English-predominant word-order, because of the influence of English structure on the clause.

Other interpretations of the clause in (21) are possible, however, and are worth briefly discussing. First, Amranwen inserts a brief pause (less than 0.5 seconds) between the words *drws nesa* and *pobl*, which may suggest that she is aware that she is producing an unusual word order, but chooses not to correct herself. I would still interpret the clause as showing convergence, since a speaker can presumably produce convergence either consciously or unconsciously.

Second, alternatively, the word order in (21) could be due to interpolation of an adverbial *drws nesa*, with the pause flagging this, in which case the sentence would mean something like "there were, next-door, people making sloe gin". In this case the word order of *drws nesa*

pobl would not be anomalous and the clause would not have a dichotomous ML.[121] In standard grammar this kind of structure in Welsh requires soft mutation of the constituent that follows the interpolation, i.e. *roedd, drws nesa, **b**obl yn wneud sloe gin*, but such mutation is not produced by Amranwen, which could be taken as an argument against this clause showing interpolation of *drws nesa*. My experience as a native Welsh speaker, furthermore, is that interpolation of this kind (between the finite verb and the subject) is rare in contemporary spoken Welsh. Overall, I suggest that convergence is the most salient interpretation of the word order found in this clause.

The NP *Cinderella type peth* in (22), meanwhile, consists of morphemes from both Welsh and English occurring in a predominantly-English word order. This also fits in with my definition of convergence, and so I propose that this phrase also shows word-order convergence towards MH order, which is more common in English. It does not tally with Myers-Scotton's statement that constructions showing convergence have morphemes "all from one language", but I suggest that the structural dichotomy identified in the clause does nevertheless point to convergence as having taken place.

In this section I have discussed the dichotomous ML clauses found in the dataset analysed, and have identified that they probably show word-order convergence. However, I have also noted that dichotomous ML clauses are extremely rare in the data. According to my application of the MLF model to this dataset, there is effectively no word-order convergence in the output of these speakers.

In the next section I will discuss what these findings suggest about word-order convergence in Welsh in general.

---

[121] Actually (21) would then be divided into two clauses, neither of which would have a dichotomous ML.

c. What this analysis suggests about word-order convergence in Welsh

As concluded in the previous section, given that the results from the analysis presented in this chapter suggest that there is, with two exceptions, no word-order convergence in the speech of these six Welsh-English bilinguals, the question must be raised as to whether there is equally no word-order convergence in the speech of <u>all</u> Welsh-English bilinguals.

One point, which I will address in the next section, is whether or not the rarity of word-order convergence found in the data is due to limitations in using the MLF model to identify convergence.

First, however, to what extent is there convergence to English in Welsh word order in these data? I will argue that whilst, strictly speaking, these speakers do produce some word-order convergence (i.e. two clauses), the indication is that there is <u>not</u> convergence overall in the analysed data.

In chapter 2 section 3 I noted that many studies have shown that multiple occurrences of a similar type of convergence often result in a more widespread grammatical shift in the language of the speaker community analysed; e.g. Sandalo (1995), Schmitt (2000), Toribio (2004). The data I have analysed show that this is clearly not happening with Welsh, because only one clause containing convergence was identified. If there were frequent examples of the same type of

convergence seen in the NP *drws nesa pobl*, where Welsh morphemes have English word order, then it could be argued that Welsh is converging towards English MH word order in head/modifier NPs.

However, as I have already discussed, there is no evidence of a widespread HM > MH word-order shift in these data. In fact, the great majority of head/modifier NPs examined have the expected word-order.

In order to gauge how representative the speech of the 6 speakers (3 conversations) I analysed is, I looked more informally at the remainder of the Siarad corpus data for instances of dichotomous ML clauses. Due to time constraints it was not possible to undertake as detailed an analysis of the transcripts as was made for the three transcripts I analyse above. Instead, I read and/or listened to each transcript/recording and isolated every clause which I identified as having a dichotomous ML.[122] In table 4 below I list each conversation and outline what dichotomous ML clauses, if any, were found in it.[123]

Conversation	# of speakers	# of clauses with a dichotomous ML
Davies1	2	2
Davies2	2	0
Davies3	2	0
Davies4	2	0
Davies5	3	0
Davies6	2	0
Davies7	2	0
Davies9	2	0
Davies10	3	0
Davies11	3	0

[122] I reiterate that such a casual methodology is not watertight, and it may be that I have failed to identify some dichotomous ML clauses in this process, even though I believe I have noted all such clauses that occur.

[123] This list includes the three conversations I analyse above in this chapter (Davies6, Fusser6 and Fusser27), and therefore the two dichotomous ML clauses I have already identified as having been found in those conversations.

Davies12	2	0
Davies13	2	0
Davies14	2	0
Davies15	2	0
Davies16	2	0
Davies17	2	0
Deuchar1	2	0
Fusser3	2	0
Fusser4	2	0
Fusser5	3	0
Fusser6	2	1
Fusser7	2	1
Fusser8	3	0
Fusser9	2	0
Fusser10	2	0
Fusser11	2	0
Fusser12	3	0
Fusser13	3	1
Fusser14	2	0
Fusser15	2	0
Fusser16	2	0
Fusser17	2	0
Fusser18	2	0
Fusser19	2	0
Fusser21	2	0
Fusser22	2	0
Fusser23	2	0
Fusser25	2	0
Fusser26	2	0
Fusser27	2	1
Fusser28	2	0
Fusser29	2	0
Fusser30	2	0
Fusser31	2	0
Fusser32	3	0
Lloyd1	4	0
Robert1	2	0
Robert2	2	0
Robert3	2	1
Robert4	2	0
Robert5	2	0
Robert6	2	0
Robert7	3	0
Robert8	5	0
Robert9	2	0

Roberts1	2	0
Roberts2	2	0
Roberts3	2	2
Roberts4	2	3
Smith1	2	0
Stammers1	2	0
Stammers2	2	0
Stammers3	2	0
Stammers4	2	0
Stammers5	2	0
Stammers6	3	0
Stammers7	2	0
Stammers8	2	0
Stammers9	2	0
TOTAL	153	12

Table 4. Dichotomous ML clauses observed in entire Siarad corpus.

As can be seen, very few dichotomous ML clauses are found in the data. Only 12 were found (including the two already discussed in this chapter), spread across several recordings, although three conversations contained more than one dichotomous ML clause, Davies1, Roberts3 and Roberts4 (in Davies1, the two dichotomous ML clauses produced are by the same speaker; in Roberts3 each speaker produces one dichotomous ML clause each; in Roberts4 the same speaker produces all three dichotomous ML clauses).

The dichotomous ML clauses found in the whole corpus all, like the two already discussed, concern a head/modifier NP with MH order in a clause with a Welsh finite verb. Examples are given in (23), (24), and (25). In each instance the head/modifier NP in question is in bold.

(23)	fi	'n	credu		ni	'di		
	1S	PRT	believe.NONFIN		1PL	PRT.PAST		
	wneud		y	**mynd**	**mas**	**i**	**Bangor**	
	do.NONFIN		DET	go.NONFIN	out	to	Bangor	
	<u>thing</u>	sut	gymaint					
	thing	such	amount					

"I think we've done the 'out to Bangor' thing so much."

[DAV1-NON576]

(24)	timod	mae	gyn	I #	dipyn	bach	o	
	know.2S	be.3S.PRES	with	1S	little	little	of	
	gypyrddau #	yn	rhydd	yn	y [/]	yn	y	gegin
	cupboards	PRT	free	in	DET	in	DET	kitchen
	mewn	ffordd	'lly	xxx #	*set* o	droriau	wrth #	drws
	in	way	thus	xxx	set of	drawers	by	door
	i	fynd	i	*passage* #	a	cwpwrdd		
	to	go.NONFIN	to	passage	and	cupboard		
	double	lle	mae	'r	**<u>microwave</u>**	**pethau**		
	double	where	be.3S.PRES	DET	microwave	things		
	ar	ei	ben	o				
	on	POSS.3SM	head	3SM				

"You know, I have a few cupboards free in the kitchen, in a way, then, [...] a set of drawers by the door to go into a passage and a double cupboard where the microwave things are on top of it."

[FUS13-CRI1496]

(25) geith Kate cymryd lle fi ond
get.3S.NONPAST Kate take.NONFIN place 1S but
mae 'n well gennyn nhw ##
be.3S.PRES PRT better with.3PL 3PL
certain [?] **amser**de?
certain time TAG
"Kate can take my place, but they prefer a certain time, you know." [FUS7-BLO500]

In (23) the embedded clause *ni 'di wneud y mynd mas i Bangor thing sut gymaint* 'we've done the "going out to Bangor" thing so much' contains the collocation *mynd mas i Bangor thing*, where *mynd mas i Bangor* modifies the English morpheme *thing* in MH word order.[124] A deleted first person plural present auxiliary *ŷn* is assumed in this clause (as indicated by *ni*), which would point to Welsh as being the ML, but the MH word order of *mynd mas i Bangor thing* does not match, since MH is more usual in English, and so this clause is assigned a dichotomous ML. Similarly, *microwave pethau* 'microwave things' in (24) where English *microwave* modifies Welsh *pethau* in English-predominant MH word order, but the finite verb *mae* would point to Welsh, so that clause has a dichotomous ML too. Lastly the construction in (25) has *certain*[125] preceding the Welsh word *amser* in MH word order whereas the finite verb *mae* indicates Welsh, resulting in the clause ML being labelled as dichotomous. Other examples of similar MH word order in

[124] That this is a complex NP of this kind is reinforced when the same speaker contrasts *mynd mas i Bangor thing* with *y mynd mas adre thing* 'the "going out at home" thing' in the next utterance (NON578)—a clause which is also identified as having a dichotomous ML for similar reasons.

[125] *Certain* is attested in both the Welsh and English dictionaries (albeit with different orthography), but its positioning in relation to *amser* is still unusual for Welsh.

clauses in the corpus include *set dyddiad* 'set date' and *average oed* 'average age' (Roberts4).

Interestingly, two of the examples above involve the word *thing* or *peth* 'thing' in a MH word order NP, and this is true of other dichotomous ML clauses found in the corpus, e.g. two in Roberts3 (*y* ***raw peth*** *'ma* 'this raw thing' [MED389], *y* ***cwrs thing*** *'ma* 'this course thing' [LER804]) and another in Davies1 (see footnote 116 above), as well as ***Cinderella type thing*** in Fusser27 (see example 22 above). In fact, while there are probably too few examples in the corpus overall to start looking for patterns within dichotomous ML clauses, it is perhaps noteworthy that seven dichotomous ML clauses out of 12 involve this word in either Welsh or English. An explanation for its occurrence in this way might be that some Welsh speakers have borrowed (either in its English form or as a calque) this specific word along with its word-order properties, so that it is more likely to be positioned in MH word order than other words. It may be being used as a filler word (cf. *thingummy* in English) with freer word-order properties. In either case, this can still be interpreted as convergence, where English word order influences the positioning of words within a clause which would otherwise have Welsh ML.

There is no space here (and there are insufficient data) to undertake a detailed sociolinguistic analysis of the dichotomous ML clauses found in the rest of the corpus, but while they happen to have all been produced by women, they are produced by speakers of varying ages (e.g. Non in Davies1 is 18 while Medeni in Roberts3 is 56), so it is still unclear whether or not dichotomous ML clauses happen more frequently in the speech of younger or older speakers.

Although the number of speakers in each conversation, and the number of clauses produced by each speaker within a conversation,

varies, the above table gives a general indication of the frequency of dichotomous ML clauses in these data. They are extremely rare, as can be seen. 12 clauses across 69 recordings is fewer than one in five dichotomous ML clause per conversation. This indicates that the two dichotomous ML clauses found in the three conversations I analysed in detail actually represents a higher than average frequency when compared to the Siarad corpus as a whole (though this is, of course, due to chance). If it is assumed that the corpus is representative of the Welsh-English bilingual population as a whole, then it is a strong and conclusive indicator that word-order convergence in head/modifier NPs is practically non-existent in Welsh today, and any such convergence that does occur is sporadic and does not pattern to any obvious extralinguistic variable, although the reoccurrence of *peth/thing* in MH word order is striking and may benefit from further research.

Therefore, the indication from the data in this corpus is that Welsh is not converging towards English in the word order of head/modifier NPs, even though it is the case that there is some convergence of this kind visible, albeit rare and sporadic.[126] A diachronic study would indicate whether speakers in the future are shown to use MH order more interchangeably with HM than the speakers analysed here, but that is a matter for future research. Indeed, the inference made from these data is that such a change is not predicted to occur very soon.

---

[126] It is, however, not entirely absent from people's speech. I have overheard Welsh speakers produce sentences such as: ffoniais i'r service pobl 'I phoned the service people', where pobl service would be the word order if Welsh were the ML, and *mae o efo'r wrong goriad* 'he has the wrong key', where *goriad wrong* would be the word order if Welsh were the ML. I have heard similar examples in the speech of people whom I would classify as very proficient bilinguals. Nevertheless, I observe that such constructions are notable by their idiosyncracy and rarity, and the paucity of such clauses in the Siarad corpus supports this observation.

5. Reviewing the effectiveness of the MLF model in identifying word-order convergence

In this chapter so far I have presented the results of applying the MLF model to a set of data from Welsh-English bilinguals; this was based on my adaptation of Myers-Scotton's (2002) version of the model. My innovation was to introduce the concept of a dichotomous ML to account for clauses where the MLF's principles do not point to the same language as ML.

The aims of the analysis were (1) to find out what language is the predominant ML for Welsh-English bilinguals' clauses, and (2) to ascertain how useful the MLF model is in identifying word-order convergence in Welsh-English data. I have already discussed the first aim above; I now respond to the second aim.

The extent of the success of applying the model to the data can be measured by seeing to how many clauses the model could be successfully applied. I call this the model's 'scope', by which I mean its effectiveness in terms of the number of clauses where it can identify an ML as Welsh, English or dichotomous.

An ML can be identified in all finite clauses in the dataset, indicating that the MLF model has a very broad scope when applied to finite clauses produced by Welsh-English bilinguals. 56.85% of the total clauses in the dataset have a finite verb. When the model is applied to clauses without a finite verb, however, its scope is much narrower.

13.74% of the clauses in the dataset are nonfinite, and the MLF model can identify an ML in only four (0.89%) such clauses. Meanwhile, verbless clauses constitute 29.40% of the dataset, and the MLF model can identify an ML in only 0.83% of such clauses. The combined number of finite, nonfinite and verbless clauses in the data in which an ML is identifiable is 1874 out of the 3275 clauses in the data (57.22%). The remainder of the clauses in the dataset do not have an identifiable ML because the MLF model could not be successfully applied to them. The scope of the MLF model with this dataset is reasonably wide—an ML is identifiable in 57.22% of all clauses analysed—but it clearly does not have full applicability.

The two principles of the MLF model, the SMP and the MOP, both use a clause's finite verb as a means of identifying the source of that clause's ML. The SMP looks at the outside late system morphemes in finite verb morphology in Welsh and English verbs, whilst the MOP looks at, amongst other things, the word order of the finite verb and the subject. Self-evidently, if there is no finite verb in a clause, the SMP cannot be tested on that clause at all, and the MOP can only be tested if the clause contains a head/modifier NP of a certain type (as discussed in this chapter). This limits the applicability of the model when tested on nonfinite or verbless clauses.

I do not identify this as an inherent flaw in the MLF model, however. Rather, I suggest that the MLF model is probably 'easier' to apply—that is, it has a wider scope, able to identify an ML in a higher proportion of clauses—in languages which have a greater variety or higher within-clause frequency of outside late system morphemes.[127]

---

[127] Of course, I presume that speakers of these kinds of languages will also produce e.g. fragments and nonfinite clauses in natural speech, like these Welsh-English bilinguals.

Since the only outside late system morpheme which can be used to satisfy the SMP in Welsh and English is finite verb morphology, clauses which lack a finite verb cannot satisfy the SMP. On the other hand, the MLF model might have a comparatively wider scope when tested on languages which contain a greater variety of types of outside late system morphemes, e.g. languages which are morphologically richer than Welsh or English, since, presumably, more clauses in data from speakers of such languages would contain outside late system morphemes.

A factor which I will return to in the next chapter is related to the issue of finite verb deletion, which I raised briefly in chapter 3. In Welsh, certain auxiliary constructions are often produced with a deleted auxiliary verb. In such clauses there is no overt finite verb on which to test the SMP. In analysis I chose to analyse such clauses as having an ellipted Welsh verb with Welsh outside late system morphemes, though without an analysable surface position in the clause structure. This meant that the SMP could be tested on auxiliary-deleted clauses (identifying Welsh as ML), but not the subject/verb criterion of the MOP. This was a decision based on the assumption that Welsh is probably supplying morphosyntax to such clauses. It may, however, be possible that such clauses actually show word-order convergence which the MLF model fails to identify. I discuss such constructions in more detail in chapter 6, in a separate analysis.

Further issues that arise from the application of the model include clause division. As noted in chapter 3, Myers-Scotton (2002:54) posits the Complementiser Phrase as the unit of analysis for the MLF model, and so I isolated all clauses in the data for the purposes of

---

Such clauses in those languages would still be less straightforward to test the MLF model on than finite clauses.

analysis. However, in order to divide the data into clauses I had to separate out embedded clauses from within larger complex clauses. I ignored, for example, the relationship between the MLs of individual clauses within a complex clause, following the parameters of the MLF model. The MLF model is mostly concerned with code-switching at the smallest clausal level: it does not, in its current formulation, make predictions about language-mixing at the larger level (although, perhaps ironically, this was supposedly a function of the model in its early conception [e.g. Myers-Scotton 1993]). As far as the concept of the dichotomous ML is concerned, it can only identify convergence at the simple clausal level, and if, for instance, a language were to show convergence of larger structures (e.g. of the position of embedded structures in complex clauses, such as a right-branching language converging to being a left-branching language), the MLF model could probably not identify these.

Given that this analysis process has highlighted some issues with the MLF model, then, how successful has my application of it been to this dataset? Has it been proven useful? I argue that it has. To demonstrate this, I now return to considering the primary aims of the analysis as noted at the beginning of this chapter.

Applying the MLF model to the dataset has shown that Welsh is the provider of structure (ML) in the great majority of these speakers' finite clauses, and, notably, that finite bilingual clauses with English ML simply do not occur in this dataset. Note also that the results show that this pattern is similar throughout the speech patterns of all six speakers analysed. With regards to the predictions of the MLF model as to the nature of the code-switching exhibited by these speakers, moreover, I have shown that they exhibit classic code-switching in all but two instances (the dichotomous ML clauses) in their bilingual clauses, and

that they adhere, the great majority of the time, to the prediction that one language is the source of morphosyntax in a clause in which there is code-switching: Welsh. This is in line with the findings made by Deuchar (2006) for her Welsh-English data.

The second aim of this analysis was to see if the MLF model could be used a tool to identify convergence. Two dichotomous ML clauses were identified, and I have demonstrated how it can be interpreted to show word-order convergence. But more notably, according to this method, I have shown that there is almost no word-order convergence in this Welsh-English dataset. A language-specific ML (i.e. either it is Welsh or it is English) is identifiable in the great majority of finite clauses, which means that those clauses do not contain word-order convergence. More specifically, my method allows a quantification of the word-order convergence present in the dataset: in terms of clause proportion, there is only word-order convergence in 0.06% of the output of the Welsh-English bilinguals I have analysed.

In the next section I will demonstrate how my adaptation of the MLF model can be successfully applied to data from a language pair other than Welsh-English, specifically Austrian German-English, where further examples of dichotomous ML clauses showing word-order convergence are found.

6. Looking for convergence in Austrian German-English data using the MLF model

So far in this chapter I have shown that applying the MLF model to Welsh-English data results in the identification of very few dichotomous ML clauses showing convergence. Whilst this may be an issue with the model as I have adapted it, it may also be due to a genuine lack of word-order convergence involving the relative subject and the finite verb or of the head and modifier in NPs in Welsh. If the same method were tested on a different dataset, involving different languages, and such an analysis showed more clauses to have a dichotomous ML, then this would show that the issue lies with the languages in question more than with the tenets of the model.

In this section I present a new analysis, applying the MLF model to data collected by Eppler (2004) from bilinguals in Austrian German and English, to discover the extent of dichotomous ML clauses in those data, and thereby further ascertain the effectiveness of the model as I have adapted it.

a. The Austrian German-English data

The data I use for this smaller analysis is a subset of the data collected by Eppler in 1993 (Eppler 2004).[128] She recorded Austrian Jewish immigrant refugees then living in north-west London, who had moved from Austria to England in the 1930s, prior to the Second World War (2004:128). All the participants were in their late 60s or early 70s when

[128] Eppler's transcribed data were sourced from http://talkbank.org, the same database that holds the Siarad Welsh-English corpus. The transcriptions were made, like my data, using the CHAT system and the LIDES format.

they were recorded. Their L1 was ("standard") German, phonetically influenced by Viennese German. Their L2 was English, acquired in their late teens to early twenties. Eppler (2004:129) notes that all the participants were highly proficient in both their languages, and used a high-prestige in-group mixed code (which Eppler refers to as "Emigranto"), a characteristic of which is inter- and intrasentential code-switching.

Eppler made many of her recordings at social meetings between members of this group, which included two or more participants involved in various activities. Eppler herself, as the researcher, was present, and often participated in the discourse.

I expected that the prolonged period of time that the speakers in Eppler's data had spent living in London (approximately 60 years), and having their community language changed from being Austrian German to being English, would result in a change occurring in their German output because of contact with English. That is, I hypothesised that the "Emigranto" they used consisted of frequent convergence from Austrian German to English.

b. Applying the model

Using a similar methodology to that of the Welsh-English analysis, I tested the MLF model on the Austrian German-English data by subdividing one transcript (Jen3a) into simple clauses and using the SMP and the MOP to identify the ML of every given clause. The options for ML were German, English, dichotomous, or unidentifiable.

Eppler (2004:128–9) states that the participants in her study spoke standard German with some phonological aspects of Viennese

(i.e. Austrian) German. For the purpose of analysis, I assume these speakers' dialect is essentially Austrian Standard German (ASG), which is very similar to the Standard German of Germany (GSG) as far as syntax is concerned, though ASG possesses some phonological, semantic and lexical properties which distinguish it from GSG (Martin 1997).

The main differences between the grammars of English and Austrian German which I will be using as a means of identifying the which language is the ML source are as follows:

- Outside late system morphemes: English finite verb morphology, as noted in chapter 3, includes outside late system morphemes. I concur with Luescher (2008:26), who identifies Austrian German finite verb morphology as being outside late system morphemes, along with determiners, pronouns, adjectives and certain nouns, because they are (usually) inflected for person, number etc. as finite verbs are, and have grammatical relations external to their constituent. These will be the criteria used to identify the source of ML as German or English with regards to the SMP.
- Morpheme order: Both English and German head/modifier constructions usually have MH order, so are not useful as a differentiator. However, clause word order in German and English is notably different. Whereas English is SVO, German is a V2 language, meaning that while SVO is a frequent word order, the verb must occur in second position, which often results in VSO word order. German also has rules which mean that past participles are positioned at the end of main clauses, and finite verbs are positioned at the end of subordinate clauses (e.g. Eppler 2004:132). These word order differences relating to the

position of the verb will be the criteria used for identifying the language which is the source of ML according to the MOP.

To demonstrate how I tested the MLF model on these data, examples (26) and (27) below show two bilingual clauses from the analysed data. English words are underlined in the examples that follow.

(26) und heuer fahren wir nach Harringate #
and this year go.1PL.PRES 1PL to Harringate
for a long weekend
for a long weekend[129]
"And this year we're going to Harringate for a long weekend."
[DOR-136]

(27) vorgestern you weren't there
day before yesterday you weren't there
"The day before yesterday, you weren't there."
[TRU-230]

In (26) the finite verb *fahren* is German, thus identifying German as the source of outside late system morphemes in the clause. The word order of the verb *fahren* and the subject *wir* is VS, because in German the verb is the second element of the clause; here, *heuer* is the first element. English does not have VS order in this context, and so German is clearly providing word order. Since both the SMP and MOP identify German as the source of morphosyntax, German is identified as the source of the ML of (26). In (27), the finite verb *weren't* is English. Its position after the subject *you* is SV. If German were providing word

[129] The glossing of Eppler's data in this example, and the examples that follow, is mine.

order, then the presence of *vorgestern* 'the day before yesterday' would make the finite verb the next element. It is not, and so English is identified as supplying word order for the clause. Both SMP and MOP point to English, and so English is the ML of (27).

The data I have analysed is the transcript titled Jen3a, of data recorded and transcribed by Eppler using the CHAT system. Jen3a, which is 52 minutes 9 seconds in length, is approximately one half of the total session ("Jen3") originally recorded by Eppler. The participants are four women, Dorit (74 years old, Eppler's main participant), Lilly (75), Trude (74) and Melly (69), as well as Eppler herself, and an unnamed sixth woman (whose speech has not been transcribed). They are playing cards together during the recording but maintain a conversation throughout. The transcription header, presumably written by Eppler, notes that approximately 55% of the words in the discourse are German and that approximately 45% are English. This means that both languages are used in approximately the same proportion by the participants in general (though not necessarily by the individuals concerned), as indicated by Eppler's comment that these speakers use frequent code-switching (2004:129).

c. Results

The data consist of 2142 clauses in total, produced by the four selected speakers (I exclude Eppler herself and the sixth card player from the analysis). Out of those clauses, 1671 (78.01%) had an identifiable ML. The results for the ML distribution across all clauses is given in figure 9 below.

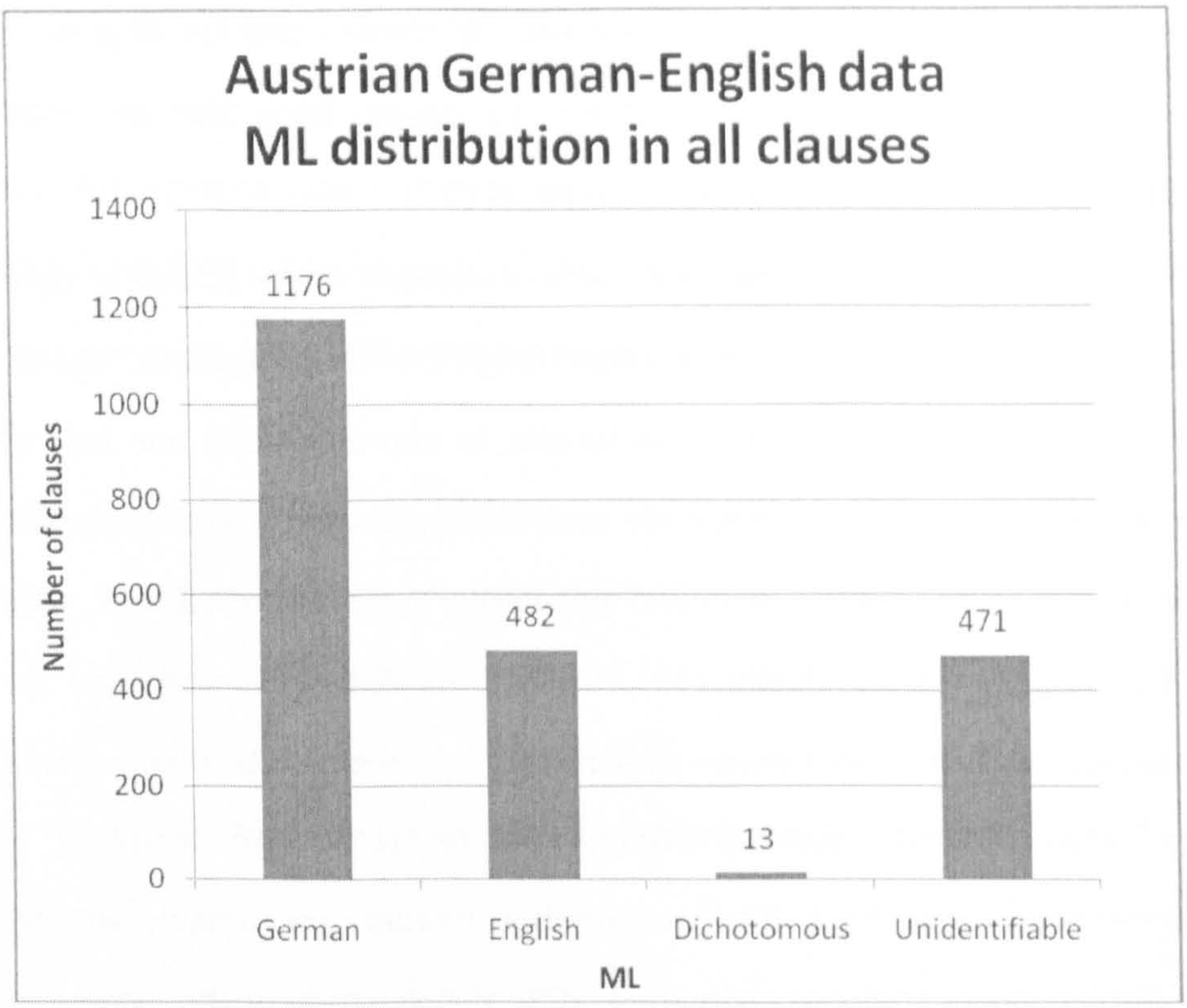


Figure 9. The Matrix Language distribution across all clauses (monolingual and bilingual) in the Austrian German-English dataset.

First looking at monolingual and bilingual clauses together, 1176 clauses (70.38% of those with an identifiable ML) were found to have German as the source of the ML. 482 clauses (28.85%) had an English ML. Finally, 13 clauses were identified as having a dichotomous ML (0.78%). German supplies the morphosyntax most frequently in the clauses of these four speakers according to these data.

I now focus on the bilingual clauses in the dataset. The ML distribution in bilingual clauses are illustrated in figure 10 below.

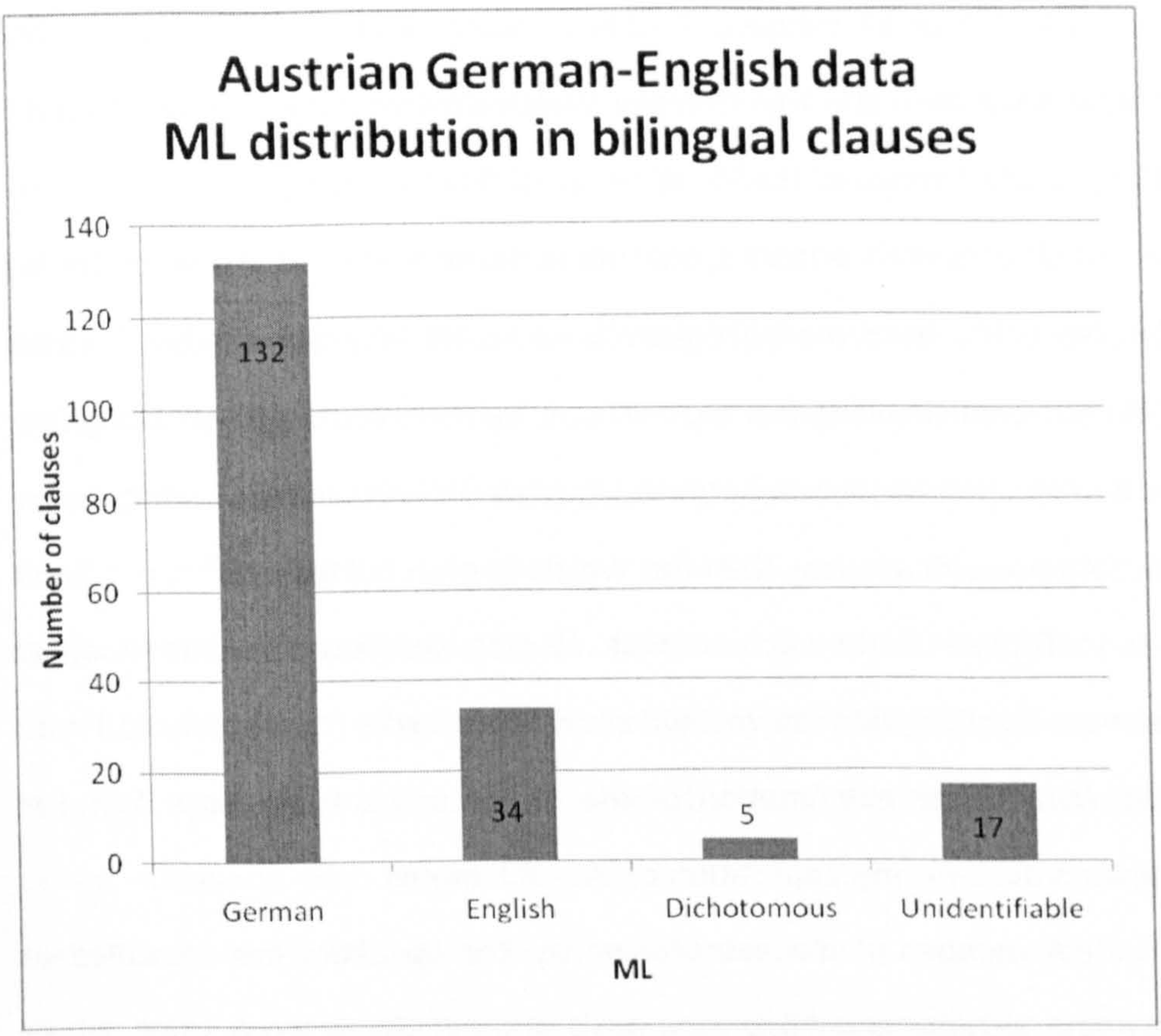


Figure 10. The Matrix Language distribution across all bilingual clauses in the Austrian German-English dataset.

There were 188 bilingual clauses in the dataset (8.78% of all clauses). Of these, 171 (90.96%) had an identifiable ML. Of those clauses where an ML was identified, 132 (77.19%) had a German ML, 34 (19.88%) had an English ML and 5 (2.92%) had a dichotomous ML. German again is frequently the source of the ML in the bilingual clauses of these speakers. Of interest, however, is that almost one fifth of the bilingual clauses analysed had English as the ML; see example (27) above for a bilingual English ML clause from the Austrian German-English data. Recall that, in my Welsh-English data, English was never found to be the ML in bilingual clauses.

Note also that the proportion of dichotomous ML to clauses which do not have a dichotomous ML found in these Austrian German-English

data (0.78% of all clauses; 2.92% of bilingual clauses) is considerably higher than the ratio found in the Welsh-English data I analysed above (0.06% of all clauses; 0.40% of bilingual finite clauses).

The overall observations made from these results point to (a) English being used more frequently as an ML source by these Austrian German-English bilinguals than it was by the Welsh-English bilinguals, and (b) the Austrian-German English bilinguals producing more dichotomous ML clauses than the Welsh-English bilinguals.[130]

To see if the dichotomous ML clauses found in the Austrian German-English data do in fact show word-order convergence, I now turn to a closer examination of the dichotomous ML clauses found in Eppler's data via my application of the MLF model.

A number of clauses uttered by the speakers are identified as having a dichotomous ML on the basis of main clause word order, where some elements in the clause indicate German as providing word order and other elements indicate English as providing word order. A frequently-found form of this phenomenon is, as noted above, where the past participle is not positioned at the end of the clause as in standard German. An examples are given in (28) below.

(28)	im	achtundachziger	jahr	#	hab(e)	ich
	in	eighty-eighth	year		have.1S.PRES	1S
	gemacht		eine	fünfzigjährige		<u>reunion</u>
	make.PASTPART		DET	fifty year		reunion

"in the eighty-eighth year I made a fifty year reunion" [DOR-116]

---

[130] These two facts may be connected. Perhaps if bilinguals in a certain speech community source the ML of their bilingual clauses from either of their languages, then they will also be more likely to produce dichotomous ML clauses. To verify this, it would be necessary to analyse further bilingual corpora to see if a pattern emerges; this will be a matter for future research.

The SMP clearly identifies German as supplying finite verb inflection in this clause, e.g. *habe*. The MOP too identifies German as being the source of some of the word order in the clause, e.g. the VS order of *habe ich* following the PP *im achtundachziger jahr*, typical of German V2 order. Some clause word order, however, is not typically German. The object of *habe gemacht* is *eine fünfzigjährige reunion*, which contains an English insertional CS. In the grammar of ASG, *gemacht* here should be positioned at the end of the clause. However, *gemacht* in (28) appears after the subject *ich* and before the object *eine fünfzigjährige reunion*. This is a more English-like order, where the nonfinite verb complement of an auxiliary verb comes between the auxiliary and the object constituent. The MOP identifies more than one language as supplying clause word order in (24): English supplies the word order of *gemacht* and German supplies the word order of *habe ich*. Since there is conflicting evidence concerning which language the MOP identifies as being ML, the clause has a dichotomous ML.

Other main clauses in the data which similarly lack a past participle appearing at the end of the clause are (29), (30) and (31) below.

(29) dann haben sie müssen warten
then have.3PL.PRES 3PL must.3PL.PRES wait.NONFIN
auf ein(en)[131] bus xxx
for DET.SN bus xxx
"Then you had to wait for a bus [...]" [DOR-263]

(30) und voriges jahr sind wir g(e)fahren
and previous year be.1PL.PRES 1PL go.PASTPART
nach Cambridge
to Cambridge
"And the year before, we went to Cambridge" [DOR-795]

(31) hab(e) ich gelesen in der
have.1S.PRES 1S read.NONFIN in DET.SM
zeitung # von einer lorry
magazine of DET.SM lorry
"I read in the magazine about a lorry." [MEL-1297]

There are a number of unusual aspects to the clause in (29). First, the infinitive verbs *warten* and *müssen* appear directly after the subject, instead of at the end of the sentence (because of *haben*) as would be expected of ASG. Secondly, those two morphemes are not in the expected order: when positioned at the end of the clause, they should be ordered *warten müssen*. The expected German form of this clause

[131] These brackets are in the original transcription, and represent an assumption on the part of the transcriber (i.e. Eppler) as to the usual determiner ending, which was not pronounced by the speaker. It would be possible to interpret the production *ein bus* rather than *einen bus* as representing convergence towards English (zero) case marking, as a bare form.

would be *dann haben sie auf ein' bus warten müssen*.[132] (29) and (30) similarly have German finite verbs but show an English-like word order of the nonfinite verb. In (30), the verb *gefahren* is found immediately following the subject *wir*, instead of after the PP *nach Cambridge* as it would be expected in standard German, to give *sind wir nach Cambridge gefahren*. In (31) as well, *gelesen* is found directly after the subject *ich* but preceding the PP *in der zeitung von einer lorry* rather than following it as would be expected in German (i.e. *habe ich in der zeitung von einer lorry gelesen*).

In all three examples the word order positioning of the nonfinite verb is more English-like than German-like, because that verb is found immediately after the subject, rather than at the end of the clause. Compare some equivalent English forms of these sentences: 'then you have waited for a bus'; 'the year before, we went to Cambridge'; 'because I have read in a magazine'. Here the finite auxiliary and the subject are in SV order, as expected for English; in German the order would be VS. However, the nonfinite verbs *waited*, *went* and *read* here immediately follow these morphemes, just like *müssen warten*, *gefahren* and *gelesen* in the German-English sentences above. Nevertheless, in the three clauses there is still some word order that is typical of German, such as the VS order of *haben sie*, *sind wir* and *habe ich* following the first element of the clause (*dann* and *und voriges jahr* respectively in [29] and [30] and the whole preceding clause in [31][133]). So, when the MOP is applied to each clause, it cannot identify one language uniquely as supplying morpheme order, since English and

---

[132] I have consulted a native German speaker who confirms that this would be the expected form.

[133] That clause is "sie ist ueberfahren worden".

German both appear to provide structure to this clause. These three clauses are assigned a dichotomous ML because of this.

Further examples of this phenomenon occurring in main clauses are shown in (31) and (32).

(31)	einmal	hab(e)		ich	gekriegt	eine	
	once	have.1S.PRES		1S	get.3S.PAST	DET.SF	
	singleton[134]...						
	singleton						
	"Once, I got a singleton..."				[DOR-1019]		

(32)	sie	sind	gefahren	mi(t)	(eine)m	taxi	
	3PL	be.3PL.PRES	go.NONFIN	with	DET.SN	taxi	
	"And they go by taxi."				[DOR-800]		

In (31) the past participle *gekriegt* is found following the subject *ich* but preceding the indirect object *eine singleton*, whereas in standard German it would be expected to be positioned at the end of the clause. This word order appears to have been influenced by English word order: where German has [Adv Aux SO Participle] word order, English has [Adv S Aux VO]. But the word order of *habe ich* is German, so again this clause has morpheme order provided by German and English (or at least influenced by the latter), and thus the MOP cannot identify one language as providing structure, so the clause is identified as having a dichotomous ML.

In (32) *gefahren* is not positioned after the PP *in einem taxi* as expected, but sits in the position following *sind*, which makes the

---

[134] The English word 'singleton' is a gaming term in Bridge and other card games, referring to the only card held by a player in a particular suit.

structure more English-like (cf. 'and they have gone'), so the MOP again indicates English whilst the SMP indicates German (cf. *sind*). Both (31) and (32) are assigned a dichotomous ML.

Example (33) below is of two co-ordinating clauses, one of which contains a subordinate clause. The first clause has a dichotomous ML.

(33)	und	wir	sind	gesessen	in	King's	
	and	1PL	be.1PL.PRES	sit.NONFIN	in	King's	
	Cross	und	haben	gewartet	bis	sie	
	Cross	and	have.1PL.PRES	wait.NONFIN	for	3PL	
	aufgemacht						
	open.NONFIN						

"and we sat in King's Cross, and waited until they opened."

[DOR-802/803]

Here the expected German word order for the first clause *und wir sind gesessen in King's Cross* would be for *gesessen* to be at the end of the clause after the PP in a [S Aux Complement Participle] structure, i.e. *und wir sind in King's Cross gesessen*. What the speaker has produced, however, is clause word order which is much more similar to English [S Aux V Complement], in that *gesessen* is positioned between the finite auxiliary and the adverbial PP. When applying the MLF model to this first clause, we see that the SMP points to German as the source of outside late system morphemes (i.e. the inflection of *sind*), but the word order is more English-like than German-like, thus the MOP indicates English as being the source of morpheme order. Thus this clause in (33) is labelled as having a dichotomous ML.

The above examples all show the English-like word order position of a participle in main clauses, and this is the phenomenon found in the

majority of the dichotomous ML clauses identified in these data (12 out of 13). The other dichotomous ML clause found shows a similar phenomenon, but this time in a subordinate clause where the finite verb, which in German is expected to be positioned clause-finally, is instead found following the subject. This is shown in (34); the dichotomous ML clause is the second clause beginning with *weil*.

(34) die [/] die muss doch
DET.SF DET.SF must.3Sf.PRES nevertheless
Wien kennenlernen, weil die
Vienna know.NONFIN because DET.3PL
eltern sind beide aus Wien
parents be.3PL.PRES both from Vienna
"Nevertheless, she must get to know Vienna, because both parents are from Vienna." [LIL-68/69/70]

In the subordinate clause in (34), the finite verb *sind* indicates that German is the source of outside late system morphemes. However, the position of this verb is not clause-final as would be expected of ASG (due to the presence of *weil* at the beginning of the subordinate clause), which would be *weil die eltern beide aus Wien sind*. Instead word order seen is the English-like order where the verb follows the subject. The MOP points to English while the SMP points to German. The clause is assigned a dichotomous ML because of this.

All the dichotomous ML clauses I have identified have some part of the word order which shows similarity to English. In the next section I will discuss to what extent this suggests convergence of German to English word order in the output of these speakers.

d. The extent of convergence in the Austrian German-English data

The dichotomous ML clauses I discussed above mostly show similarly unusual word order, where a participle or nonfinite verb which would be expected to occur clause-finally in ASG is instead positioned directly after the finite verb, i.e. SAuxVO is found rather than the more standard SAuxOV. This word order is more English-like than German-like, since in English the participle would be expected to follow the finite verb (SAuxVO). I have suggested that English word order has influenced the structure of all these clauses. I argue that this is convergence towards English-predominant SAuxVO word order.

Main clauses of this type in ASG would have a nonfinite verb clause-finally, following the object or complement, whereas in English that verb would follow the finite verb in SAuxVO order. VO order is nevertheless available in ASG, e.g. in main clauses where the basic word order is SVO; cf. the fabricated example in (35) below.

(35)	ich	mag	das	Buch
	S	V	O	
	I	like	DET	book

"I like the book."

An interpretation of the use of SAuxVO not SAuxOV in dichotomous ML clauses in these data is that the speakers have extended the frequency of German VO word order, via the influence of English, to auxiliary constructions in main clauses, resulting in VO word order in Austrian German main clauses. I identify this as being word-order convergence to

English, where the prevalence of English VO word order has influenced the increase of its use in the German of these bilinguals.

One subordinate clause with a dichotomous ML is found in the data, shown in (31) above. Here the finite verb is found directly following the subject, rather than clause-finally as would be expected of ASG. This clause has VO word order, which is typically English, since German subordinate clauses are expected to have OV order. Again, I suggest that the speaker has selected the English VO word order for German subordinate clauses, on the basis that this word order is available elsewhere in German (e.g. in main clauses, as shown in [35] above). I identify this too as word-order convergence.

All the 13 clauses in Eppler's data which I have identified as having a dichotomous ML show signs of word-order convergence from German to English due to the prominence of English SVO. It occurs in the speech of three out of the four analysed participants: Dorit (9 clauses), Melly (2 clauses) and Lilly (2 clauses). I infer from this that the clause-final positioning of the verb (nonfinite in main clauses; finite in the subordinate clause) has ceased to be obligatory in the speech of these bilinguals, due to the structural influence of English. Whilst such convergence is not at all common in the data—dichotomous ML clauses still represent less than one percent of the total clauses analysed—it is found, and probably too frequent and regular to be classed as a speech error.

I argue that this is an example of word-order convergence in the speech of these bilinguals. These speakers have been living in England for around 60 years when the data were recorded, and have apparently not spent long periods of time in Austria since emigrating. Their German has been affected by bilingualism in English, their L2 which they acquired after emigrating from Vienna. They would not have been

exposed to much English before they were teenagers (apart from perhaps at school in Vienna). The speakers keep regular company with other Austrian German-English bilinguals of similar backgrounds (and are "a close-knit network", according to Eppler [2004:129]), but regular contact with other German speakers from outside their group is presumably minimal, which implies that the German that these speakers come into contact with on a regular basis is similarly-affected by English contact, which may have a 'reinforcing' effect on their word order.

Though I argue that the positioning of a participle before the object (or PP etc.) found in some clauses in Eppler's data is due to influence from English clause word-order, it might be the case nevertheless that this word order was already present in colloquial varieties of the Austrian German of **monolinguals** in the Vienna region where Eppler's speakers grew up and acquired their **L1**. That is, the nonstandard word order might be a monolingual trait rather than evidence of any convergence deriving from the speakers' bilinguality. I consider this briefly here.

Some recent studies of Austrian German indicate that nonstandard word order is present in Austrian German to an extent, indeed more so than in the German dialects of Germany (Sapp 2006). Sapp—whose study examines several German varieties and not only Austrian German—refers to a certain German subordinate clause word order shown in (36) where the finite verb precedes the nonfinite verb, which is distinct from the standard German word order, where the finite verb follows the nonfinite verb.

(36) ...dass Klaus heute das Buch **will** **lesen**

that Klaus today DET book wants read

"...that Klaus wants to read the book today."

(taken from Sapp 2006:125)

[cf. Standard German: ...dass Klaus heute das Buch **lesen will**

Sapp (2006:125) states that such word order is "disallowed by prescriptive grammars" of contemporary German and is judged to be "completely ungrammatical" by native German speakers. Whilst this word order is different to the word order I identify above,[135] it is of relevance that several studies indicate such word order to be more frequent in the speech of Austrian dialectal German than in other dialects of German: Maurer (1926:60–63) finds that the above word order is found in the Tyrolean dialect (western Austria), while Patocka (1997:275 and 290) shows that the German spoken in Styria (south-eastern Austria) allows this word order. Sapp (2006:127) notes that the regions of eastern Austria form a "traditional zone" where this word order is used where the nonfinite verb is perfective but not where it is an infinitive. Austrian German, then, is perhaps syntactically different to other varieties of German to some extent. Nevertheless, recent research by Sapp (2006:154), using grammaticality judgements, indicates that this word order is currently declining in use in Austrian dialects.

Other nonstandard word orders have been found in the dialectal German of monolinguals by e.g. Auer (2007)[136], and while colloquial

---

[135] Most of the dichotomous ML clauses found in the analysed German-English conversation are main not subordinate clauses, for example, and the nonstandard word order found therein is the positioning of the nonfinite verb before the object (in this regard the order shown in (36), where *das Buch* precedes *lesen,* is standard)

[136] Auer's (2007) study of Kanaksprak—a variety of German spoken by young people who are descendants of Turkish immigrants (2007:325)—shows occurrences of a similar SAuxVO word order in a nonstandard context (2007:341). Auer's speakers are apparently

German deviates somewhat from standard German in all parts of the grammar, including syntax (cf. Barbour and Stevenson 1990:147), generally the differences between the syntax of dialectal and standard German are considered to be slight (cf. the discussions in the volume edited by Russ [1990], which highlight little in the way of notable word-order differences between the various German dialects described). I infer from Sapp's (2006) conclusions, that the word order deviation from the German standard that was once found in Austrian dialects has declined in frequency over the 20th century, that nonstandard word order is not very common in dialectal Austrian German.

What, then, of the extraposition of an argument or adjunct (where that element appears to the right of the verb cluster)? A hypothetical example of this (not necessarily Austrian German but rather colloquial German generally) is given in (37) below.

(37) ...dass Klaus heute lesen will **das Buch**
that Klaus today read wants DET book
"that Klaus wants to read the book today."

(taken from Sapp 2006:125)

In (37) the NP *das Buch* appears to the right of the verb cluster *lesen will* in a VO word order, contrary to the standard German word order *das Klaus heute* ***das Buch*** *lesen will* (OV word order). Sapp (2006:125) comments that this word order is possible "in colloquial forms" of standard German, albeit in marked contexts, but he further notes that it is not an usual word order and is judged acceptable only under certain discourse conditions (cf. Duden 1995:790-791). Note that Barbour and

---

monolingual but their German is highly stylised and may therefore not be representative of colloquial German in general.

Stevenson (1990:196) refer to word order where a nonfinite verb in a construction containing a modal verb fails to be moved to the end of the clause as "deviant" and a feature of *Gastarbeiterdeutsch* 'Guest Worker German', the speech of immigrants to Germany. It can be presumed, then, that this kind of word order is rare and quite unusual in monolingual speech, although further research is necessary to see the extent of this specific phenomenon in colloquial German (Austrian or otherwise).

I have been unable to find any literature which specifically discusses the situation of Viennese German syntax in the years immediately preceding the Second World War beyond Eppler's (1999, 2004) brief descriptions of it. She notes that the only Viennese dialect trait in the German of the speakers she recorded is the phonology (Eppler 2004:128), and, aside from some Yiddish influence of the phonological and lexical kind (2004:129), the variety they speak (and presumably therefore the variety they acquired as children living in Vienna) is "the Standard German spoken by the highly educated Jewish communities in the urban centres of the Austro-Hungarian monarchy" (Eppler 1999:292). Eppler therefore appears confident that the syntax of the German these speakers acquired whilst still living in Austria was the standard variety, in which one presumes the object extraposition discussed above did not occur.[137]

It is quite likely, therefore, that such word order was minimally present (if at all) in the German syntax acquired by Eppler's participants, and any instances of SauxVO word order found in Eppler's data—as in the clauses I identified to have a dichotomous ML—are

---

[137] That is, Eppler would presumably say that the standard Austrian German acquired by these speakers would not demonstrate extraposition of this type, even though it is possible that it may occur sporadically in colloquial Austrian German.

innovations based on English, and possible instances of convergence. This point is arguably reinforced by the fact that Eppler (e.g. 2004) has identified other parts of these speakers' speech as having been influenced by English, and therefore any other anomalous features—like the one discussed here—may similarly be due to the influence of English.

In fact, other studies of German word order—particularly the use of VO at the expense of OV—produced by L1 speakers who emigrated from Germany to an English-speaking country seem to support the notion that attrition of some kind (see e.g. Pavlenko 2004) is the main influence behind such nonstandard word order. Schmid (2002:168), comparing the language of German Jews who emigrated to Australia with the German of monolinguals, finds a trend towards a higher frequency of main clause SVO word order than in monolingual German, and the data also shows some adherence to English rather than German rules for adverbial placement: the clause shown in (38) has the time adverbial *einige Tage* following the place adverbial *hier*, which Hutz (2004:202) notes is expected in English but not German (in which adverbials of place follow adverbials of time, i.e. the opposite order).

(38) Sie waren hier **einige** **Tage**
they were here several days
"they were here for several days."

Schmid ascribes this structural overgeneralization by her speakers—perhaps to be equated with convergence—to English exposure, where SVO is used in preference to SOV in German by bilinguals in contexts where SVO is not used by monolingual German speakers. Other German immigrants to Anglophone countries have been demonstrated to show

English influence in their L1 (e.g. Waas's (1996:165) study of German immigrants in Australia).

Another factor that supports the notion that convergence, more than pre-existence in the monolingual variety, is salient in the increase in frequency of VO rather than OV is that Eppler's speakers had limited contact with any German speakers other than those in their own 'Emigranto' group. Such a situation might be conducive to a "vicious circle" of attrition (Schmid 2002:25) whereby speakers come to prefer the nonstandard structures they hear from other immigrants to any standard forms they used before emigrating. This differs from a scenario where the same speakers that demonstrate contact-induced convergence have contact with speakers of their L1 who demonstrate no contact-induced changes, a scenario which is argued to slow down the convergence process (Hutz 2004:203).

An assumption that can be made here is that the German these speakers learnt before they emigrated to England did not display anomalous word order in subordinate clauses. While word order of the kind found in dichotomous ML clauses in Eppler's data is arguably found in the German of monolingual Germans, it is rare and very marked (Sapp 2006). It follows, then, that such a word order must have appeared in Eppler's speakers' speech **after** they emigrated, a period which is surely characterisable as a period of intense English influence and sociolinguistic dominance. An explanation for the innovation of this word order as being due to the influence of English appears to be a valid one, in the absence of more details on pre-existing SauxVO word order in German in the speech of speakers in 1930s Austria.

The situation of these four Austrian German-English bilinguals analysed is very different from that of the Welsh-English bilinguals studied in this thesis. The former have had some contact with Austrian

German speakers since moving to England, but primarily only within their, presumably narrow, Austrian German refugee group. They live outside the country in which they grew up, in a community where German is not an official language. The majority language used in their present community (London) is English. The Welsh-English bilinguals I analysed, meanwhile, live in the same country as they grew up and have spent most of their lives in. It is a country where both English and Welsh are languages with official status, and where the minority language, Welsh, is spoken by a considerable proportion of the population and is also, in some cases, the community language. They are also all bilinguals who had learnt both their languages by an early age, before adolescence.

It seems that the Austrian German-English bilinguals have more intense exposure to English and, crucially, less exposure to their native language than the Welsh-English bilinguals do. I suggest that this is why there is more word-order convergence in the speech of the former, but a more in-depth study would be required in order to ascertain more precise sociolinguistic differences between the two groups, which would help understand why English would have more apparent structural effect on the output of the Austrian German-English bilinguals than of the Welsh-English bilinguals.

7. Returning to the evaluation of the MLF model

Earlier in this chapter I evaluated the scope of the MLF model when applying it to Welsh-English data, and showed how, whilst I identified one dichotomous ML clause as showing convergence, the data generally displayed adherence to the parameters of classic code-switching and of the MLF model. In order to determine whether the scarcity of dichotomous ML clauses in those data was an issue with my methodology, I applied the same methodology of analysis to a small dataset of Austrian German-English data. I identified 14 dichotomous ML clauses in those data, representing a considerably higher proportion of clauses showing convergence than was found in the Welsh-English data.

I argue that this shows that the concept of a dichotomous ML language has validity, and it can be said to quantify the extent of word-order convergence in a dataset. If the model, with this adaptation, were applied again to a dataset which is already known to show convergence, then I would expect a substantial proportion of clauses from those data to have a dichotomous ML.

For the purposes of measuring the extent of word-order convergence in the speech of Welsh-English bilinguals, meanwhile, I have shown that it is very scarce, if not practically absent. According to the model, the Welsh-English bilinguals studied display classic code-switching rather than convergence when using their two languages. From the viewpoint of language change, which is difficult to examine with synchronic data and few participants, the MLF model indicates the

probability that widespread word-order convergence is distant on the horizon for the grammar of Welsh.

8. Conclusion

In this chapter I have shown the results of my application of the MLF model to a Welsh-English dataset consisting of six bilinguals, to see whether they display word-order convergence. The results clearly show that these speakers favour Welsh as a source for their clause structure, producing Welsh clauses more frequently than English clauses, but, more crucially, always using Welsh as the ML in clauses in which there is code-switching. Two clauses are found to have a dichotomous ML clause, one where a monolingual Welsh head/modifier NP has word order which is more prominent in English than it is in Welsh (where the modifier precedes, rather than follows, the head noun), and another where morphemes from both Welsh and English appear in an English-predominant word-order in a clause where other information points to structure being sourced from English. Both these are analysed as showing convergence. The scarcity of dichotomous ML clauses in the Welsh-English data would seem to indicate the rarity of word-order convergence in these speakers' grammars.

To show how my method of applying the MLF model can identify dichotomous ML clauses and convergence in data other than Welsh-English, I described an analysis of data collected from Austrian German-English bilinguals who had spent most of their lives in England. The same methodology applying the MLF model was used, and the results

showed that considerably more clauses in these data had a dichotomous ML, showing the influence of English on German word order. This I interpreted as convergence.

Whilst I have shown that my method can be effective in identifying convergence, I acknowledge that it is possible that the MLF model cannot find all instances of word-order convergence in a dataset. What, for instance, of clauses where there is deleted or ellipted material? What if convergence involves deletion? The MLF model works best with overt surface morphemes on which to base its identification of clausal ML, and so clauses showing morpheme deletion which might not have a dichotomous ML may nevertheless contain signs of convergence.

To this end, in the next chapter I discuss this apparent weakness in the MLF model, particularly with reference to a type of construction which appears to be common in spoken Welsh—auxiliary deletion of *bod* 'to be'—which I believe shows word-order convergence. I will analyse and examine a sample of such clauses from the Welsh-English corpus to see how common these constructions are, and whether or not they show word-order convergence despite the fact that they do not have a dichotomous ML.

Chapter 6 — Analysing clauses with auxiliary deletion for signs of convergence

1. Introduction and justification

In the previous chapter I applied the MLF model to clauses in a Welsh-English dataset to measure the number of dichotomous ML clauses found, and I identified one clause as apparently showing word-order convergence from Welsh toward English. The inference is that there is almost no word-order convergence in this dataset, because the number of dichotomous ML clauses found was so low.

However, a number of clauses were analysed in the data where only one of the two principles of the MLF model (either the MOP or the SMP) was testable on that clause, because there was insufficient material in the clause to test them. These clauses vary in type and construction, and none of them have a dichotomous ML, but many are similar, in that they lack an overt (Welsh) finite verb and have a clause-initial subject pronoun. In this chapter I will argue that these clauses appear to show convergence despite the fact they do not have a dichotomous ML. I will argue that this convergence reflects a shift in progress from VS to SV in Welsh main clause word order due to the influence of English SV order.

My focus will be on clauses with auxiliary constructions which have a null, or deleted, finite verb form of *bod* 'to be' and a subject 2nd singular pronoun *ti*, assessing how frequently they occur in the speech of Welsh-English bilinguals compared to non-deleted forms.[138] Table 5 below illustrates a standard verb paradigm for *bod* 'to be' in Welsh (adapted from King 1996:146) although of course there are many dialectal variations.

Number	Person	Form
Sg.	1st	ydw, wyf
	2nd	wyt
	3rd	mae, ydy
Pl.	1st	ydyn
	2nd	ydych
	3rd	maent, ydynt

Table 5. A paradigm for the Welsh finite verb forms of *bod* 'to be'.

I then proceed to discuss reasons for the occurrence of the phenomenon. I will also show a similar phenomenon which is reported to be occurring in another verb-initial language, the Central Salish of western Canada, also argued to reflect convergence (Shaw et al 2008), to indicate that this kind of linguistic phenomenon may not be isolated to Welsh.

As discussed in chapters 3 and 5, when applying the MLF model to clauses in the data, a clause's ML was identifiable if there was enough

[138] I justify my focus on constructions with *ti* as the subject in section 5 below.

evidence within that clause to satisfy either one or both of the Morpheme Order Principle and the System Morpheme Principle. Most clauses could have (and were expected to have) both principles satisfied by the material they contained. However, the data analysed in chapter 5 show that there are some clauses or clause fragments in which an ML is identifiable on the basis of just one principle. A certain subset of such clauses are clauses (or clause fragments) which have no overt finite verb in a construction where a finite verb would be expected, but do have an overt subject and nonfinite verb. Here the SMP is testable on the deleted verb's outside late system morphemes, based on the assumption that a finite verb has been deleted. An example is given in (1) below.

(1)	ti	'n	byw	yn	y	gym
	2S	PRT	live.NONFIN	in	DET	gym

"You live in the gym." [FUS27-MAB179]

In (1) there is no visible finite verb, but there is an overt subject (*ti* 'you') and an overt nonfinite verb (*byw* 'live'), so a finite auxiliary verb is expected in such a context to build a Welsh periphrastic verbal construction. I assume that there is a null verb here, which is presumably *wyt*, a form of *bod* 'to be' which matches *ti* in person and number. The language of this verb can be presumed from the Welsh aspectual pre-verbal particle *yn*, which is Welsh and typically found in Welsh construction using a finite form of *bod*. This satisfies the SMP and identifies the clause ML as Welsh. I suggest that the position of the finite verb in this clause is not identifiable, so the MOP cannot be tested on this clause. The clausal ML is identified as Welsh by virtue of the SMP.

The description of a dichotomous ML clause given in chapter 3 captures the premise that the two principles of the MLF model point to different language as supplying the ML, this indicates that both participating languages are a source for clause structure. Enough clause material is required to test both principles for a clause to be labelled with a dichotomous ML. If only one principle can be tested using the information present in a clause, e.g. the clause type outlined above where only the MOP can be tested, then a dichotomous ML cannot occur as a result of the application of the MLF model to that clause. The clause type shown in (1), then, does not have a dichotomous ML, but it does not necessarily follow that no such clauses show convergence. It could be the case, I argue, that clauses which were not assigned a dichotomous ML but on which only one principle could be tested nevertheless may show word-order convergence.

In the next section I will discuss this with respect to the clause type illustrated in (1) above, to see to what extent they can be argued to display signs of word-order convergence. While other types of clauses are found in the data on which only one principle can be tested,[139] those will not be the focus here.

2. Welsh-English clauses with a deleted auxiliary verb

Instances in the data of the clause type outlined above, where there is no overt finite verb, were analysed using the MLF model as containing

[139] For example, a clause fragment which has no finite verb but has a bilingual head/modifier NP, which could have the MOP tested on it.

null elements.[140] Consider a typical example of this kind of construction in a clause from the data analysed in the analysis discussed in chapter 5. It is an interrogative construction, shown in (2).

(2)	ti	'di	cwrdd	â	Lily?
	2S	PRT.PAST	meet.NONFIN	with	Lily
	"Have you met Lily?"				[FUS27-LIS73]

The clause in (2) has a 2nd person singular pronoun *ti* as its subject, and the only overt verb is nonfinite: the verbal noun *cwrdd* 'meet'. Lisa, the speaker, produces no audible finite verb, but a null finite auxiliary verb *wyt* is assumed, based on the presence of the perfect particle (*we*)*di*. The ML of (2) is identified as Welsh. Note that the initial overt element of the clause is the subject *ti*. Compare this clause with an alternative form which does have a finite verb, the 2nd singular present form of *bod* 'be' *wyt*, as shown in (3) below.

(3)	wyt	ti	'di	cwrdd	â
	be.2S.PRES	2S	PRT.PAST	meet.NONFIN	with
	Lily ?				
	Lily				
	"Have you met Lily?"				

[140] Myers-Scotton (2002:55) argues that the CP, the unit of analysis she favours, can contain null elements in any clause position, and cites English ellipsis phrases such as "What?" and "Never!" as examples of clauses containing nulls. Such clauses, she claims, are "simply monolingual CPs that contain a number of null elements" (ibid.). However, I suggest that the status of such clauses are not, as Myers-Scotton suggests, "clear" (ibid.), since the language from which null elements are sourced can surely not be clear, and so it cannot be certain whether or not a CP containing nulls is a monolingual or bilingual. Myers-Scotton further claims that null elements can be used to identify the ML of a clause (cf. 2002:57), but if, as I would suggest, the language from which null elements are sourced in a CP is not known, then it seems arbitrary to use unknowns as evidence for clausal ML. Clauses with null elements are, in my view, problematic for the MLF model.

Here there is a finite verb, *wyt*, inflected for person, number and tense to agree with the subject *ti*. This finite verb precedes the subject NP *ti*, in line with the VS main clause surface word order of Welsh, so in (3) it is the finite verb *wyt* which is the first overt element of the clause, rather than the subject. These morphemes form part of an auxiliary construction with the nonfinite verbal noun *cwrdd* 'meet' as the main verb. It is the finite verb *wyt* which is deleted to produce the construction found in (2). I will refer to this process from here on as auxiliary deletion (AD).

The chief point of interest here is that clauses in the data such as (1) and (2) above, i.e. which show AD, all have a clause-initial subject. Unmarked Welsh main clause word order is VSO. Elsewhere in Welsh, in clauses without null elements, subject-initial word order (SVO) is indicative of emphasis, as described in chapter 3. In English, however, SVO is a normal main clause order. It is possible to interpret clauses with AD (-A, auxiliary deleted) in these data as being more English-like than clauses in which the auxiliary verb is not deleted (+A, overt auxiliary).

Before proceeding to present an explanation for this phenomenon, I will assess in the next section the extent of AD found in the dataset analysed in chapter 5.

3. Examples of auxiliary deletion

There are clauses in data analysed in chapter 5 that contain an auxiliary construction involving the 2nd singular pronoun *ti* (which I shall henceforth refer to as B2S clauses[141]), some of which show retention (henceforth +A) of the 2nd singular present auxiliary verb *wyt* and some which show its deletion (henceforth -A). Examples of positive declarative constructions with –A are shown in (4), (5) and (6) below.

(4) ti mynd pasio 'r industrial estate
2S go.NONFIN pass.NONFIN DET industrial estate
"You go past the industrial estate." [DAV6-DAN827]
[+A form: **wyt** ti mynd pasio'r industrial estate]

(5) ti 'di siarad ybyty fo gyda fi
2S PRT.PAST talk.NONFIN about 3Sm with 1S
"You've talked about him with me." [FUS27-LIS328]
[+A form: **wyt** ti 'di siarad ybyty fo gyda fi]

(6) ti 'n gwybod
2S PRT know.NONFIN
"You know." [FUS6-ANT14]
[+A form: **wyt** ti'n gwybod.]

[141] i.e. the verb *bod* + a second person singular pronoun.

There is no example of a positive declarative clause with +A in this dataset.[142]

Examples in the dataset of positive interrogative constructions with both +A are given in (7) and (8), and examples with -A are given in (9) and (10). Interrogative constructions are identified on the basis of rising intonation and/or the presence of a WH morpheme.

(7) wyt ti 'di # cofrestru a
be.2S.PRES 2S PRT.PAST register.NONFIN and
bopeth felly er eto?[143]
everything thus IM again
"Have you registered and everything then, er, yet?"
[FUS27-MAB800]

(8) wyt ti byth yn darllen y
be.2S.PRES 2S ever PRT read.NONFIN DET
llyfrau...?
books
"Do you ever read the books....?" [FUS27-LIS374]

(9) ti gweld y stori 'ma?
2S see.NONFIN DET story here
"Do you see this story?" [DAV6-DAN726]
[+A form: **wyt** ti gweld y stori yma?]

---

[142] Note that the dataset is not large as far as auxiliary constructions of this type are concerbed. I believe Welsh speakers would in fact sometimes produce +A positive declarative clauses.

[143] Recall that the word order of questions like this one in Welsh, VSO, mirrors the word order of some interrogatives in English.

(10) faint o discount ti 'n gael?
amount of discount 2S PRT get.NONFIN
"How much of a discount do you get?" [FUS6-AMR236]
[+A form: faint o discount **wyt** ti'n gael?]

Negative constructions can be formed in a variety of ways in Welsh, some examples of which are seen in the examples below. The negative in (11) below is formed with both a negative prefix on the auxiliary verb (*d-wyt*) and a negative particle *ddim*, whereas when there is -A, as in (12) and (13), only the particle *ddim* signifies that the construction is negative.[144] The negative declarative construction in the second clause in (11) below, *dwyt ti'm yn licio plant*, is +A.

(11) oh dw i 'n gwybod
IM be.1S.PRES 1S PRT know.NONFIN
dwyt ti 'm yn licio plant,
be.2S.PRES.NEG 2S NEG PRT like.NONFIN children
Lisa!
Lisa
"Oh, I know that you don't like children, Lisa!
[FUS27-MAB661]

There are no examples of +A negative interrogatives found in these data.[145]

---

[144] There are other ways of forming a negative construction in Welsh, which will be discussed later in this chapter.

[145] As with +A positive declarative clauses, I believe the absence of +A negative interrogatives from this dataset is more indicative of the smallness of the dataset than of the rarity of this construction. The more extensive study presented in this chapter will address whether these two types of construction really are rare.

-A is found in clauses with both declarative (12) and interrogative (13) constructions, shown below.

(12) ti 'm yn cael <u>lectures</u> 'ma nag
2S NEG PRT get.NONFIN lectures here NEG
wyt[146]
be.2S.PRES
"You don't have lectures here, do you." [DAV6-HEC994]

[+A form: **dwyt** ti'm yn cael <u>lectures</u> 'ma nag wyt]

(13) ti ddim yn licio dreifio?
2S NEG PRT like.NONFIN drive.NONFIN
"You don't like driving?" [FUS6-ANT426]

[+A form: **dwyt** ti'm yn licio dreifio?]

In auxiliary constructions involving *bod* 'to be', a particle denoting aspect, viz. (*y*)*n* or (*we*)*di,* is expected to precede the nonfinite verbal noun (e.g. Borsley et al 2007:12), as seen in e.g. (5) and (6) above. However, this particle is sometimes completely deleted by speakers in B2S clauses. For example, (14) shows a +A positive interrogative construction with a (presumably) deleted *yn*, whilst (15) shows a -A positive declarative construction with a deleted particle (presumably *wedi* given the presence of the agreement morpheme *do*) and (16) a -A positive interrogative construction with a deleted particle.

[146] Arguably a clause which has a TAG 'question' at the end, e.g. *nag wyt* 'aren't you' here, could also be interpreted as being an interrogative. I have chosen not to interpret them as such for the purposes of this analysis, unless such clauses also use a WH morpheme and/or rising intonation.

(14) yeah wyt ti licio fo xxx be?
yeah be.2S.PRES 2S like.NONFIN 3Sm xxx what
"Yeah, do you like it [...] what?" [FUS27-MAB577]

(15) ti dorri fo do
2S break.NONFIN 3Sm yes
"You've broken it, haven't you." [DAV6-DAN114]
[+A form: (wyt) ti **wedi** dorri fo do]

(16) ti dal watsied o?
2S still watch.NONFIN 3Sm
"Are you still watching it?" [DAV6-DAN50]
[+A form: (wyt) ti dal **yn** watsied o?]

-A negative constructions with particle deletion are not found in these data.[147]

I will focus in this chapter on AD in present and perfect tense constructions,[148] and I will ignore the periphrastic construction clauses found in the dataset which have an overt imperfect or conditional (etc.) finite auxiliary. This is because it is not possible to identify the tense of a

[147] Presumably to be interpreted as another coincidence, rather than indicating that such a construction is impossible.

[148] Given that the present tense is often hypothesised to be the 'default' tense in language (e.g. Declerck 1991:69—it is "the tense that is used when the speaker wishes to describe a situation without bothering to locate it in time"), it seems apt to focus a study of a given phenomenon on clauses which are in some sense default (or unmarked) before progressing to examining that phenomenon in other (marked) types of clauses. The context of all the -A clauses examined here is consistent with the interpretation that the deleted auxiliary is a present tense verb, although note that clauses including a perfect tense construction formed using a present tense auxiliary and the particle *wedi* are included in this analysis, because they are overtly different to -A clauses which are present tense and have *yn*.

deleted verb, and I assume that every -A B2S clause is present tense by default.[149]

As discussed, there are many examples of B2S clauses with both -A and +A in the data on which the MLF model was tested in chapter 5. There are also numerous examples of -A clauses where the subject is not *ti*, though these are less prevalent. Examples are given in (17), (18) and (19) below. Example (17) has a deleted 1st person singular auxiliary verb, (18) has a deleted 1st person plural auxiliary, and (19) has a deleted 3rd person plural auxiliary.

(17) fi *just* yn dweud "toes"
1S just PRT say.NONFIN toes
"I just say 'toes'" [FUS27-:LIS1142]
[+A form: **wyf** fi just yn dweud "toes"]

(18) ni mynd i fynd rownd ffordd 'na
1PL go.NONFIN PRT go.NONFIN round way there
i [//] ar gyfer o
to for 3Sm
"We're going to go round that way for it" [FUS27-LIS856]
[+A form: **ŷn** ni mynd i fynd rownd ffordd 'na i [//] ar gyfer o]

[149] I do not imply that -A B2S clauses of a different tense are impossible; see e.g. the example in footnote 117, which, whilst it has a 3rd singular masculine pronominal subject, appears to perhaps have a deleted imperfect tense auxiliary. Ways speakers differentiate between tenses in -A clauses will be an avenue for future study.

(19) a nhw 'n edrych mor *cute*

and 3PL PRT look.NONFIN so cute

"And they look so cute." [FUS27-LIS131]

[+A form: a **maen** nhw'n edrych mor *cute*]

Note that these three examples are produced by the same speaker, Lisa, born in south Wales, and similar examples produced by speakers born in north Wales are much less frequent.[150]

Furthermore, in (20) and (21) below, -A is found in clauses where the subject is a noun or demonstrative pronoun. In (20) the subject is a proper noun *Lily*, whilst in (21) the subject is the feminine demonstrative pronoun *honno*, here indicating a woman. In both examples the 3rd singular present form *mae* is apparently deleted.

(20) Lily 'n byw 'da Kristen

Lily PRT live.NONFIN with Kristen

"Lily lives with Kristen." [FUS27-LIS82]

[+A form: **mae** Lily'n byw 'da Kristen]

(21) honno casáu fi ydy

that-one.F hate.NONFIN 1S be.3S.PRES

"She hates me, doesn't she." [DAV6-DAN980]

[+A form: **mae** honno ('n) casáu fi ydy]

---

[150] Nevertheless AD involving pronouns other than *ti* is sometimes found in northern speech in the data. For example, in Davies6, Hector produces the following –A clause:

o	'di	dal	hi	ddigon	ciaidd	y	nos
3SM	PRT.PAST	catch.NONFIN	3SF	enough	brutal	DET	night
Sadwrn	'lly!						
Saturday	thus						

"He was quite fiercely drunk Saturday night, then!" [DAV6-HEC552]

[+A form: **oedd** o 'di cael hi ddigon ciaidd y nos Sadwrn 'lly!]

Note that whilst (20), like (17), (18) and (19), is spoken by Lisa, who is from south Wales, (21) is spoken by Daniel, a young man from north Wales, and there are other examples in these data of AD before a proper/lexical noun spoken by speakers from north Wales.

I have shown that B2S clauses in the dataset analysed for chapter 5 display both +A and -A, indicating that there is variation by speakers in this regard. I have also shown that it occurs in several different contexts, including both affirmative and negative constructions and declarative and interrogative constructions. In the study I present below I will examine the variation of AD according to these linguistic variables.

Having shown that -A is prevalent in the speech of speakers in a small dataset I have already analysed, proceed now to discuss the previous studies that have been conducted into this phenomenon in Welsh. Before coming to that, I will briefly clarify my position on the finiteness of clauses which have –A.

a. –A clauses: finite or nonfinite?

Even though –A clauses result in the deletion of the finite verb, I do not assume that this makes them nonfinite clauses.[151] Note that Welsh speakers already interpret some clauses which lack an inflected verb to be finite: cf. e.g. what Borsley et al (2007:81) call "*i*-clauses", which are dependent clauses which have an initial *i* morpheme (which is homophonous with the preposition *i* 'to') and a general SVO word order

[151] Recall that for the analysis presented in chapter 5 I analysed all –A clauses as being finite.

(where V is nonfinite).[152] An example of this kind of clause is given in (22).

(22)	meddyliodd	Aled	**i**	**Mair**	**fynd**		**adre**
	think.3S.PAST	Aled	to	Mair	go.NONFIN		home

"Aled thought that Mair had gone home"

(taken from Borsley et al 2007:81)

In the dependent clause in (22) *Mair* is the subject and *fynd* is the V. Though *fynd* is nonfinite, Borsley et al (2007:82) argue that *i*-clauses which are complements to epistemic or declarative predicates are "interpreted as tensed", even though the verb with inflectional marking (*meddyliodd* in [22] above) is in the main clause and not the *i*-clause. Borsley et al (2007:98) conclude that morphological tense marking "is not a necessary condition for finiteness" in Welsh clauses.[153]

If this general facet of Welsh is accepted, it may be reasonable to infer that Welsh speakers are able to interpret –A B2S clauses as being finite even though they have no morphological tense marking, given that the inflectionally-marked finite auxiliary verb has been deleted. In light of this it is appropriate to label both +A and –A clauses as finite.

4. Review of the literature on AD

AD in Welsh has not been widely discussed in the research literature to date.

---

[152] Strictly speaking the word order is [I-S-V(nonfinite)-O/Complement-Adjunct] (Borsley et al 2007:81).

[153] Indeed, verbless clauses in Welsh can also be interpreted as being finite (cf. Borsley et al 2007:66–7).

Borsley et al (2007) note that a feature of *bod* is the omission of its finite form in clause-initial position (i.e. AD) with certain pronominal subjects, such as *ti*.[154] They suggest that sentences showing AD have a phonologically-empty form of *bod* (2007:260), and support this with evidence from associated tag questions, as in (23) below, where the finite tag form is overt (undeleted) but the main clause auxiliary is not overt (deleted).

(23)	ti	'n	mynd,	ynd	wyt?
	2S	PRT	go.NONFIN	Q.NEG	be.2S.PRES[155]
	"You're going, aren't you?"				

Borsley et al do not elaborate on the specific variation or extent of AD in spoken Welsh.

King (1996), claiming to represent "written colloquial Welsh" based on native speech patterns (1996:3), cites the 2nd person singular present form of *bod* as *ti*, i.e. no overt auxiliary, as the form "most frequently heard" for declaratives by both northern and southern speakers of Welsh. An example King gives is in (24)

(24)	ti	'n	edrych	yn	union	fel	dy
	2S	PRT	look.NONFIN	PRT	exact	like	POSS.2S
	dad						
	father[156]						

"You look exactly like your father." (taken from King 1996:147)

---

[154] Borsley et al (2007:260–1) also note that AD is found where the pronoun subject is 1st person plural *ni* and 2nd person plural *chi*, as well as 1st person singular *fi* "in the speech of some speakers of southern dialects".

[155] My gloss.

[156] My gloss.

King says this is also the form used often for interrogative constructions (1996:147), although the paradigms he provides give *wyt ti?* as the usual interrogative form (1996:146). Jones and Thomas (1977:23), writing twenty years earlier, also give *'ti* (their apostrophe) as the contracted spontaneous spoken form, at least for northern speakers, and give *'rwyt ti* as the full spontaneous spoken declarative form; I give a hypothetical example of this form in (25).

(25)	r-wyt	ti	yn	mynd	
	DET-be.2S.PRES	2S	PRT	go.NONFIN	
	"You're going."				

King (1996:147) agrees with *(r)wyt* being the full form of the verb when it is heard ("perhaps the only *r-* form of the verb that truly is part of the spoken language"). Thorne (1996), however, gives *wyt* or *ydwyt*[157] as the form (1996:257), and cites sentence (26) below (1996:272).

(26)	yr	wyt	ti	'n	dysgu
	DET	be.2S.PRES	2S	PRT	teach.NONFIN
	"You teach."				

Thorne apparently believes AD is not part of the usual grammar of Welsh speakers. Thorne (1996) is a grammar based on written sources, though he claims his description represents "Cymraeg cyfoes", contemporary Welsh (1996:11). A more recent work by Thorne

[157] *Ydwyt* is a synchronically a variant of *wyt*, along with e.g. 1st person singular *ydwyf* (*fi*), 2nd person plural *ydych* (*chi*), etc. Such forms are not to be confused with forms such as *dwyt*, which are negative.

(2000:65) seems to recognise that sentences such as (27) exist, which he says may be shortened to (28), but otherwise AD is not mentioned.

(27) ti 'n gweld
2S PRT see.NONFIN[158]
"You see."

(28) t' wel'
2S see.NONFIN[159]
"Y'see."

Earlier Welsh grammar books all give 'full' forms of the 2nd person singular present form of *bod*, i.e. *wyt* or *ydwyt* in Williams (1980), *rwyt ti* in Uned Iaith Genedlaethol Cymru (1976), *wyt* or *ydwyt ti* in Rowland (1853), and *wyt* or *ydwyt (yn caru)* 'you love' in Spurrell (1870). None of these earlier works suggest that AD occurs in this context in Welsh, though it may be assumed that these earlier works describe written rather than colloquial Welsh.

Whilst most authors writing on Welsh grammar, particularly those writing some time into the past, state that the 2nd person singular present form of *bod* is *(r)wyt* or *ydwyt*, more recent authors describe the contracted form *ti*, i.e. the AD form. It could be inferred from this that AD of *wyt* in Welsh is a fairly recent, variable phenomenon.

While I have shown above that AD may be found in clauses involving a pronoun subject other than *ti*, this is not commented on by most of the authors above. However, Borsley et al (2007:261) note the presence of AD in clauses where the subject is *ni*, *chi* or *fi*, while King

[158] My gloss.

[159] My gloss.

(1996:146) cites AD as being possible in 1st plural forms of the *bod* construction: he gives the 1st plural present form for southern varieties of Welsh as *(ŷn) ni* (his bracketing), which I interpret as King noting that -A of *ŷn* is also sometimes found. For other pronominal subjects, King gives the full +A forms, from which I infer that he does not indicate AD is usual in these contexts. Other authors (e.g. Jones and Thomas) give no indication that AD occurs in clauses involving other pronominal subjects.

References to AD in linguistic studies of colloquial Welsh are rare, but examples from phonetically-transcribed data, collected by Roberts (1988:119, based on data collected for her research from 1973), of Welsh-English bilinguals living in Pwllheli, north Wales, include constructions such as that in (29) below.

(29) [ɬe ti n mɨnd i?]

=	lle	ti	'n	mynd	i?
	place	2S	PRT	go.NONFIN	to[160]

"Where are you going to?"

The clause in (29), spoken by a participant of unspecified age and sex but native to the town, shows clear deletion of the auxiliary verb *wyt* which would be expected to be positioned between *lle* and *ti*. Roberts notes the effect of English contact on the construction, but only in the stranding of the preposition *i* (paralleling English constructions like *where are you going to?*. She does not mention the AD in the clause. This suggests two things.

[160] My gloss of Roberts' broad phonetic transcription.

i. AD in this context was present in a north Welsh dialect as early as 1973, when the data were collected, at least in some constructions;

ii. the AD in this clause, and presumably AD in general, was not considered unusual enough a phenomenon to warrant comment by a researcher writing in 1988.

Phillips (2007) is one of the few authors to specifically discuss AD in Welsh, analysing primarily the speech of Welsh-speaking children. He uses data collected by B.M. Jones in the 1990s, as well as some data from the Welsh-English Bangor Pilot corpus.[161] An example of a -A clause Phillips gives is shown in (30) below.

(30)	ti	'n	gweud	popeth	(w)y(f)	fi
	2S	PRT	say.NONFIN	everything	be.1S.PRES	1S
	'n	hoffi[162]				
	PRT	like.NONFIN				

"You're saying everything that I like."

(from Phillips 2007:274)

[+A form: **wyt** ti'n gweud popeth (w)y(f) fi'n hoffi]

Phillips notes that constructions such as (30) are common in the data he analysed (2007:274). He argues that this sort of construction in the output of young Welsh speakers reflects a shift in Welsh main clause order from VSO to SVO. Such a shift, Phillips argues, is partially influenced by the existence of subject-initial clauses in Welsh, as I

---

[161] These data are located on the same website as the Siarad corpus, http://www.talkbank.org. They were collected and transcribed in 2004–5.

[162] My gloss.

discussed in chapter 5, but he claims that it is also internally-motivated. I attempt to summarise his proposal for the internal conditions that lead to AD below.

For Phillips (2007:275), in constructions where there is +A auxiliary verb which is a form of *bod* 'to be', this auxiliary has a limited function. It conveys tense as well as what Phillips calls "bodolaeth" ('existence', presumably related to the literal meaning of *bod*), but the clause's primary meaning is conveyed by the verbal noun (*gweud* 'say' in [30] above). Phillips concludes that, in the speech of young Welsh-English bilinguals, the auxiliary verb (e.g. *wyt*) is subsequently reanalysed as a verbal particle. Phillips (ibid.) suggests that primary stress then shifts from the auxiliary verb to the aspectual particle *yn* or *wedi*, e.g. *wyt ti yn mynd* 'you are going', paralleling English patterns such as *he is going* (Phillips' emphasis). The next step in Phillips' proposed sequence of word-order change is the deletion of the initial auxiliary verb (or "particle") due to the primary stress shift, and presumably because it has become optional for some speakers, resulting in the subject becoming the first overt clause-element, as in (30). This is what I call AD. A logical further step would be for Welsh to be reanalysed as having SVO as unmarked main clause order, perhaps with the verbal noun becoming analysed as the main verb of clause (Phillips ibid.).

I agree in principle with Phillips' argument that stress shift is partially behind the existence of subject-initial clauses of this kind in Welsh, though I think it unlikely that the primary shift moves to the particle *yn/wedi* in such clauses. Having listened to the corpus data, none of the auxiliary-deleted clauses cited thus far in this chapter or in chapters 5 have primary stress on the aspectual particle *yn* or *wedi*, except in a context where emphasis is suggested (e.g. *ti yn mynd* 'you

are going'). Though an acoustic analysis is beyond the scope of the present study, I have observed that, in these data, the main stress in unmarked constructions appears to fall most commonly on the nonfinite verbal noun. I disregard Phillips' assertion of where the stress has moved to, but agree that an argument for a stress shift away from the auxiliary verb is persuasive. I concur with Phillips' suggestion that English SVO main clause word order is influencing word-order shift in Welsh from VS to SV, and will examine it below.

Existing literature on AD, therefore, indicates that this phenomenon is present in some varieties of Welsh, and, according to some sources, particularly in the speech of young bilinguals, but no quantitative study on the extent of AD in Welsh currently exists. This is an issue I intend to address in this thesis, by presenting a comprehensive study of AD in a subset of Welsh-English data taken from Siarad comprising of 28 bilinguals.

5. Hypothesis

To recap: I have shown in this chapter so far how, in data from six Welsh-English bilinguals analysed in chapter 5, the auxiliary *wyt* in auxiliary constructions where the subject is *ti* is sometimes found deleted (-A) and sometimes not (+A). This is found in both declarative and interrogative constructions, both affirmative and negative constructions, and both where there is and where there is not a pre-verbal particle *yn* or *wedi*. Since -A Welsh clauses have a clause-initial subject, I have suggested (cf. Phillips 2007) that this may reflect the influence of English SVO word order in the output of Welsh-English bilinguals.

While I have found that AD occurs in clauses with any pronoun or a lexical noun as the subject, I will focus in this analysis on constructions where the subject is the pronoun *ti* (B2S clauses), because there appears to be more variation in speakers' production of B2S clauses with regard to +A or –A than with other subjects. Also, given that the conversations in these data took place in a familiar discourse context between (usually) two people, B2S constructions are more common than those with other subjects, thus permitting a more comprehensive study of AD in constructions using this particular pronoun.

I present my hypothesis for this chapter here. AD is a form of variation in the speech of Welsh-English bilinguals. That change is the increasing selection of -A over +A in B2S clauses, which is (at least partly) triggered by convergence to a language in contact, English, which has SVO unmarked main-clause word-order. This influences deletion of the clause-initial Welsh finite verb to result in that clause having an initial subject. I will look for evidence of this by analysing a greater set of data (described below) from the Siarad Welsh-English corpus, and conduct variation analysis based on the independent variable of age. Evidence of language change in progress—word-order shift from VSO to SVO—will be found if a different frequency of –A is found in the speech of younger bilinguals compared to the speech of older bilinguals, e.g. a higher frequency of -A in younger bilinguals than in older bilinguals will be interpreted to indicate an increase in –A in Welsh.

My study will consist of a quantitative analysis of AD in a dataset, focusing on the extent to which AD occurs in given linguistic contexts. I will also analyse AD with regards to independent extralinguistic variables, particularly the age of the speakers. I then proceed to assess

whether or not –A clauses indicate word-order convergence in these data.[163] I will do this by considering to what extent the clauses with –A in these data pattern with English SVO word order.

Note that, for the sake of economy, I exclude clauses from this analysis which have the pronoun *ti* and a finite verb if they are not auxiliary constructions. I therefore do not include -A clauses which contain adjectival or prepositional constructions where the deleted finite verb is a 2nd person singular form of *bod* (as shown in [31] below) in my analysis.

(31) ti 'n lawn?
2S PRT fine
"Are you alright?" [DAV6-HEC292]
[+A form: **wyt** ti'n lawn?]

Given that –A in such constructions as this do occur in these data, it could be the case that AD is not limited to auxiliary constructions. It may be, for example, that AD is available in all Welsh-English clauses where the finite verb is a form of *bod*. While I focus on B2S auxiliary clauses in this analysis, I will bear the above in mind in the discussion below.

6. Methodology of analysis

In chapter 5, where I described applying the MLF model to a data from six speakers in order to identify clauses with a dichotomous ML, I

[163] My focus is on B2S clauses. Nevertheless, I will also briefly comment on AD in clauses where the subject is a pronoun other than *ti* or where it is a lexical noun.

explained that I had used three recordings/transcriptions totalling 88 minutes 24 seconds, or 3275 clauses. For this analysis of AD, however, I chose to analyse the output of a greater number of speakers representing a wide range of ages, genders, etc., so as to provide a reliable indicator of the presence or not of language change in these data.

Using data from the corpus described in chapter 4, I analysed transcribed conversations by 28 speakers from 13 recordings.[164] Each recording consisted of a conversation between two people, with the exception of Davies10 and Davies11, which each involved three speakers. As with the speakers analysed in chapter 5, all the participants are Welsh-English bilinguals. They vary in age from 12 to 81[165] and come from various regional backgrounds (though primarily north Wales, where 18 of the 28 speakers were raised). 14 speakers are male and 14 are female.

Table 6 below gives details of all 28 speakers. The relevant independent variables are noted for each speaker.

Speaker	Recording	Gender	Age	Region they lived in 1st year of life
BRW	Fusser31	M	12	Mid-Wales
NON	Davies1	F	18	South Wales
SAR	Davies1	F	19	South Wales
HEC	Davies6	M	23	North Wales
DAN	Davies6	M	25	North Wales
LON	Fusser30	F	25	North Wales

[164] The recordings were selected so as to provide as equal a distribution as possible of males and females and across three age groups—as detailed below in section 7.c.iii.

[165] The mean speaker age is 43.

MOR	Robert6	F	27	South Wales
MEL	Fusser30	F	28	North Wales
OLW	Fusser19	F	28	North Wales
NER	Stammers3	F	33	North Wales
SND	Stammers5	F	36	South Wales
GUT	Stammers3	M	37	North Wales
TRE	Fusser19	M	38	England
RHO	Stammers5	M	39	North Wales
MRL	Fusser15	F	40	South Wales
BEL	Fusser14	M	43	North Wales
ARF	Fusser31	M	43	North Wales
AWE	Fusser14	F	47	South Wales
GFR	Fusser15	M	50	North Wales
MIC	Davies10	M	52	North Wales
RAC	Davies11	F	52	South Wales
EIR	Robert6	F	56	South Wales
CLE	Davies10	M	58	North Wales
HIL	Davies10	F	62	North Wales
OWA	Davies11	M	67	North Wales
DER	Davies11	F	72	North Wales
EML	Robert8	M	79	North Wales
GOR	Robert8	M	81	North Wales

Table 6. The 28 speakers analysed for variation in auxiliary deletion.

A transcript (transcribed using the CHAT system, as before) of each recording was searched for occurrences of clauses involving the pronoun *ti* using the KWAL function in CLAN. After manually eliminating those constructions which were not auxiliary constructions (see example

35 above), the remaining B2S clauses were entered into a Microsoft Excel spreadsheet file, and separated according to recording and speaker. For each B2S clause produced by a given speaker, it was noted in the analysis spreadsheet whether or not there was an overt auxiliary verb. Identification of this (-A or +A) was based on the transcripts.[166] The spreadsheet was also used to note whether or not clauses were declarative or interrogative constructions, affirmative or negative. Details concerning the pre-verbal particle *yn* or *wedi* was also noted, i.e. whether or not this particle was deleted or retained in the construction. A sample of this analysis spreadsheet is provided in Appendix D.

After entering the data and analysing each clause for AD, I used equation functions in Microsoft Excel to categorise the data[167] and to produce statistics representing the ratio of -A to +A across the data. I also used SPSS to conduct Chi-square statistical analyses on the data which indicate statistical significance. These results are presented in the next section.

---

[166] I rely on the transcribers having transcribed whether or not there was an audible auxiliary verb in such constructions.

[167] This included using the function COUNTIF, which counts the number of clauses where a certain trait is found (e.g. deletion of the auxiliary), and provides a total over a set of data. Thus frequencies for individual speakers, individual recordings and for specific variables could be calculated from one core clause analysis spreadsheet.

7. Results

b. Overall results

I present here the overall results for the dataset. 648 B2S clauses were produced in total by all the 28 speakers analysed.[168] Of the 648 clauses analysed, 601 (92.75%) clauses showed -A. The remaining 47 (7.25%) clauses showed +A.[169] Figure 11 below shows these results.

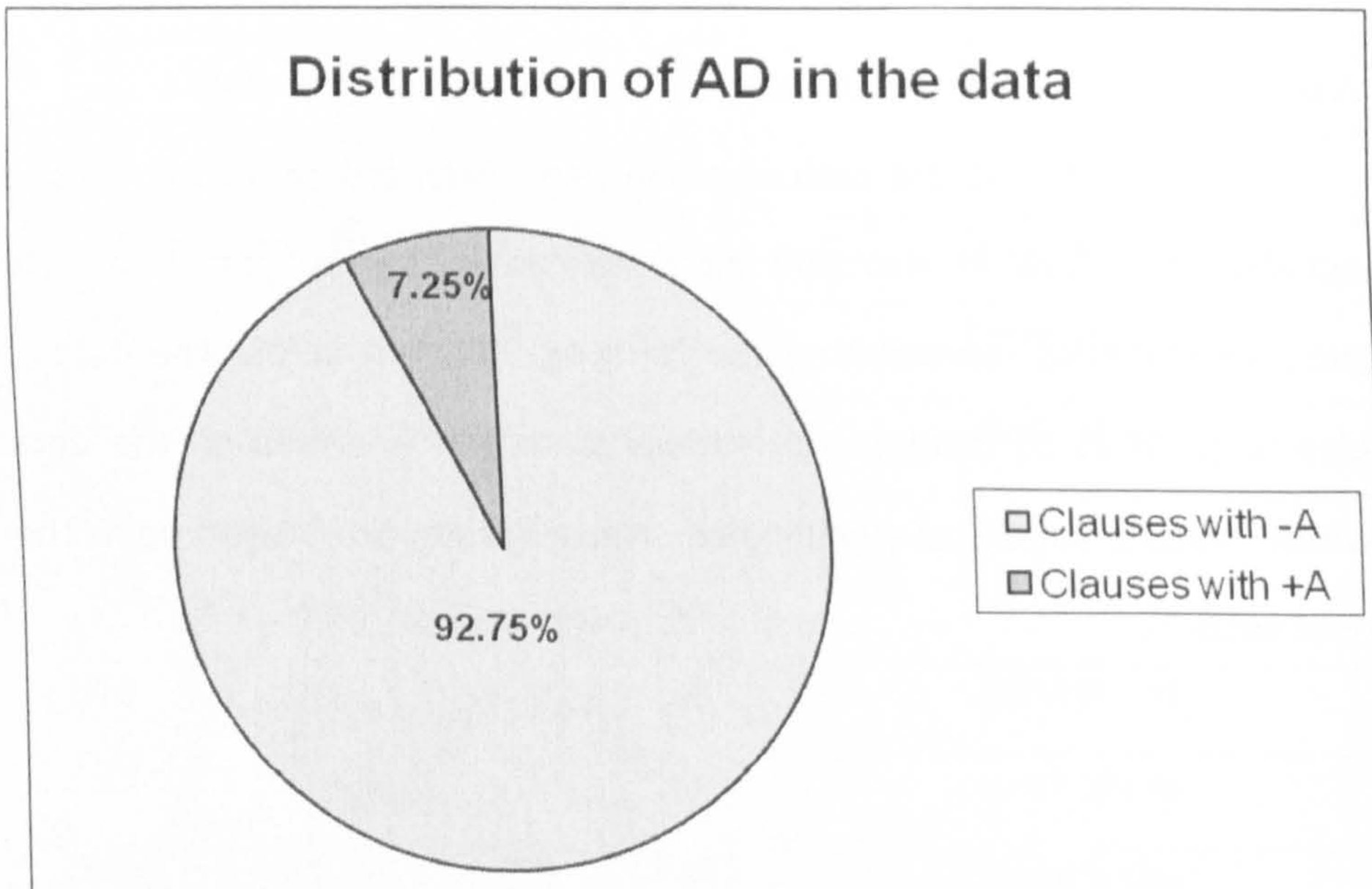


Figure 11. The proportion of +A (auxiliary retained) versus –A (auxiliary deleted) B2S clauses in the dataset as a whole.

This overall trend shows clearly that the speakers analysed delete the finite auxiliary verb in their B2S constructions in the great majority of

[168] This does not include clauses with an overt form of the verb *bod* in imperfect, perfect or conditional tense.

[169] The number of B2S clauses produced by an individual speaker ranged from 1 (DER) to 77 (ARF), with each speaker producing a mean 23.14 B2S clauses. Each speaker produced a mean 21.46 -A clauses, compared to a mean 1.68 +A clauses; see below for a more detailed discussion of individual speakers' AD patterns.

clauses (92.75%), indicating that -A is the norm in B2S clauses for these speakers.

Examples of clauses with +A and -A, involving various types of constructions, found in my data are shown below in examples (32) through (46).

Affirmative declarative constructions:

(32) ±A

wyt	ti	'n	ddistaw	lawn	wan
be.2S.PRES	2S	PRT	quiet	very	now
d-wyt					
NEG-be.2S.PRES					

"You're very quiet now, aren't you." [ROB8-EML3322]

(33) ±A

a	wedyn	wyt	ti	mynd	i
and	then	be.2S.PRES	2S	go.NONFIN	to
chwith					
left					

"And then you go to the left." [DAV11-OWA1479]

(34) ±A

wyt	ti	'n	gorod	mynd		i	
be.2S.PRES	2S	PRT	must	go.NONFIN		to	
rywle	lle	mae		nhw	'n	roid	[//]
somewhere	place	be.3PL.PRES		3PL	PRT	put.NONFIN	
sgrwbio		chdi	drosodd		gynta		
scrub.NONFIN		2S	over		first		

"You have to go to somewhere where they put... scrub you over first." [FUS30-MEL978]

(35) -A

ti	mynd	pasio	'r	industrial	estate
2S	go.NONFIN	pass.NONFIN	DET	industrial	estate

"You go past the industrial estate." [DAV6-DAN827]

[+A version: **wyt** ti mynd pasio'r industrial estate]

(36) -A

ti	'di	siarad	ybyty	fo	gyda	fi
2S	PRT.PAST	talk.NONFIN	about	3SM	with	1S

"You've talked about him with me." [FUS27-LIS328]

[+A version: **wyt** ti 'di siarad ybyty fo gyda fi]

(37) -A

ti	'n	gwybod
2S	PRT	know.NONFIN

"You know." [FUS6-ANT14]

[+A version: **wyt** ti'n gwybod]

Negative declarative constructions:[170]

(38) ±A

os	nad	wyt	ti	'n	*glear*	o	dy [//]
if	NEG	be.2S.PRES	2S	PRT	clear	of	POSS.2S
beth	ydy	dy		brief	dy		hun
what	be.3S.PRES	POSS.2S		brief	POSS.2S		self

"If you're not clear about what your own brief is."

[FUS15-GFR1705]

(39) -A

ond	ti	ddim	wedi	sylwi	hynny
but	2S	NEG	PRT.PAST	realise.NONFIN	that

"But you haven't realised that." [FUS14-BEL422]

[+A version: ond **d-wyt** ti ddim wedi sylwi hynny]

(40) -A

ti	'm	isio	sefyll	'na	am	hanner
2S	NEG	want	stand.NONFIN	there	for	half
awr						
hour						

"You don't want to stand there for half an hour."

[FUS15-MRL4653]

[+A version: **d-wyt** ti'm isio sefyll 'na am hanner awr]

---

[170] Note that I first analyse embedded clause negative declarative constructions together with main clause negative declarative constructions for convenience. I discuss the variation of AD according to different kinds of negation morpheme (e.g. main clause *nid* versus embedded clause *nad*) below.

Affirmative interrogative constructions:

(41) +A

wyt	ti	'n	agree-io	efo	be
be.2S.PRES	2S	PRT	agree.VBZ	with	what
mae	nhw	'n	trio	wneud?	
be.3PL.PRES	3PL	PRT	try.NONFIN	do.NONFIN	

"Do you agree with what they're trying to do?" [FUS19-OLW1943]

(42) +A

wyt	ti	'di	gweld	e?
be.2S.PRES	2S	PRT.PAST	see.NONFIN	3Sm

"Have you seen it?" [ROB6-MOR2073]

(43) -A

ti	'n	gwybod	lle	mae	hwnna?
2S	PRT	know.NONFIN	place	be.3S.PRES	that one

"Do you know where that is?" [STA5-SND1949]

[+A version: **wyt** ti'n gwybod lle mae hwnna?]

(44) -A

ti	isio	siarad	efo	fi	rwan?
2S	want	talk.NONFIN	with	1S	now

"Do you want to talk with me now?" [FUS14-BEL345]

[+A version: **wyt** ti isio siarad efo fi rwan?]

Negative interrogative constructions:[171]

(45) -A

ti	'm	yn	meddwl	'sa	fo
2S	NEG	PRT	think.NONFIN	be.3S.CONDIT	3SM
'n	byd	boring	iawn?		
PRT	world	boring	very		

"Don't you think it would be a very boring world?"

[FUS19-OLW3025]

[+A version: **d-wyt** ti'm yn meddwl 'sa fo'n byd boring iawn?]

(46) =A

ti	'm	'di	synnu?
2S	NEG	PRT.PAST	wonder.NONFIN

"You're not surprised?" [STA3-NER2730]

[+A version: **d-wyt** ti'm 'di synnu?]

c. Auxiliary deletion according to speaker

I now turn to examining the AD activity by individual speakers in the data, comparing them to the mean for the speaker set as a whole as shown above. Figure 12 below shows the proportion of B2S clauses with -A produced by each speaker. Table 7 below gives more details on speakers' AD activity, including the total number of B2S clauses a

[171] In these data I found no +A clauses which were negative interrogative constructions; see section 7.d.iii below for a comment on this.

speaker produced and what proportion of those had +A/-A, as well as the gender and age of each speaker.

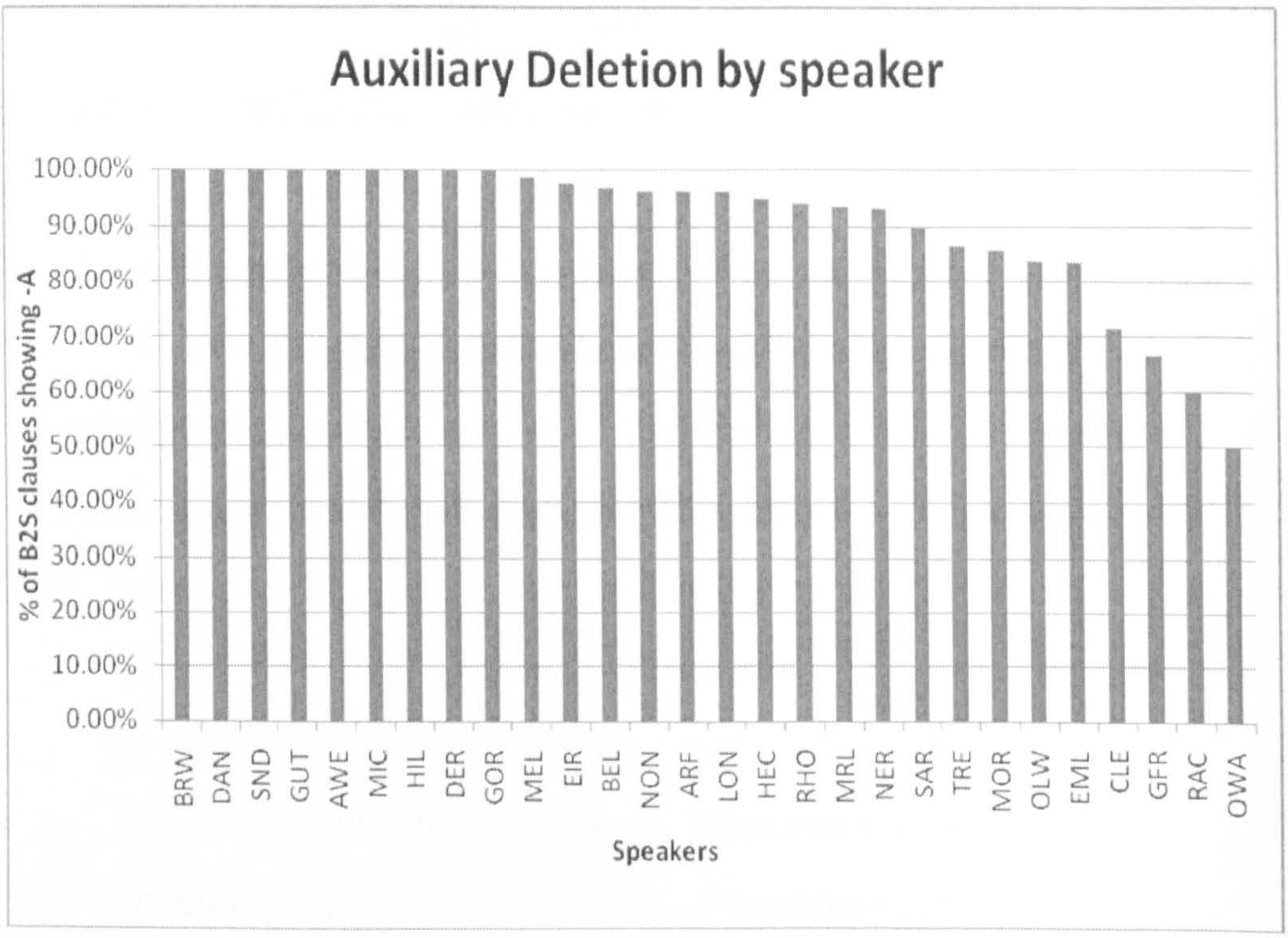


Figure 12. The distribution of clauses with +A versus −A according to speaker.

Speaker	Gender	Age	# B2S clauses	# +A	# -A	% -A
BRW	M	12	31	31	0	100.00%
DAN	M	25	23	23	0	100.00%
SND	F	36	22	22	0	100.00%
GUT	M	37	12	12	0	100.00%
AWE	F	47	26	26	0	100.00%
MIC	M	52	2	2	0	100.00%
HIL	F	62	3	3	0	100.00%
DER	F	72	1	1	0	100.00%
GOR	M	81	4	4	0	100.00%
MEL	F	28	62	61	1	98.39%
EIR	F	56	37	36	1	97.30%
BEL	M	43	28	27	1	96.43%
NON	F	18	25	24	1	96.00%

ARF	M	43	77	74	3	96.10%
LON	F	25	25	24	1	96.00%
HEC	M	23	19	18	1	94.74%
RHO	M	39	32	30	2	93.75%
MRL	F	40	30	28	2	93.33%
NER	F	33	29	27	2	93.10%
SAR	F	19	19	17	2	89.47%
TRE	M	38	22	19	3	86.36%
MOR	F	27	21	18	3	85.71%
OLW	F	28	49	41	8	83.67%
EML	M	79	6	5	1	83.33%
CLE	M	58	7	5	2	71.43%
GFR	M	50	27	18	9	66.67%
RAC	F	52	5	3	2	60.00%
OWA	M	67	4	2	2	50.00%

Table 7. The distribution of clauses with +A versus –A according to speaker.9

Nine speakers were found to produce -A in 100% of their B2S clauses (BRW, DAN, SND, GUT, AWE, MIC, HIL, DER, GOR). Most of the remaining speakers produce a high percentage of –A clauses. No speaker produces fewer than 50% -A clauses. The speakers in general produce –A B2S clauses frequently.

So far I have looked at overall AD patterns in the data, finding that the average proportion of -A is very high, and I have examined AD patterns from speaker to speaker, finding that most speakers have a high rate of -A. Next I will look at a range of extralinguistic variables and consider the effect they have on AD patterns in the output of these speakers. The three variables I consider are gender, place of birth and age. Note that these are independent variables that have been shown to correlate to linguistic variables (e.g. Tagliamonte 2006).

d. Independent variables

i. Results according to gender

The first independent variable analysed is the gender of the speaker. Of the 28 speakers analysed, 14 were male and 14 were female. The male speakers show a mean 88.49% -A, whilst the females show a mean 92.36% -A. Whilst females in these data appear to delete the auxiliary a little more frequently than males, there is no significant difference in AD activity in regards to the gender of the speaker.

Labov (2002) highlights the importance of gender over other sociolinguistic factors as a significant variable in language variation. His work indicates a correlation with gender, including in situations of language change (e.g. Labov 1990:210–13 and 215, 2001:366). Labov's (1990, 2001 etc.) study of phonological change by English speakers living in Philadelphia in the 1970s shows that variation was female-dominated and that women were shown to be "a full generation ahead of men" in many phonological changes (Labov 2001:501). Maclagan, Gordon and Lewis' (1999) study of stigmatized and non-stigmatized diphthongs in New Zealand English indicates that middle-class women are found to be more conservative than men in use of stigmatized forms but more innovative in use of non-stigmatized forms (1999:38). The study of historical English by Nevalainen and Raumolin-Brunberg (2003) suggests that women in the early modern English period were proactive in the spread of innovative forms such as *my* and *thy* instead of *mine* and thine, a change that started in the English of the lower classes which was predominated by men in the authors' data (Nevalainen and Raumolin-Brunberg 2003:119–20). Cheshire's (1998) analysis of the

English of adolescents living in Reading found that girls used nonstandard forms of English (e.g. the use of *ain't* for the auxiliary *have* and *be*) less frequently than boys, and thus there was a correlation between gender and use of nonstandard English (1998:38). These studies all give weight to the Labovian hypothesis that gender correlates with age in linguistic variation. It is noteworthy, then, that in my data there is no significant gender variation of AD in B2S clauses, and therefore differs from the situation found in many other communities.

An explanation for this lack of correlation might be that there are insufficient tokens to indicate gender variation—it may be the case that more data is required to show significant gender variation, and, indeed, several of the studies I cite above are concerned with phonological variables, of which there will usually be many tokens in a dataset. In the future it may be worthwhile to analyse data from more Welsh-English bilinguals (e.g. from other conversations in the Siarad corpus) in order to see if gender is found to correlate with AD where more tokens are available.

ii. Results according to the region where a speaker lived first

The second independent variable analysed is the region where the speakers lived for the first year of their life.[172]

---

[172] Note that many of the speakers analysed have not lived in the same area as they were born in. TRE, for example, who was born in England, moved to Wales when he was about 6 years old; RAC was born in south Wales but had lived in north Wales for over 20 years by the time she was recorded). Nevertheless, I interpret the region in which the speaker

Of the 28 speakers analysed, 18 lived in north Wales, 8 in south Wales, one in mid-Wales and one in England for the first year of their life. Speakers who spent the first year of their life in north Wales showed a mean 90.20% -A whilst speakers who spent the first year of their life in south Wales showed a mean 90.23% -A. Speakers from the north and south in these data therefore produce –A with almost identical mean frequencies, indicating that the region where a speaker grew up has no effect on the frequency of -A in B2S clauses. [173]

The other two regional subgroups only contain one speaker each, so it is not appropriate to examine them more closely here, since no useful generalisation can be made.[174]

iii. Results according to age

The last independent variable to be examined is that of age. Apart from indicating the difference between the linguistic output of younger bilinguals compared to older bilinguals, the age variable is a useful tool in studying diachronic language change. Whilst the data I discuss in this thesis are synchronic, collected over a roughly two year period, this synchronic data can be interpreted via an application of the apparent time principle (e.g. Labov 2001, Chambers 1995; Chambers, Trudgill and Schilling-Estes 2004) to give an indication of diachronic variation.

---

lived at the beginning of their life to reflect the region where that speaker was living during the crucial (i.e. early) stages of language acquisition.

[173] However, this does not mean that AD outside of B2S clauses may not be patterned differently according to region (or any other variable). I will suggest below that AD in clauses where the subject is a different pronoun may occur more frequently in the speech of speakers from south Wales than in the speech of speakers from north Wales.

[174] For reference, BRW is the speaker who lived in mid-Wales for the first year of his life, and showed 100% -A, whilst TRE is the speaker who lived in England for the first year of his life, and showed 86.36% -A.

According to this paradigm, a difference in the frequency of a construction in younger speakers compared to older speakers is assumed to signify that language change has occurred in the period between the points at which those two sets of speakers acquired their languages. If, for example, phenomenon P is found in 75% of constructions in the speech of a group of 20 year old bilinguals, and P is found in 50% of constructions in the speech of a group of 50 year old bilinguals, and that P is found in only 25% of constructions in the speech of a group of 80 year old bilinguals, then it can be inferred that P has become more frequent in the speech of bilinguals of these languages during the past 60 years.

To test apparent time I stratified the speakers into three distinct age groups: i) those under 30 years old at the time of recording, ii) those 30 to 50 years old inclusive, and iii) those over 50 years old. With this regard, the speaker set is reasonably balanced in terms of age: there are 9 speakers in the younger age bracket (under 30 years old), 10 in the middle age bracket (30 to 50 years old) and 9 in the older age bracket (over 50 years old).

In the data there were 9 speakers in the <30 age group, 10 in the 30-50 age group and 9 in the >50 age group, so they are quite balanced in number. Speakers in the youngest age group (<30) showed a mean 93.78% -A. Speakers in the middle (30-50) age group showed a mean 92.57% -A. The speakers in the oldest age group (>50) showed a mean 84.76% -A. The difference in -A between the younger group and the middle group is only 1.2%, which is negligible. However, the difference in -A between he between the youngest group and the oldest

group is 9.1%, which would appear to be significant.[175] The percentage of AD in the three groups is illustrated in Table 8 and Figure 13 below.

	under 30	30-50	over 50
+A	6.22%	7.43%	15.33%
-A	93.78%	92.57%	84.67%

Table 8. AD across the three age groups.

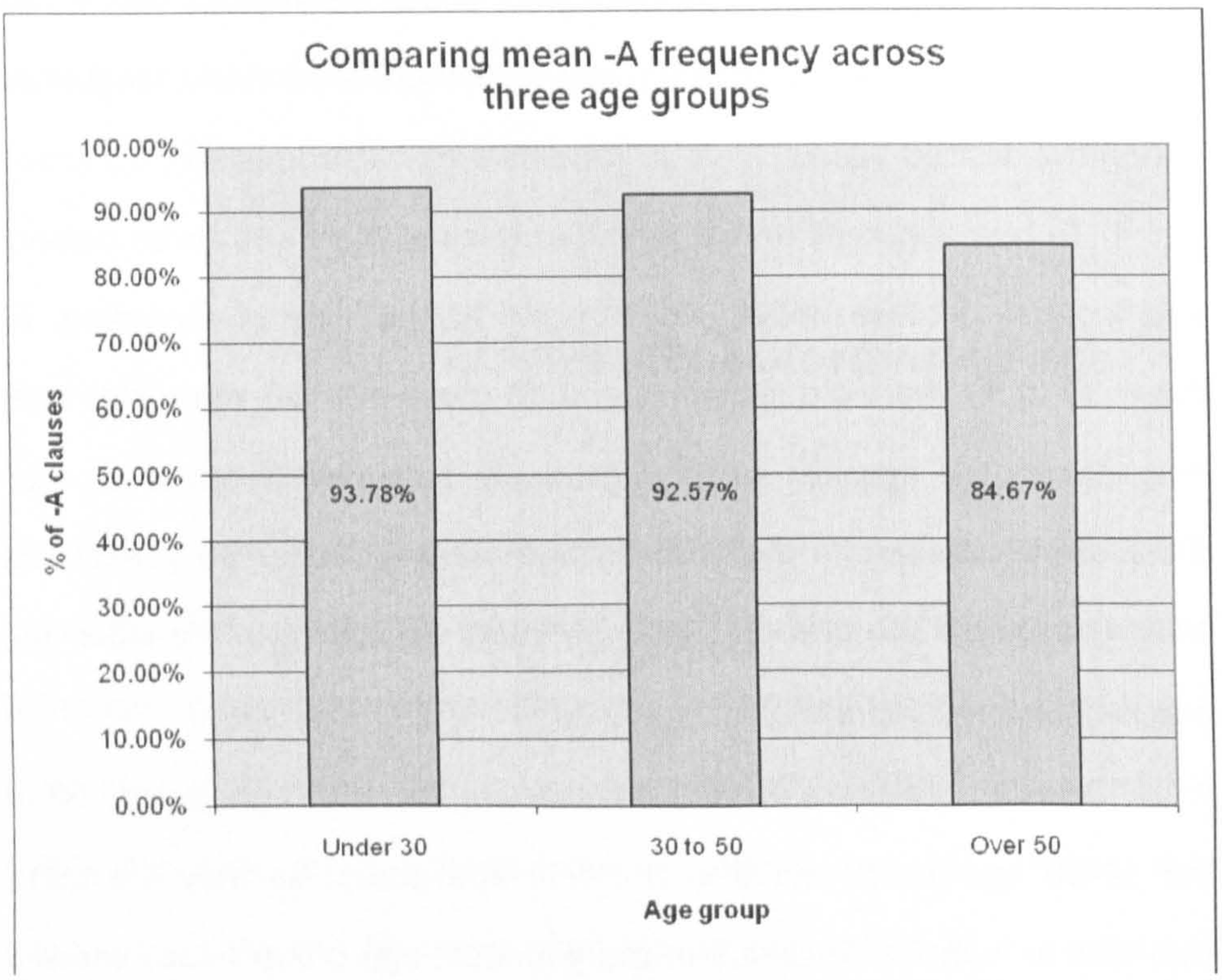


Figure 13. Percentage of clauses across the three age groups which were -A.

To determine whether or not these differences are statistically significant, I checked the correlation of AD with three age groups using a Chi-square test within the program SPSS.[176] The correlation was significant with $p<0.05$ ($p=0.0005$, where the significance level (alpha

[175] The difference in –A between the middle group and the oldest group is 7.9%.

[176] See e.g. Pallant (2005) for more details on the Chi-square test as a method of ascertaining statistical significance of the relationship between sets of data.

value) is set at 0.05). There is therefore a significant relationship between AD and age.

The above results indicate a progression across the age groups: whilst speakers in the oldest group (born before c.1954[177]) produce -A in 84.67% of B2S clauses, speakers who were born later (during or later than c.1955[178]) produce -A in 92.57% (30-50 year olds) and 93.78% (younger than 30) of B2S clauses, which is significantly more frequently.

The difference in AD activity between younger speakers compared to older speakers in the data indicates language change in the Welsh of these speakers according to the apparent time method. The change in question is an increase in the deletion of the auxiliary verb (-A) in B2S clauses over the past half century.

Despite this trend, it should be noted that the frequency of -A in the >50 age group is still considerable: over three quarters of their B2S clauses. The indication is that -A in B2S clauses was already an established feature of Welsh when the speakers in the oldest group acquired the language, more than 50 years ago. Indeed, the two oldest speakers analysed (GOR, 81, and EML, 79) have quite high -A rates (100% and 83.33% respectively), from which I infer that -A was present in Welsh as far back as the 1920s. If so, the increase in frequency of -A perceived in this analysis is the 'tail end' of a change which has possibly underway for a long period. Whilst descriptions by other authors of B2S clauses, as discussed at the beginning of this chapter, do not seem to indicate -A as being present (or, at least, not 'acceptable') in Welsh until quite recently (e.g. King 1996), recall how Jones and Thomas (1977) state that, some 30 years ago, deletion of *wyt* was acceptable in

---

[177] Assuming that these speakers were at least 51 years old when the recording took place, and that data collection began in 2005.

[178] Assuming that these speakers were 50 years old or younger when the recording took place, and that data collection began in 2005.

informal speech, and that Roberts (1988) also indicates that -A was found in some informal speech in the 1970s. My findings support these indications, and they suggest that -A in B2S clauses was common before the 1950s, even when Welsh grammar texts from that period do not show this. I will return below to discussing the concept of -A being a change in progress.

Before concluding this analysis, however, a brief discussion of the validity of the concept of apparent time is in order. The *a priori* assumption which must be made when applying the apparent time paradigm is that speakers do not change their speech patterns over the course of their lives from when they first acquire their language(s). This phenomenon is referred to as age grading (Chambers 1995, Boberg 2004:251), in which "generational differences represent the effect of aging rather than change in the language" (Boberg ibid.), and features which appear to be new become less frequent as the speakers who innovated them grow older. The only surefire method by which to verify that age grading has or has not taken place in a dataset is by using real-time data from the same (or similar) speech community from a previous point in time, from which it can be seen whether young speakers from the historical data retain their linguistic characteristics in the later data when they are in an older generation.

Age grading has been attested in some data (e.g. the Ontario English study cited in Chambers 1995:188–9), and is often found alongside other factors; e.g. Labov (1972:24) compares his Martha's Vineyard study from the 1960s with data from the *Linguistic Atlas of New England* and finds that while change seems to have occurred, age grading may also be present as a confounding factor. Labov (2001) distinguishes age grading, where the individual's language changes but the community's does not, from generational change, where the

individual's language does not change but the community's does (i.e. the assumption made when examining apparent time), but accepts that age grading was a factor in some real-time studies he examined (Labov 2001:76). In this regard, therefore age grading may be present in these Welsh-English data, in which case the lower frequency of –A in the speech of older generations may be because young innovators abandon (to an extent) the –A innovation as they grow older.

Boberg's (2004) study of change in Montreal English (analysing two datasets collected some years apart) generally corroborates the apparent time hypothesis, since synchronic age patterns match real-time change (e.g. the increase in frequency of /sk/ in *schedule* is maintained by the speakers who innovate the increase, and later generations continue the change; 2004:259). Boberg finds little evidence for age grading (according to the "classical" model) in his data (2004:265), and uses this to suggest that age grading "is not as serious an obstacle to apparent-time analyses of change in progress as has previously been assumed" (2004:266). Boberg's conclusion is that a combination of real-time and apparent time data is the optimal solution for analysing a language change process.

Practicalities restrict me from testing for the presence of age grading in the speech of these bilinguals, since substantial pre-existing corpora of Welsh informal speech are not available. There is therefore little to compare the present data to apart from itself, via apparent time (as outlined above). I use apparent time as an indicator of the directionality of any possible change that is occurring in these data (cf. Jones 1998:102, who adopts a similar position on her data analysis), and I am also able to make reference to what existing evidence of this aspect of Welsh in the literature (see section 4 above), but cannot disprove the existence of effects such as age grading. If another Welsh-

English corpus recorded at an earlier point in time were available, a real-time study could be conducted to bolster the indications of apparent time in my data, but a suitable other dataset is not available.

A recent suggestion for something of a compromise between apparent time and age grading is offered by Boberg (2004) in his "late adoption" model.[179] Boberg suggests that there are scenarios where older speakers adopt the linguistic trends innovated by the contemporary younger generation, rather than rejecting new variants, and thus their speech becomes more like the younger speakers'—thus the older speakers are key players in the spread of a linguistic change, accelerating a change in progress, resulting in the more rapid increase in use of an innovative linguistic feature than would be inferred from an apparent time analysis of real-time data (Boberg 2004:258). He calls this concept late adoption (2004:257), and his hypothesis is supported by most of his data: older speakers are frequently found to convert to using an innovative form, regardless of the form they learnt as children. This indicates that older speakers (i.e. a post-acquisition point in time) are not automatically excused from the process of language change, and do <u>not</u> become more conservative as they grow older as the age-grading hypothesis would have it. Rather, older speakers "drive changes to completion" (2004:266).

Given, then, the absence of real-time data for analysing AD in my data, and that therefore the possible presence of age grading cannot be demonstrated, the possible presence of late adoption can be postulated (although not proven without real-time data). Consider: while there is a significant difference between -A by speakers from the

---

[179] Boberg (2004:257) describes late adoption as "a third model of age-based linguistic variation ... really another kind of age grading, in that it involves change rather than stability in postacquisition grammars".

younger age group and the older speakers, all speakers produce -A clauses a large majority of the time—that is, speakers aged 30 and over in the data do not appear to be especially conservative in this regard. It could be the case that -A in B2S clauses is an example of late adoption. -A was innovated by younger speakers and then adopted post-acquisition by older generations, until it has become, by today, the norm in informal speech of speakers of all ages, albeit still more frequent in the speech of young Welsh speakers. If this is the situation, then older Welsh speakers may be driving the change to completion, by accelerating the spread of -A which was introduced by younger generations. I qualify the above suggestion, however, by noting that only by collecting a second dataset in some decades' time can the presence or absence of late adoption (or indeed age grading) be established.

In this section I have reported how I tested independent variables on the results of the analysis of AD in B2S clauses. The only noteworthy independent variable turns out to be age: younger speakers produce utterances with -A significantly more frequently than older speakers. I take this as an indicator that use of -A in Welsh has become more frequent in recent years. In the next section I examine the linguistic constraints of AD, and whether or not the type of linguistic context in which the construction occurs has bearing on the deletion or retention of the auxiliary verb.

e. Linguistic variables

Next I consider the linguistic variables which may also affect the application of AD. The variables I discuss are (1) the type of

construction, declarative or interrogative (and differences between different types of interrogative construction), (2) the linguality of the clause, whether monolingual or bilingual, (3) the presence or not of negation in the clause, and (4) the deletion or retention of the pre-verbal particles *yn* or *wedi*.

i. Type of construction

The clauses in the dataset were categorised into one of two types of construction: declarative or interrogative. A Welsh B2S interrogative construction[180] was identified as an interrogative if either a) it had a pre-verbal interrogative WH morpheme such as *beth*, *pam* etc., or b) the utterance had rising intonation at the end. Of the 648 B2S clauses analysed, 422 (65.12%) contain declarative constructions. The remaining 226 clauses (34.88%) contain interrogative constructions. Declarative constructions are the majority.

Of the declarative clauses, 390 clauses (92.42%) have -A whilst 32 (7.58%) retain the auxiliary (+A). -A the clear norm in declarative clauses produced by these speakers.

Of the 226 clauses with an interrogative construction, 211 (93.36%) have -A, whilst 15 (6.64%) have +A. There is only a very small difference between –A in declarative and interrogative clauses (0.95%), shown to be statistically insignificant according to a Chi-square test where $p>0.05$.

The comparison between AD in declarative and interrogative constructions is illustrated in Table 9 and Figure 14 below.

---

[180] By 'interrogative' I simply mean 'question'; I do not imply e.g. constituent movement.

	Declarative	%	Interrogative	%
+A	32	7.58	15	6.64
-A	390	92.42	211	93.36
Total	422	100.00	226	100.00

Table 9. AD distribution in declarative vs. interrogative constructions.

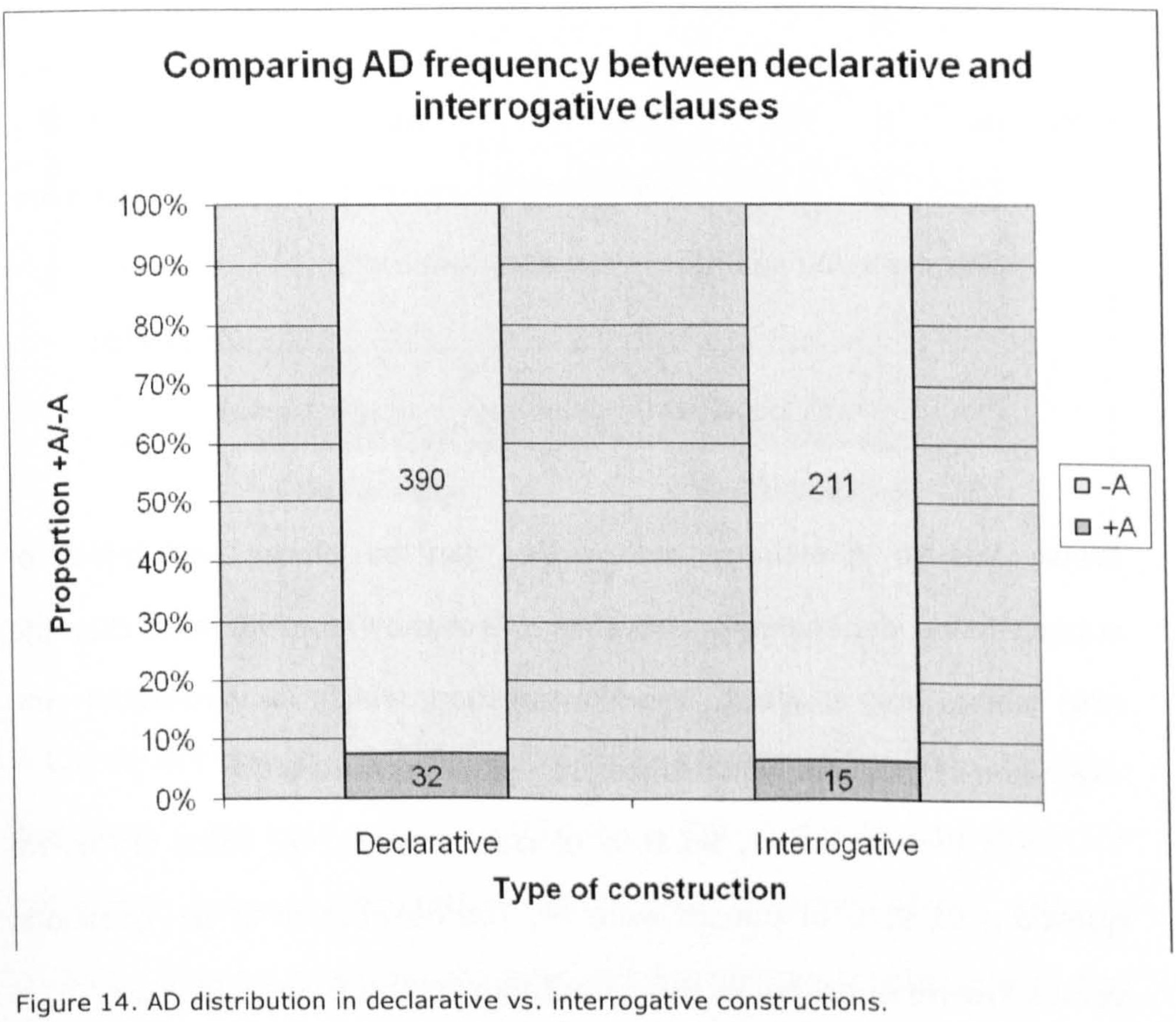


Figure 14. AD distribution in declarative vs. interrogative constructions.

The indication is that the type of construction has no effect on the speaker's likelihood to delete the auxiliary verb: a declarative B2S clause is as likely to have -A as an interrogative B2S clause.

I now examine AD in different types of interrogative construction. Welsh, like English, can form interrogatives as either WH-questions or as Yes-No questions. WH questions, as noted above, are question word-initial, whether or not they are +A as in (47) below or -A as in (48).

(47) be wyt ti wneud **weekend** 'ma

what be.2S.PRES 2S do.NONFIN weekend here

Daniel?

Daniel

"What are you doing this weekend, Daniel?"

[DAV6-HEC2268]

(48) be ti 'n galw mwy nag un enfys?

what 2S PRT call.NONFIN more than one rainbow

"What do you call more than one rainbow?"

[STA5-RHO2075]

[+A version: be **wyt** ti'n galw mwy nag un enfys?]

Welsh Yes-No questions, meanwhile, can be either verb-initial or subject-initial depending on whether or not they have AD. See example (39) above for a Welsh Yes-No question which is verb-initial and example (41) for one which is subject-initial because of -A.

In WH questions, 96.10% of clauses were -A, while in Yes-No questions 91.95% of clauses were -A. The distribution of WH questions versus Yes-No questions is shown in Table 10 and Figure 15 below.

	WH question	%	Yes-No question	%
+A	3	3.90	12	8.05
-A	74	96.10	137	91.95
Total	77	100.00	149	100.00

Table 10. AD distribution in WH questions vs. Yes-No questions.

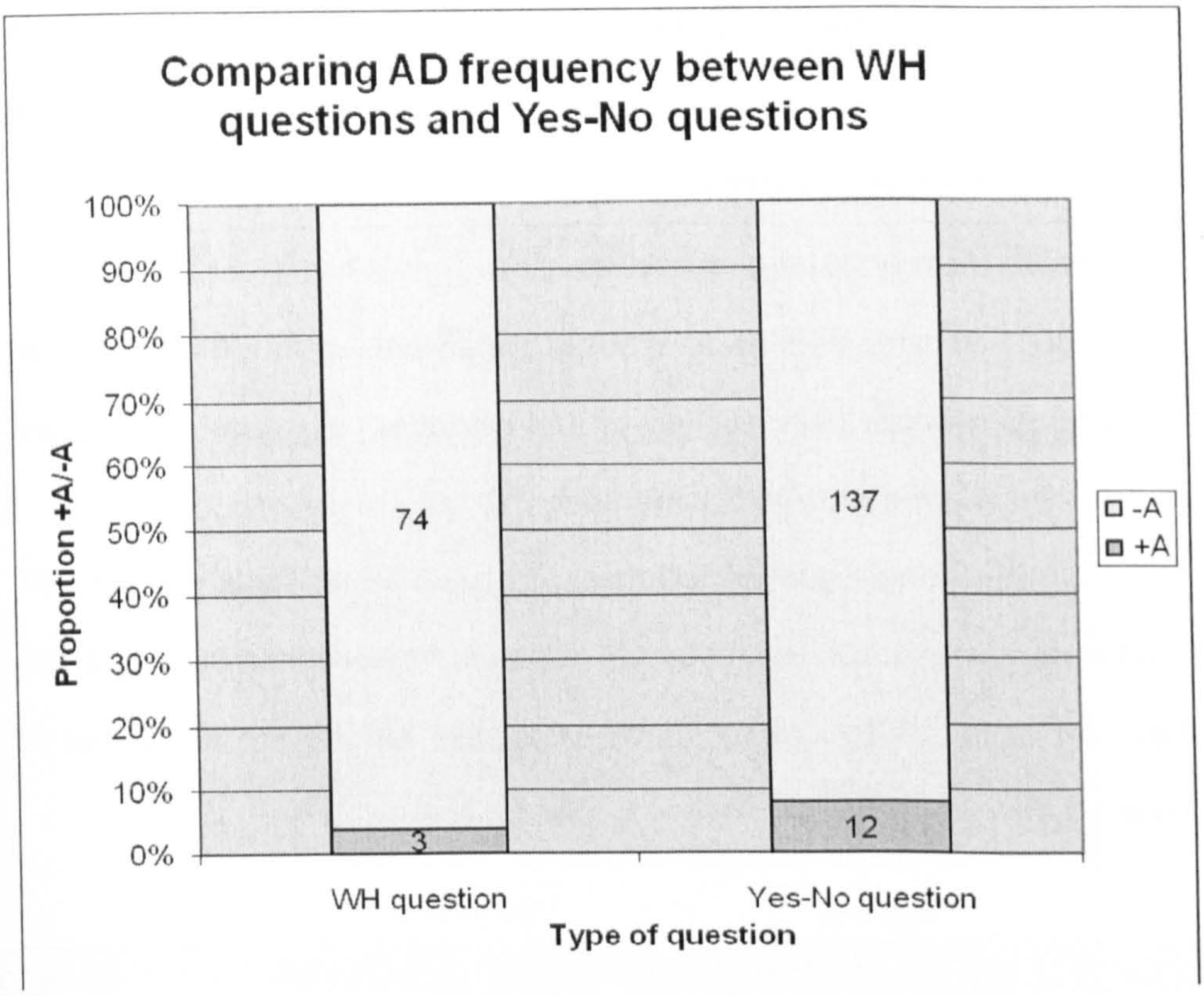


Figure 15. AD distribution in WH questions vs. Yes-No questions.

A Chi-square test to check the correlation between declaratives and WH-questions gives a value of $p>0.05$, and a Chi-square test to check the correlation between declaratives and Yes-No questions also a value of $p>0.05$. This indicates that the type of interrogative construction is also unrelated to -A frequency.

ii. Linguality of clause

In chapter 2 section 7, the connection between convergence and language mode was discussed. To briefly recap, it has been argued (e.g. Toribio 2004) that a speaker's bilingual mode increases the likelihood of convergence occurring, and it follows that a bilingual clauses involving

code-switching are more likely to show convergence than monolingual clauses. In light of this, I will now examine the connection between AD and clause linguality, i.e. whether or not a B2S clause is monolingual (containing morphemes from only one language) or bilingual (contains morphemes from both languages).

Out of the 648 B2S clauses analysed, 546 (84.26%) are monolingual, whilst 102 (15.74%), are bilingual. Bilingual clauses are therefore in the minority in this dataset.[181]

Of the monolingual B2S clauses, 511 (93.59%) show -A, while 35 (6.41%) show +A. Of the bilingual clauses, meanwhile, 90 (88.24%) show -A, while 13 (11.76%) show +A. These results are illustrated in Table 11 and Figure 16 below.

	Monolingual	%	Bilingual	%
+A	35	6.41	12	11.76
-A	511	93.59	90	88.24
Total	546	100.00	106	100.00

Table 11. AD distribution in monolingual vs. bilingual clauses.

[181] Bilingual clauses are generally in the minority in the data, it seems; cf. the data analysed in chapter 5, where 18.5% of clauses were bilingual.

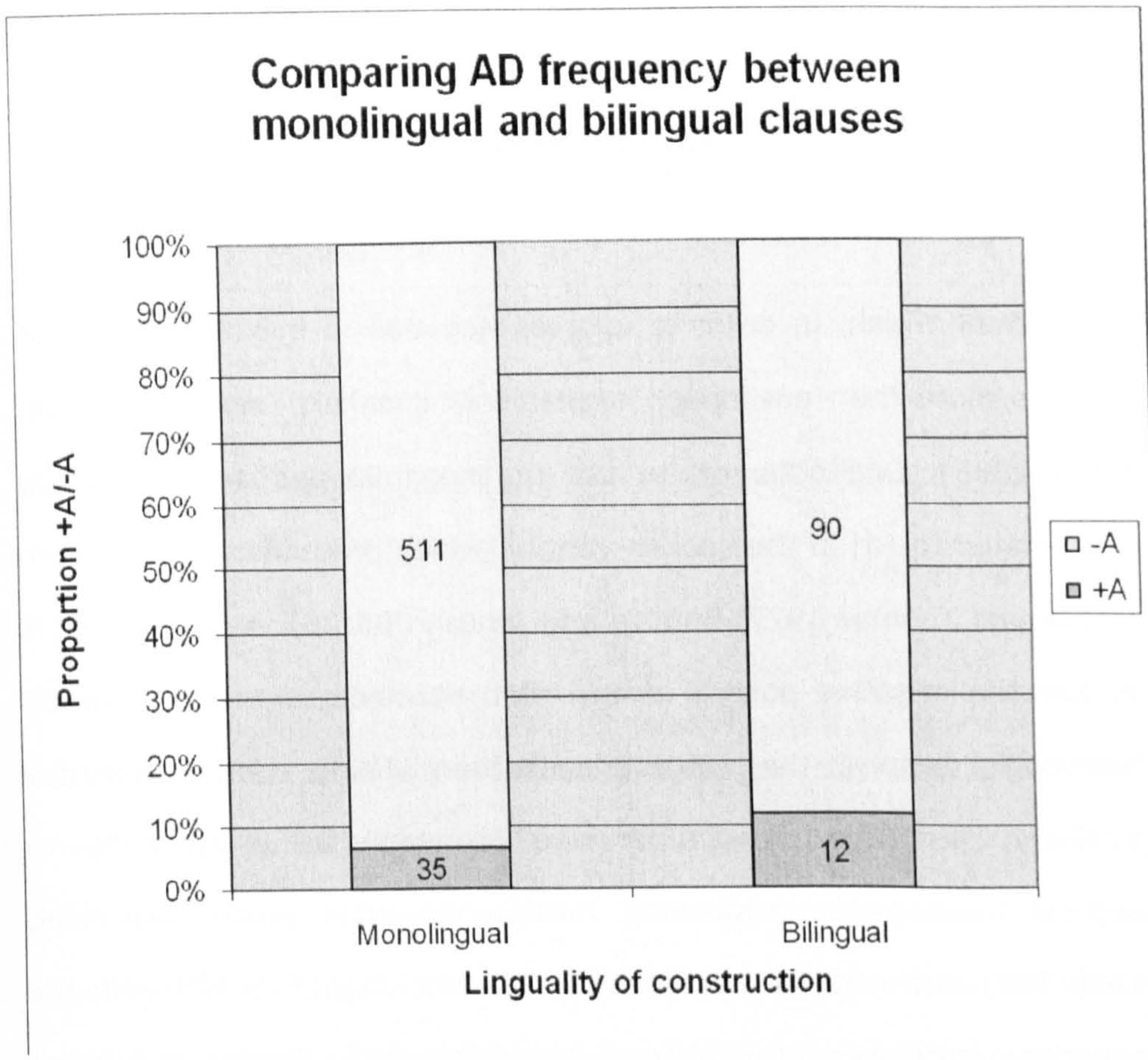


Figure 16. The proportion of +A versus –A clauses in monolingual versus bilingual clauses in the dataset as a whole.

-A is more frequent in monolingual clauses than in bilingual clauses (93.59% compared to 88.24%), but the difference just fails to reach significance ($p=0.056$, $p>0.05$). I infer from this that there is no relationship between the linguality of a B2S and the likelihood of -A occurring in that clause. A clause having CS, and thereby being a bilingual clause, does not apparently encourage –A.[182]

[182] In fact, as shown above, -A is slightly more common in monolingual clauses than in bilingual clauses.

iii. Role of negation

In informal Welsh, in order to express negation in a B2S construction there is more than one option available to speakers (see Borsley and Jones 2005). One option is to use the negative particle *nid*, usually abbreviated to [d] in the spoken variety, which precedes the finite verb (Jones and Thomas 1977, Borsley and Jones 2005:26). An alternative is to use the negative particle *ddim,* often abbreviated to [m], directly following (usually) the subject constituent (King 1996). Negative auxiliary constructions can be formed acceptably using either one or both of these negation systems. Further, negative subordinate (King 1996:307) and negative conditional clauses (King 1996:178) use the negative particle *na* or *nad* (e.g. King 1996:307, Borsley and Jones 2005:30). The invented examples below illustrate the difference between the various types: (49) uses *nad*, (50) uses *ddim*, and (51) uses both *nid* and *ddim*. (52), which lacks *ddim* and only has an abbreviated form of *nid*, is considered ungrammatical (Borsley and Jones 2005:28). See (38) through (40) and (45) to (46) above for examples of different negative constructions from the data.

(49) Os nad wyt ti 'n gwybod
if NEG be.2S.PRES 2S PRT know.NONFIN
"If you don't know."

(50) Wyt ti ddim yn gwybod
be.2S.PRES 2S NEG PRT know.NONFIN
"You don't know."

(51) Dwyt ti ddim yn gwybod
NEG-be.2S.PRES 2S NEG PRT know.NONFIN
"You don't know."

(52) *Dwyt ti 'n gwybod
NEG-be.2S.PRES 2S PRT know.NONFIN
"You don't know."

There is also a negative morpheme used only in southern varieties of Welsh, which has a variety of forms, including *so*, *sa*, smo, sdimo, etc. (King 1996:146; they are what Borsley and Jones [2005] call "s-forms"; Borsley [2005:3] calls *so/smo* an "extra-strong negative verb", a negative form of *bod* 'to be'). This morpheme occurs in the same position as *nid*, though it cannot co-occur with *ddim* (Borsley and Jones 2005:63). A hypothetical example of the use of *so* is given in (53).

(53) So ti 'n gwybod
NEG 2S PRT know.NONFIN
"You don't know."

In this section I examine the connection between AD and negation. 81 B2S clauses in the dataset feature a negative construction, which is 12.5% of the total B2S clauses. Of these, 79 clauses (97.53%) have -A, whilst 2 (2.47%) have +A. The negative constructions with -A represent 12.19% of the total B2S clauses in the data, whilst negative constructions with +A represent only 0.31% of the total data.

Affirmative B2S clauses, meanwhile, constitute the majority of the data (567 clauses or 87.5%). 522 (92.06%) of these clauses have -

A, whereas 45 (7.94%) have +A. These results are shown in Table 12 and Figure 17 below.

	Affirmative	%	Negative	%
+A	45	7.94	2	2.47
-A	522	92.06	79	97.53
Total	567	100.00	81	100.00

Table 12. AD distribution in affirmative vs. negative constructions.

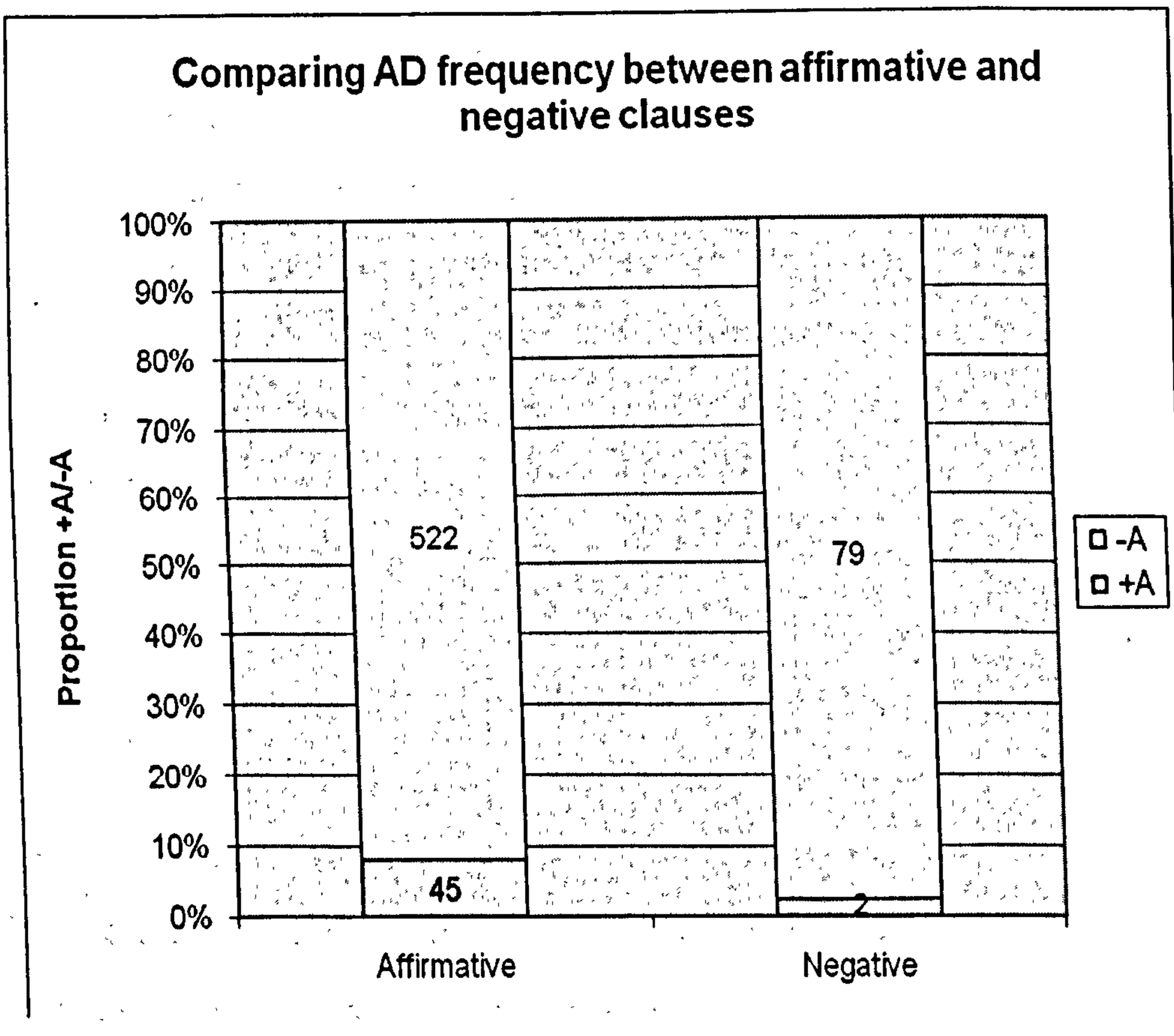


Figure 17. The proportion of +A versus -A clauses in affirmative versus negative clauses in the dataset as a whole.

-A occurs slightly more frequently in negative B2S clauses (97.53%) than in affirmative B2S clauses (92.06%), but the difference is not statistically significant given that p=0.076, so p>0.05.

I noted above that –A negative declarative clauses are rare in the data (only two occurrences) and that –A negative interrogative clauses are not found at all. The combined distribution of AD in declarative/interrogative and affirmative/negative clauses is illustrated below in Table 13, where +D means declarative, -D means interrogative, +N means negative and –N means affirmative.

	+D -N	-D -N	+D +N	-D +N	Total #
+A	30	15	2	0	47
-A	329	193	61	18	601
Total #	359	208	63	18	

Table 13. AD distribution according to both declarative/interrogative constructions and affirmative/negative constructions.

It appears that retention of the auxiliary (+A) is especially inhibited in negative interrogative constructions in the speech of these speakers, though this gap might be explained by there not being many –D, +N clauses in these data. Generally, I interpret the results in Table 9 to represent the general lack of -A in negative clauses in this dataset.

These results clearly show that, in B2S negative constructions, -A is the norm, and that B2S negative constructions where the auxiliary is not deleted are very rare (with only 2 occurrences in these data).

Example (54) below is a -A negative declarative clause from the dataset.

(54)	tɨ	(ddi)m	yn	cofio	hynna
	2S	NEG	PRT	remember.NONFIN	that
	"You don't remember that."				[FUS14-AWE735]

This clause forms negation using [m], an abbreviated form of *dim*.

There are two -A negative clauses in the data which use the *na*(*d*) particle, like (55) below. Here the final [d] of *nad* has either been elided or assimilated with the onset [t] of *ti*.

(55) os na(d) ti 'di roid o fewn
if NEG 2S PRT.PAST put.NONFIN 3Sm in
ffor' wrong yli
way wrong look.2S
"Unless you've put it in the wrong way, see."

[DAV6-DAN379]

The negation system using *sa/so* is found only in two clauses in the dataset. An example is given in (56) below, which is direct speech where the speaker is quoting another person.

(56) "achos so ti gallu cael blood
because NEG 2S can.NONFIN have.NONFIN blood
test tan dydd Llun beth bynnag"
test until day Monday whatever
"Because you can't have a blood test until Monday anyway."

[DAV1-SAR3205]

If the negation systems used in B2S negative clauses showing AD are categorised, 74 out of 81 (93.67%) -A clauses use *dim* on its own.[183]

-A clauses using other negation systems are much less apparent in the dataset. *Sa/so* occurs in only two clauses, both by the same

[183] *Ddim* is usually abbreviated to [m] by the speakers in these data.

speaker, who was born in south Wales. *Na* or *nad* occurs in three clauses, twice on its own and once, possibly, co-occurring with *dim*; this clause is given in (57) below.

(57)	na	ti	(dd)im	yn	gwybod	lle
	NEG	2S	NEG	PRT	know.NONFIN	place
	wnaeth	y	deg	punt	yna	ddod
	do.3S.PAST	DET	ten	pound	there	come.NONFIN
	o	off	y	*top*?		
	from	off	DET	top		

"You don't know where that ten pounds came from off the top?"

[STA5-SND2778]

It is difficult to be certain whether or not the *na* that precedes the subject in this clause is the negative particle *na* or the negative morpheme *na* 'no' (i.e. "no, you don't know where..."). This morpheme is not stressed in the audio recording, so it is possible that it is indeed the negative particle. However, given there is only one occurrence in this dataset of such a construction, where both negation systems present and it is -A, I suggest that this is an infrequent phenomenon, and that using only one system is most prevalent. These results show that *dim* is by far the most common negation system found, supporting the notion that pre-subject negative particles are not common in AD clauses, potentially because they are considered 'part' of the auxiliary verb and are deleted along with it.

The two +A negative clauses in the data both use *nad* to express negation.[184] These clauses are shown in (58) and (59) below. Note that

[184] Note that there are clauses in the data which use *byth* and *erioed* 'never' as a form of negation. *Byth*, along with *erioed,* is a negative adverb which is used in imperfective

both clauses were produced by the same speaker, Gwynfor, who was born in north Wales.

(58)	os	nad	wyt	ti	'n	*glir*	o	dy [//]
	if	NEG	be.2S.PRES	2S	PRT	clear	of	POSS.2S
	beth	ydy	dy		<u>brief</u>	di	dy	
	what	be.3S.PRES	POSS.2S		brief	2S	POSS.2S	
	hun							
	self							

"If you're not clear of your...what your own brief is."

[FUS15-GFR1705]

(*byth*) or perfective contexts (*erioed*) as well as in "important non-negative uses" (Borsley et al 2007:263). Unlike *ddim*, *byth* (and *erioed*) can follow a negative subject (2007:269), e.g.

Does	neb	byth	yn	y	stafell
be.3S.IMP.NEG	nobody	never	in	DET	room

"There's never anybody in the room." (taken from Borsley et al 2007:269)

[cf.

*Does	neb	ddim	yn	y	stafell
be.3S.IMP.NEG	nobody	NEG	in	DET	room

"There's nobody in the room."]

Because of this difference between the syntax of *ddim* and *byth/erioed* I have decided not to include clauses involving *byth* or *erioed* as instances of negation in these data.

(59) mae 'n debyg nad wyt ti *just*
be.3S.PRES PRT likely NEG be.2S.PRES 2S just
yn mynd i allu xx identify-io
PRT go.NONFIN to can.NONFIN xx identify-VBZ
xxx...
xxx
"It's likely that you're not going to be able to [..] identify [...]..."
[FUS15-GFR1934]

Only 4 occurrences of *nad* are found in the dataset; 50% of these occurrences are +A and 50% are -A. As noted, both the –A occurrences are by the same speaker. I suggest that this indicates that *nad* is used infrequently in general (and is in any case used in fewer contexts than e.g. *nid*), rather than indicating it is necessarily more resistant to AD than other negation forms.

To sum up the issue of the connection between negation and AD, the results reveal that while -A occurs more frequently in the negative constructions of the speakers analysed than in the positive constructions (97.53% versus 92.06%), this relationship is not statistically significant. With regards to the different negation systems found in B2S clauses, the use of d*im* is by far the most frequent, with 93.67% of negative constructions which delete the auxiliary using this negation morpheme.

iv. Particle deletion

As noted in section 3 above, in Welsh an aspectual particle is expected to precede the nonfinite verbal noun in periphrastic auxiliary constructions, including B2S clauses (King 1996:166, 292). This is either

the progressive particle *yn* (abbreviated to *'n* [n] when following a vowel) or the perfective particle *wedi* (often abbreviated to *'di* [di] in informal spoken Welsh); see Borsley et al (2007:12) for more on these terms. The dataset shows that these particles occur in +A clauses and -A clauses. Examples are given below: (60) has +A with (*y*)*n*; (61) has +A with (*we*)*di*; (62) has -A with (*y*)*n*; (63) has -A with (*we*)*di*. The particle is in bold in each instance.

(60)	wyt	ti	**'n**	cofio	fo,	y
	be.2S.PRES	2S	PRT	remember.NONFIN	3Sm	DET
	dyn	*lollypop*?				
	man	lollypop				
	"Do you remember him, the lollypop man?"					[DAV10-CLE2639]

(61)	wyt	ti	**'di**	gweld	e?
	be.2S.PRES	2S	PRT.PAST	see.NONFIN	3Sm
	"Have you seen him?"				[ROB6-MOR2073]

(62)	achos	ti	**'n**	tynnu	fe	ar draws
	because	2S	PRT	pull.NONFIN	3Sm	across
	"Because you pull it across."					[ROB6-MOR166]

(63)	ti	**'di**	bod	mewn	*gig*	o'r blaen?
	2S	PRT.PAST	be.NONFIN	in	gig	before
	"Have you been in a gig before?"					[FUS31-ARF150]

Both particles are also sometimes found deleted before the verbal noun, and, again, both in +A clauses, as in (64), and in -A clauses, as in (65).[185]

(64) a wedyn wyt ti mynd i
and after be.2S.PRES 2S go.NONFIN to
chwith
left
"And then you go to the left." [DAV11-OWA1479]

[Form with particle: a wedyn wyt ti'n mynd i chwith.]

(65) be ti feddwl?
what 2S think.NONFIN
"What do you mean?" [FUS31-BRW1349]

[Form with particle: be (wyt) ti'n feddwl?][186]

Note that there are some verbal nouns, such as *eisiau/isio* 'want' and *angen* 'need' with which an aspectual particle is not expected, as shown in (66). Such clauses are excluded from this analysis.

(66) *so* be ti isio neud amdano
so what 2S want.NONFIN do.NONFIN about.3Sm
fo then?
3Sm then
"So what do you want to do about it, then?"

[FUS31-ARF107]

---

[185] I rely on the transcribers to have accurately transcribed whether or not they heard a particle in each instance.

[186] Note that, as the particle is deleted in clauses such as (47) and (48), it is often difficult to tell whether it is *yn* or *wedi* that has been deleted.

Since there is both particle deletion and retention in the data, I wanted to see if this correlated in any way with the fact that auxiliary deletion and retention is also present in the data. It might be expected, for example, that the particle in examples (61) and (62) is retained to express the clause finiteness lost by the deleted auxiliary. In this section I analyse whether or not particle deletion (PD) is connected to auxiliary deletion. I use the following abbreviations: -P indicates a deleted particle, whereas +P Indicates a retained, undeleted, particle.

There are logically four groups into which B2S clauses may fall: clauses with +A and +P (a retained auxiliary and a retained particle), clauses with +A but -P (a retained auxiliary but a deleted particle), clauses with -A but +P (a deleted auxiliary but a retained particle), and clauses with both -A and -P (a deleted auxiliary and a deleted particle).

Out of the 648 B2S clauses analysed, 581 (89.66%) have a construction in which an aspectual particle would usually be expected in prescriptive grammar. -P is found in 263 clauses (45.27% of those where a particle might be expected to occur), whilst +P is found in 318 clauses (54.73%). Thus +P is the most frequently-found phenomenon in these data, although -P is also common. Table 14 below shows the results of PD with AD, also illustrated in Figure 18.

	+P	% of +P	% of all	-P	% of -P	% of all
+A	6	1.89	1.03	35	13.31	6.02
-A	312	98.11	53.70	228	86.69	39.24
Total	318	100.00	54.73	263	100.00	45.27

Table 14. Frequency of particle deletion in relation to auxiliary deletion.

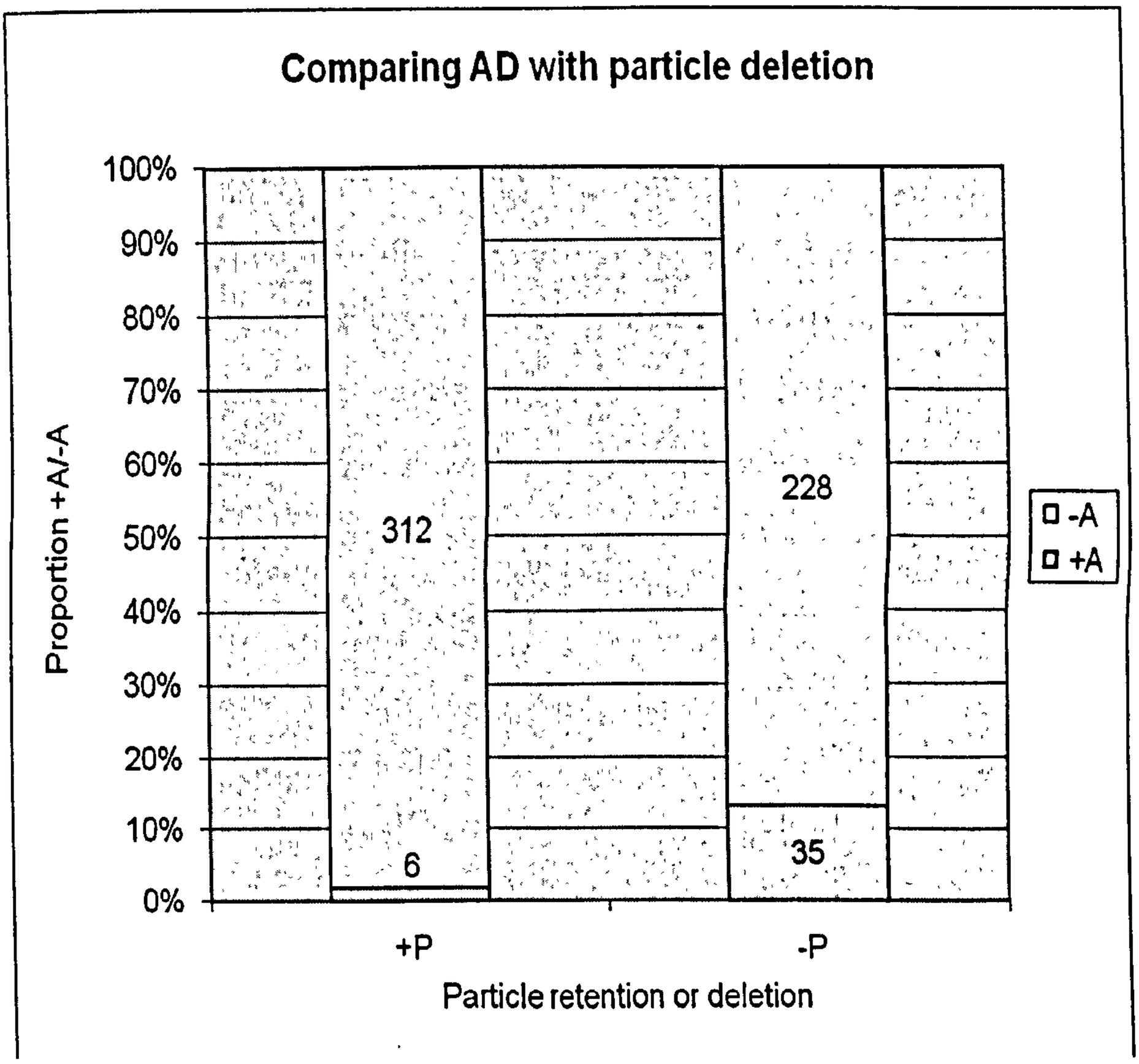


Figure 18. Frequency of particle deletion in relation to auxiliary deletion.

As Table 10 shows, the most frequent pattern in these data is -A, +P (312 such clauses; 53.70% of clauses where a particle would be expected). The next most frequent type is -A, -P (228 or 39.24%). The third most frequent type is +A, -P (35 or 6.02%). The least frequent type is where neither auxiliary nor particle is deleted: +A, +P (6 or 1.03%). In these data, -P is more common where there is +A (35 -P to 6 +P), but, conversely, that +P is more common where there is -A (228 -P to 312 +P), although a large proportion of the clauses analysed had both auxiliary and particle deleted.[187] There is a negative correlation

[187] The limited occurrence of clauses where neither auxiliary nor particle is deleted is, I suggest, indicative of the high frequency of -A in general by these speakers.

between AD and PD, and this is statistically significant at the level $p<0.01$ ($p=0.0005$).

To clarify, in -A clauses +P is found in the majority (57.78% of -A clauses), whilst in +A clauses -P is found in the majority (85.37% of +A clauses). Clauses in these data where the auxiliary verb is retained generally delete the pre-verbal particle, whilst clauses where the auxiliary verb is deleted generally retain the pre-verbal particle.

One explanation for this negative relationship is that deletion of both the auxiliary and the pre-verbal particle may remove too much overt expression of meaning from the clause. The deletion of *wyt* (-A) removes the overt present tense marker from the clause, but aspect is still marked if the aspectual particle is overt. If, however, that particle is deleted as well as the auxiliary, then all tense and aspect marking in the clause is removed. This may be why -P is not the norm in these data and particle retention (+P) is slightly more common—these speakers choose to delete morphemes from the clause only insofar as it does not eliminate too much meaning.

However, clauses with -P do occur quite frequently in this dataset, so speakers still produce them without (presumably) the other speaker having difficulty interpreting the clause meaning.[188] This may be because present tense can be considered to be the default tense in any clause, and so if there is no other evidence to the contrary provided by the morphemes in a clause, presumably speakers interpret clauses with, for example, no overt inflected verb or overt aspectual particle to be present tense, though it is not clear how they interpret the aspect expressed by a deleted particle. It is possible that the progressive

---

[188] Presumably clauses which have both -A and -P are assumed by the speakers to convey present tense and progressive aspect (i.e. *wyt* and *yn* are deleted), but this cannot be proven.

particle *yn* is assumed to be the default, since it occurs more frequently than the perfective particle *wedi* in these data. If both present tense and the progressive aspectual particle are indeed the default 'choices' for B2S clauses, this might explain why clauses where the auxiliary is retained (+A) have a deleted particle (-P) more frequently than a retained particle (+P)—that is, speakers are able to interpret that a deleted particle has progressive aspect because they associate progressiveness with the present tense of a clause. Indeed, all +A, -P clauses found in these data are non-perfective, which would support this argument. The above, however, remains conjecture, and further exploration would be required to explain speakers' rationale behind aspectual particle deletion/retention.

f. What the analysis shows

In this section I have shown and discussed the results of auxiliary deletion in these Welsh data. -A is overwhelmingly the most frequent form, occurring in 92.75% of B2S clauses.

Testing of independent variables indicates that while speakers' gender and the region where they grew up have little impact on the likelihood that they will produce -A clauses, the age of the speaker is significant, in that younger speakers produce -A more frequently than older speakers. I interpret this to indicate that -A has become more prominent in Welsh over the past 50 years or so, and that it can be seen as a change in progress.

Testing of linguistic variables indicates that the type of construction found, declarative or interrogative, the linguality of the clause and the presence or not of negation in the clause are all

insignificant in terms of AD. An analysis of the deletion or retention of the pre-verbal particle *yn* or *wedi* showed an interesting inverse correlation, in that -A clauses are more likely to have a retained particle but +A clauses are more likely to have a deleted particle, possibly because speakers elect not to delete excessive tensual/aspectual information from a clause, as it renders opaque the meaning expressed by the sentence.

In the next section I will proceed to a discussion of why AD occurs, arguing that it is due in part to the influence of English word order.

8. Discussion: why does AD occur?

In this chapter so far, I have discussed the phenomenon of AD, and given examples of it from a Welsh-English dataset, showing that those speakers produce -A in 92.75% of B2S clauses overall. I have also shown how younger speakers produce -A more frequently than speakers from older generations, which implies that -A is increasingly becoming the norm in Welsh B2S constructions. In this section I will suggest an explanation of why AD occurs in Welsh.

I propose that there are two factors which contribute to AD: (1) phonological weakening of the auxiliary verb to [Ø], and (2) influence of English subject-initial word order. I will discuss both in turn.

As suggested by Phillips (2007), an initial stage in the AD process is probably phonological weakening of the auxiliary verb (*wyt*) due to shift of primary stress from this verb to later in the structure. Phonological reduction of the vowel of *wyt* > [Ø] could then occur, with the coda [t] assimilating with the onset [t] of *ti* ([tɪ]).

I suggest that a second factor which facilitates AD is the influence of English-predominant main clause SVO word order. I have already commented above that B2S clauses with AD frequently result in a subject-initial construction, e.g. (67) below.

(67) ti 'n gwybod
S V
2S PRT know.NONFIN
"You know."
[+A form: **wyt** ti'n gwybod]

Note how the subject pronoun *ti* is the first overt element of the clause due to the deletion of the initial auxiliary verb *wyt*. Compare this to the English equivalent in (68).

(68) you know
S V
2S know.2S.PRES

Whilst there are obvious differences between the clauses—e.g. the English clause has no aspectual particle; the English main verb is finite whereas the Welsh main verb is nonfinite (although both are finite clauses)—they both clearly have a subject pronoun as the initial overt element of the clause.

As discussed already in this thesis (chapter 3), Welsh and English have some similarity in word order. Whilst Welsh is primarily VS main clause order and English primarily SV, Welsh also has SV as an alternative word order, e.g. emphasis of the subject constituent by making it the first surface clause element. In chapter 2 I described the

process of convergence as the increase in the use of a structure in Language A due to the prevalence of that structure in Language B. In this instance, I suggest that the prevalence of English SV word order influences Welsh-English bilinguals to extend the frequency of SV-order clauses in Welsh, by applying -A to B2S clauses to make them S-initial. The reason this process affects B2S clauses in particular is because the auxiliary verb *wyt*, as suggested above, is susceptible to phonological erosion to [Ø].

Interestingly, in connection with the probably recent spread of AD, the 20th century, as I discussed in chapter 1, is a period notable for showing a marked increase in the number of Welsh speakers who can also speak English. Deuchar (2005b) and Jones (1993) discuss Wales census results from 1901 to 2001. The 1901 census shows that 69.8% of Welsh speakers living in Wales were bilingual in English and that the remaining Welsh speakers were monolinguals (Jones 1993:550), but the inference from the 2001 census is all Welsh speakers over the age of 3 can also speak English (cf. Deuchar 2005b:624). Bilingualism became the norm for Welsh speakers over the past century. It is tempting to attribute the rise in AD to this increase in bilingualism with English that 'triggered' the increase in AD.

Additional evidence that it is the influence of English that is largely behind AD in Welsh is that AD appears to be, as I suggested above, a relatively recent phenomenon. If the difference in AD activity between younger speakers in the data and older speakers in the data is significant, and that AD has increased over the past 50 years, then it may be assumed that AD has become increasingly present in Welsh in the years before that, from an indeterminate point in the past when there was no AD in Welsh. Given that there is apparently no record of AD in Welsh before the 20th century, I speculate that AD might have

begun to be a feature of Welsh some time towards the beginning of that century.

I have analysed AD in B2S clauses in this chapter because in my experience I observed that it occurred in the second person in the speech of many Welsh-English bilinguals, but I noted in section 4 above that AD is also sometimes found in auxiliary constructions where the auxiliary is also (presumably) a form of *bod* 'to be' but the subject is something other than *ti*. Whilst I do not quantify AD in these constructions in this study, it appears to be true that -A in clauses with e.g. a 1st person singular auxiliary verb occurs more frequently (if not exclusively) in the speech of bilinguals who speak a southern variety of Welsh than those who speak a northern variety. Consider the examples (69) and (70) below from the wider corpus. The clause in (69) has an overt auxiliary verb *dw* (+A), whereas the clause in (70) has no overt finite verb (+A), but the pronoun *fi* indicates it is a first person singular verb.

(69)	do,	dw	i	'n	gwybod
	yes	be.1S.PRES	1S	PRT	know.NONFIN
	"Yes, I know."				[DAV6-DAN323]

(70)	fi	*just*	yn	dibynnu	ar	y	style
	1S	just	PRT	depend.NONFIN	on	DET	style
	though	yn	y	diwedd			
	though	in	DET	end			

"I just depend on the style, though, in the end."

[FUS27-LIS263]

[+A form: **wyf** fi just yn dibynnu...]

There clause produced by Daniel, a northern Welsh speaker, is +A, but the clause produced by Lisa, a southern Welsh speaker, is -A. Why might there be a difference in this deletion between the output of a northerner and a southerner?

One obvious explanation refers back to my suggestion above about the phonological erosion of *wyt*. In northern varieties of Welsh, the 1st person singular present tense form of *bod* is *ydw* [ədu] or *dw* [du], while in southern varieties of Welsh the 1st person singular present form is often *wyf* or *yf*, comprising a vowel diphthong [ui] or monophthong [i] and a coda [v] (King 1996). I suggest that -A of the northern auxiliary is blocked by the presence of the onset consonant [d], whereas AD of the southern auxiliary is facilitated where it lacks a consonant: the coda [v] presumably assimilated with the onset [v] of *fi* [vi][189] and the auxiliary becomes Ø.

Similar parallels between a northern verb form with onset or medial [d] and a southern verb form without [d] are found in other parts of *bod*, e.g. northern *yd̲an* versus southern *ŷn* (1st person plural present), *yd̲ach* versus *ŷch* (2nd person plural present), *yd̲y* versus *yw* (3rd person singular present), *yd̲yn* versus *ŷn* (3rd person plural present) etc. This could explain why -A seems to occur outside B2S clauses in the speech of speakers of southern varieties of Welsh more frequently than in the speech of speakers of northern varieties.

Yet I noted in section 3 of this chapter that clauses are found where the deleted auxiliary verb is (presumably) *mae* and the subject is a lexical noun (example 20 above) or proper noun (example 21 above). It would not be expected for the onset [m] of *mae* to be phonologically eroded, yet here the verb is [Ø]. In such cases, I suggest that influence

---

[189] I assume with this that the independent pronoun *fi* is being used, rather than the form *i*. Indeed, *fi* is probably a reanalysed form of *i* + the final coda of *wyf*.

of English is a convincing explanation for the -A. If, then, AD in constructions such as these were to be found occurring frequently in the speech of Welsh-English bilinguals, then a general shift of VS > SV in clauses in which the (deleted) auxiliary is a form of *bod* might be predicted. That is, I tentatively propose that the influence of English in the form of AD could be found across the Welsh paradigm in all (colloquial) varieties, but that phonological factors inhibit which verbs are most affected. Nevertheless, in the present study I have not focused on phonology and so will wait for future research outcomes to continue with this line of discussion.

In this section I proposed that AD in Welsh is partially influenced by phonological weakening but also, crucially, convergence to English subject-initial word order due to increased bilingualism among speakers of Welsh. As supporting evidence for this, in the next section I cite research conducted into Central Salish, a language with similarities to Welsh which seems to be undergoing a similar change.

9. AD elsewhere: Central Salish of the Vancouver region of Canada

So far in this chapter I have discussed auxiliary verb deletion in Welsh B2S clauses, and, finding that this is common in the speech of Welsh-English bilinguals, I have suggested that this is a sign of a word order shift in certain Welsh constructions from V-initial to S-initial. In this section I examine an existing study of an Amerindian language, Central Salish, which, according to recent research, has undergone a word order shift from VS to SV in recent decades. I then discuss what similarities

this situation has with Welsh, and what this indicates about AD by Welsh speakers.

Central Salish (also called Central Coast Salish) is an Amerindian language from the Salish language group, whose daughter languages are currently spoken in British Columbia in Canada and areas in north-west USA (Lewis 2009). Shaw et al (2008) give the following synopsis of the history of Central Salish contact with English. Central Salish has a number of dialects, and is spoken in the Vancouver area of Canada by a small group of speakers who are now bilinguals in Salish and English. The first contact of the native speakers with white European traders occurred in the late 18th century, and English-speaking people settled in the area from the 1820s onwards, with the founding of Fort Langley as a trading post in 1827, New Westminster as the capital of British Columbia in 1859, and Vancouver in 1886. English became a significant language in the region during the 19th century. From the late 19th onwards and into the 20th century, a programme of residential English-language schools forced Central Salish speakers away from their families, and government policy prohibited the use of Amerindian languages. These factors combined to result in a breakdown in the cross-generational transmission of Central Salish. Thus almost all speakers of Central Salish dialects are currently over 60 years old, even though there are recent language revival programmes which attempt to reintroduce Salish into the younger generations (Shaw et al 2008).

According to Shaw et al, there are a number of Central Salish dialects in the Vancouver region: Hunquminum or Hə'nqəmi'nəm is the downriver dialect, Hulquminum is the island dialect, Halqemeylem or Halkomelem is the upriver dialect.

Central Salish is a strictly V-initial language, where the subject is invariably found in the second position. However, research by Shaw et al

(2008)[190] using data collected by Galloway (during 1970-80) from Central Salish speakers of the downriver Hə'nqəmi'nəm (henceforth HN) and the upriver Halkomelem (henceforth HL) dialects shows that whilst features of HN are conservative in terms of Salish grammar, HL shows innovation in phonology, morphology and word order. In particular, certain clauses in HL have surface word order that is S-initial. Here I repeat the main findings of Shaw et al (2008) on this topic, and compare it to the situation with Welsh B2S clauses.

One morpheme which shows phonetic innovation in HL is the auxiliary verb [ʔə] or [ʔɪ]. This morpheme usually is found clause-initially, preceding the subject, giving VS word order, as seen in (71) below.

(71) ʔí·-ɬ cəl t'wə ʔí·təṯ
AUX S
Aux[1]-Past 1sSu Spec be.asleep[191]
"I must have been asleep." (taken from Shaw et al 2008:9)

In an earlier study, Galloway (1977) noted that, whilst [Aux S] was a prevalent word order, some HL speakers nevertheless produced clauses where that clause-initial auxiliary was deleted, thereby making the subject the first overt element of the clause, as seen in (72).[192] Shaw et

---

[190] The research by Shaw et al (2008) was conducted under the auspices of the collaborative MIB-UBC First Nation Language program. Thanks to the late Adeline Point, Edna Grant and Arnold Guerin Sr. and to the Musqueam people for their assistance.

[191] This gloss and others below are as given in Shaw et al (2008).

[192] Note that this phonetic innovation is not limited to auxiliary particles. A similar process is identified by Shaw et al in the erosion of the Yes-No question marker, which in downriver HN is [ʔə], but in HL is found as [-ə] or even [Ø] (i.e. deleted), as shown in (a) below.

al suggest that the phonological weakness of this morpheme makes the auxiliary verb susceptible to phonological erosion > [Ø].

(72) cəl lɛm

S V

2sSu do.what

"I went / I go." (taken from Shaw et al 2008:10)

[Form with Aux:

ʔí cəl lɛm

AUX S V]

All the subject-initial clauses which Galloway identifies in HL all have a pronominal subject, as in (72), and not a lexical subject. More recent research (e.g. Wiltschko 2000, 2002), however, shows that not only are S-initial clauses in HL more frequent than when Galloway's research was

---

(a) HN: nɛm ʔə cxʷ ʔ

Aux Q 2s.Su

HL: lɛm (-ə) cəxʷ ʔ

"Are you going?" (from Shaw et al 2008)

Similarly, the oblique prepositional marker [ʔə], which has the same form as the Yes-No marker, is always deleted in HL, whereas in HN its deletion is optional; see a comparison between the HN and HL forms in (b) below.

(b) HN: niʔ wɛl-x-əs tə sqəməl stətɛs (ʔə) ƛ

Aux throw-tr-3tr.Su Det paddle Loc Prep Det

ʔɛ:nθə

1s.Indep

HL: lə wɛl-xʸ-əstə sqəməl stətɪs [Ø] ƛə ʔɛlθə

"He threw a paddle beside me" (taken from Shaw et al 2008)

Shaw et al suggest that, as [ʔ] and [ə] are unmarked in the phonology and are also vulnerable in weak prosodic positions, they are susceptible to being phonologically eroded, and this can result in restructuring of the clause.

conducted, but that they are also now found with a lexical noun as subject, as shown in (73).

(73)	tə	John	kʷə́-l-əxʷ	tə	pə́kʷ-s
	S		V		O
	DET	John	find-TR-3OBJ	DET	book-3POSS

"John found his book." (taken from Shaw et al 2008:11)

Furthermore, while the recent HL data contains frequent occurrences of this S-initial order, the same study shows conversely that HN speakers still tend to retain the VS word order of conservative Salish. Shaw et al propose that HL speakers demonstrate a word order shift from VS to SV, whilst HN speakers do not demonstrate this change. Crucially, they associate this with the influence of English (an SV language) as the majority language in the upriver community where HL is spoken. English is somewhat less pervasive in the more rural downriver HN-speaking community, perhaps explaining why that dialect remains more conservative than the upriver dialect.

I propose there are some similarities between the Central Salish phenomenon and AD in Welsh. Both Welsh and Salish are primarily verb-initial languages, and studies of data from speakers of both languages show some constructions where an auxiliary or auxiliary-like morpheme is deleted due to both phonological weakening and influence of English (specifically, bilingualism). In Central Salish the auxiliary particle [ʔə] or [ʔɪ] has been phonologically eroded to [-ə] and then [Ø]; in Welsh the auxiliary verb *wyt* (and, less frequently, other verbs) is eroded to [Ø]. The remaining overt word order in clauses where this occurs in both Central Salish and Welsh is S-initial. Note that the data

shows that, in both Central Salish and Welsh clauses of this type, this phenomenon apparently first occurred where the subject was pronominal, before spreading to clauses where the subject was lexical. Finally, both Welsh and Salish can claim extensive and long-term contact with speakers of English. These factors combine to indicate that it seems likely that both Welsh and HL Salish have been influenced by English to undergo a word order change from VS > SV in auxiliary clauses involving a pronoun.

I argued above that AD shows convergence from Welsh to English SV order. Recall that I define word-order convergence as the increase in frequency of word order in Language A on the influence of the same word order in language B. As far as I know, Central Salish does not have native S-initial word order, and so these Central Salish data probably does not show convergence to English, since the V-initial > S-initial change exhibited in HL is not interpretable as an extension of an existing Central Salish word-order. Rather, presumably the new S-initial word order is an innovation by speakers of HL, perhaps best viewed as a direct structural transfer from English, facilitated by phonological weakening of the auxiliary. Nevertheless, this comparison of Welsh and Central Salish word order suggests that typologically-similar languages can undergo similar language change. That is, if these two V-initial languages are undergoing a shift from VS to SV, perhaps this indicates that other V-initial languages are also tending towards S-initial word order. This question demands further research, but it is unfortunately beyond the remit of this thesis.

I have shown in this section that AD due to influence of English is a phenomenon that can be found in a language other than Welsh, and the similarities in the processes seen are striking. A dialect of Central Salish has been shown to be undergoing extensive word order shift to

SV, and, whilst this change is currently not so extensive in Welsh, it may be that the data I have analysed represent the beginnings of a more widespread use of S-initial constructions by Welsh-English bilinguals.

10. Conclusion

In this chapter I have noted that the application of the MLF model to the dataset in chapter 5 resulted in a number of clauses not being able to have an ML assigned to them: this was because of the lack of an auxiliary verb in periphrastic constructions, all apparently involving the verb *bod* 'to be' and mostly with the subject pronoun *ti*. I call this phenomenon auxiliary deletion, and analyse another dataset to show that this is prevalent in the speech of the great majority of the speakers analysed. Only age seems to make a significant difference to the extent of AD: younger speakers delete the auxiliary more frequently than older speakers, though there is also apparently a negative correlation between deletion of the auxiliary and deletion of the pre-verbal-noun particle.

These results indicate that AD in B2S clauses is becoming more frequent in the speech of Welsh-English bilinguals, resulting in those clauses having an overt initial subject, indicative of influence from English SVO word order. I treat this as a form of convergence in Welsh word order towards English.

Even though the application of the MLF model described in chapter 5 resulted in very little word-order convergence being identified, it seems that there is actually quite a major form of convergence underway in Welsh word order that the model fails to fully identify.

Furthermore, AD can be seen in Welsh constructions where the subject is something other than the 2nd singular pronoun *ti*, particularly in the speech of southern speakers of Welsh. It would be useful to analyse more data to see the extent of this phenomenon, but that is beyond the scope of this thesis.

Finally, comparisons with a dialect of the Amerindian language Central Salish suggests that AD in Welsh is not an isolated phenomenon, but is actually seen to occur in more than one V-initial language. As with Welsh, influence from English seems to be behind the word-order change seen in Salish (but I do not interpret the latter to be convergence).

In the final chapter I will conclude the thesis by returning to all the points raised so far, discussing the extent of word-order convergence in Welsh, the ramifications of the research, and indicating where future research stemming from the findings of this study may lead.

Chapter 7 — Conclusions and future research

1. The research presented in this thesis

The aim of the research I presented in this thesis was to examine the extent of word-order convergence in Welsh due to the influence of English word-order. In this concluding chapter I return to the proposals I made and the findings I presented, and I discuss what contribution my thesis makes to our understanding of word-order convergence in Welsh. I also suggest some future paths research on this topic may take.

I defined convergence (chapter 2) as a process which leads to the enhancement of inherent structural similarities between two languages in contact (Bullock and Toribio 2004). This entails the increase in frequency of a category in one language which is shared by the other language (Thomason 2001), and the process is manifested by a clause with morphemes from one language but structure that is more common in another (Myers-Scotton 2002, Schmitt 2000). I proposed (chapter 3) that Myers-Scotton's (2002) Matrix Language Frame (MLF) model can be used to identify not only cases of classic code-switching, where the morphosyntactic structure of a clause comes from one language, but also composite code-switching, where the morphosyntactic structure of a clause comes from both participating languages (and the morphemes from one or both). To achieve this I presented the innovative concept of a dichotomous matrix language,

which describes a clause where the two principles of the MLF model do not indicate the same language to be the source of the morphosyntactic structure of a clause.

To test this theory with regards to Welsh, I analysed (chapter 5) a dataset of informal conversational data from six speakers taken from the Welsh-English corpus Siarad collected at Bangor University (as described in chapter 4). These speakers were all bilinguals who either acquired Welsh as a first language or simultaneously with English. They all used Welsh more frequently than English as the ML in monolingual clauses, and, more notably, used Welsh exclusively as the ML for bilingual clauses, with the exception of the one bilingual clause which Amranwen produced which had a dichotomous ML. This shows that these speakers always use Welsh for the morphosyntax of their code-switching utterances, indicating that Welsh has morphosyntactic dominance over English in this respect.

With regard to word-order convergence, only two clauses were identified in this dataset with a dichotomous ML. I suggested that they shows word-order convergence from Welsh to English, whereby a head/modifier NP in those clauses (*drws nesa pobl* 'next-door people' and *Cinderella type peth* 'Cinderella type thing') has the modifier preceding the head, a more English-like word order, rather than following it, the usual Welsh order, even though other information in the clause, such as the relative position of the finite verb and subject, indicate that Welsh is the source of clause structure. Although these clauses seem to show word-order convergence, they are the only ones in the dataset found in this way. I identified, via an informal search of the whole Siarad corpus, only 10 other clauses that could be labelled as having a dichotomous ML, all also because of word order in an NP. The inference is that there is almost no word-order convergence in the

speech of speakers in the Siarad corpus, at least when using the methodology of the MLF model.

In order to show that the dichotomous ML concept is effective in identifying clauses which show word-order convergence, I presented an analysis of a small set of data taken from an Austrian German-English corpus collected by Eppler (2004). In these data I identified 13 clauses (out of 2142) as having a dichotomous ML. I argued that all those clauses showed word-order convergence towards English, where a German participle or nonfinite verb in a main clause which would be expected to occur clause-finally in Austrian German instead appears directly following the finite verb, reflecting more English-like SVO word order. The participants recorded in Eppler's data are elderly Viennese immigrants who have lived England for many decades, and their exposure to English may be said to be more intense than that experienced by most Welsh speakers, so it is not unexpected for their morphosyntax to have been heavily influenced by English, more than with the Welsh-English bilinguals. This would explain why more dichotomous ML clauses are found in that dataset, indicating that the dichotomous ML concept is effective in reflecting the extent of convergence in a given dataset.

I then (chapter 6) noted that it was possible that there were some kinds of Welsh-English clauses which might show word-order convergence but which would not be identifiable as having a dichotomous ML. These included clauses with null elements, which are not ideal from the perspective of the MLF model, since its application favours overt surface morphemes. One such type of clause include Welsh auxiliary constructions where the initial finite auxiliary is deleted, resulting in a clause-initial subject. I noted that auxiliary deletion (AD) was common in the dataset analysed for chapter 5 in clauses where the

subject was the 2nd person singular pronoun *ti*. I proposed that clauses with -A might show convergence towards English SVO word order, and, to discover the extent to which this was the case, I analysed a second dataset from the Welsh-English corpus for the extent of AD in clauses where the subject is *ti*. This dataset consisted of data from 28 Welsh-English bilinguals of varying ages, sexes and from various backgrounds.

I demonstrated how these speakers deleted the auxiliary verb in 92.75% of clauses where the subject is *ti*. There is no significant difference between men and women when it comes to this deletion, nor do northern Welsh speakers delete the auxiliary more than southern speakers, but age, on the other hand, is a factor in the extent of -A. Young speakers produce -A significantly more frequently than older speakers. I interpret this, adopting the apparent time paradigm and considering the late adoption hypothesis (Boberg 2004), to indicate that the increase of -A in the speech of Welsh speakers is a change in progress. I examined its occurrence where the subject was a morpheme other than *ti*, and showed that it also occurs in these data where the subject is another pronoun or a lexical noun, though this occurs in the speech of fewer speakers than with *ti*.

I argued that AD in Welsh is caused by both phonological factors (weakening of the vowel of the finite verb *wyt*) and the influence of English SVO word order, a word order which is already available in Welsh, albeit in limited contexts. I suggested that it could reflect an ongoing process of shift whereby Welsh auxiliary constructions in general will converge to become subject-initial, and I made comparison with Central Salish, a language also showing shift to SVO word order due to the influence of English. AD might reflect a more widespread word-order shift in Welsh towards SVO, but it is perhaps too early to predict how far this change will go.

2. Future research

Some issues were raised during this research which were of interest and were relevant to the study of convergence in Welsh, but were beyond the scope and boundaries of the focus of this thesis. They are, however, worth pursuing further in the future, and I make some suggestions here as to what form such research could take.

First, the MLF analysis presented here showed that bilingual clauses with an English ML do not exist in the dataset I analysed. It may be the case that they are found elsewhere in the Siarad corpus, though they are probably few. They may also be found in the speech of bilinguals who have English as L1 more than in the speech of those who are Welsh L1, for example.[193] If English ML bilingual clauses are found with greater frequency outside the data analysed here, then that could be interpreted to indicate that English is more morphosyntactically-dominant in the speech of those speakers than it was for the speakers I analyse here. The repercussions of this could be that Welsh would be considered less morphosyntactically-dominant in the Welsh-speaking community in general, and that my predictions on the security of Welsh structure from English influence should be modified. In order to ascertain this, one avenue of research could be to find and record more Welsh-English bilinguals for whom English is a more dominant language (sociolinguistically, psycholinguistically, etc.), test the MLF model on the data they produce, and see how the results of that analysis compare with the analysis I present here. If there is a significant difference in the

[193] As noted in chapter 4, there are some speakers in our corpus who are L2 Welsh.

number of English ML bilingual clauses found, that could have interesting repercussions on our knowledge of Welsh structure.

Second, the same analysis showed that only two clauses could be labelled as having a dichotomous ML. That was because of the influence of English word order in a head/modifier NP. A search of the entire Siarad corpus yielded very few other dichotomous ML clauses, indicating that my analysis of three conversations is representative of the scarcity of this kind of convergence in the speech of Welsh-English bilinguals in general. Future research could see if this form of convergence becomes more prevalent, by analysing the head/modifier word order of Welsh-English bilinguals as part of a diachronic study, at e.g. five-year intervals. If such convergence were to be found with more frequency, then that would be another indicator of the increase in the influence of English on Welsh constituent structure.

Third, with regard to AD in Welsh auxiliary constructions, I have noted above how I observe that the trend may be spreading in clauses where the subject is something other than the 2nd person singular pronoun. I was not able to analyse this quantitatively in this thesis, and so an opportunity for future research could be to analyse more auxiliary constructions produced by Welsh-English bilinguals, focusing on the other pronouns and on lexical-noun subjects. The proportion of -A in such clauses could then be compared to the -A proportion I found in these data (92.75%). If there were a marked difference, then it could be inferred that AD is only prevalent in clauses where the subject is *ti*. If the proportions were similarly high, that could indicate that AD is more prevalent than I have suggested it is here. The Siarad corpus would be an excellent source for such data, although a longitudinal study could also be commenced, whereby speakers some years from now could be analysed to see the extent of AD they produce (with *ti* as subject and

elsewhere). If this proportion is higher than the findings I present in this thesis, then it would be a clear indicator of the spread of this word-order shift.

My personal research interests continue to be focused on the structural effect of English on Welsh, in word order and elsewhere. My immediate intentions for research are to extend the scope of my analysis of AD to look at other clauses with subjects other than *ti*, as well as examine data from other speakers. I would like to determine the connection between language dominance and convergence, since it may be the case that speakers for whom English is a more dominant language than Welsh produce more convergence than those who are the reverse. I will use the Siarad corpus as a starting point for these analyses, but in the future I also intend to collect further data, so as to observe what diachronic effects are visible in Welsh as we proceed further into the 21st century.

3. Word-order convergence in Welsh

What do my findings in this thesis indicate about word-order convergence, then, in Welsh? To what extent were the predictions that Saunders Lewis made in 1962 true, that Welsh would be overwhelmed by English—to all intents and purposes, dead—by the dawn of the 21st century?

My analysis of the first Welsh-English dataset, applying the MLF model with my added nuance of the dichotomous ML concept, strongly indicates that Welsh is dominant for the speakers analysed, in terms of from which language they source the morphosyntax of the bilingual clauses they produce. This is not only the norm, but apparently the

universal trend for these speakers. In addition, the dearth of dichotomous ML clauses found in these data show, I argue, that the structural effect of English on the word order of these bilinguals' utterances is minimal.

Despite this, the results of the analysis of auxiliary deletion does suggest that there may the influence of English may be felt in some parts of Welsh structure. The preponderance of subject-initial clauses where the subject is the pronoun *ti*, as well as with less frequency with other subjects, indicates that auxiliary deletion is currently common in the speech of Welsh-English bilinguals. Furthermore, it may become more widespread in the future, given how I have shown that speakers from younger generations produce clauses with AD significantly more frequently than speakers from preceding generations. If the trend continues, perhaps the next generation of bilinguals will produce AD even more frequently. I have suggested that English is (partly) the influence behind this increase in AD; even if it is not, Welsh may still be undergoing a word-order shift in main clauses to being subject-initial. If this occurs, then the language will exhibit quite extensive structural convergence towards English. At the moment, nevertheless, AD is only found in a limited set of Welsh constructions, and so should not be taken as a sign that suddenly every category and construction in Welsh will converge towards English. The indication is that this is not the case.

In chapter 1 of this thesis I posited a continuum, at one end of which there was no convergence and thus the future of Welsh was secure, and at the other end of which there was widespread convergence and thus the future of Welsh was in peril. The evidence from the study I have presented in this study puts Welsh somewhere in between, perhaps more towards the secure end of the continuum. However, if the auxiliary deletion identified does indeed signify the

beginnings of a general shift in Welsh main clause word-order towards SVO, then Welsh morphosyntax in the future may yield more ground to English influence. But, for now, Welsh speakers can remain cautiously optimistic about the fate of their language. She is not dead yet.

END

APPENDICES

APPENDIX A

This is an example of one of the letters which were sent to potential participants we wished to recruit. This English version was sent back-to-back with a Welsh translation.

Dear Sir or Madam,

Bilingual Communication in Wales

We are involved with a research project on how bilingual people communicate with each other in Wales. This is being conducted by Professor Margaret Deuchar of the University of Wales, Bangor, together with a team of researchers.

We are looking for bilingual people such as you to record having an informal conversation with a bilingual member of the family or friend. You would be welcome to choose the bilingual person you would like to be recorded with, and the place you would like us to make the recording. We can come to your home or your workplace, or if you prefer you can come to the University. The recording would take about 45 minutes and we would also ask you to fill in a short questionnaire. The whole session should not take more than an hour, and we would pay £10 each to you and your friend or family member.

Your conversation would just be a natural, relaxed chat. We can however give you some ideas of topics you can talk about to start you off. After the recording you would be given the opportunity to listen to the recording in case there were any part of it you did not want us to keep. Once you have given us your permission to keep the recording, we will analyse it anonymously for research purposes.

If you are interested, please fill in the slip on this letter and return it to your school, or you can contact one of our project staff directly by telephone at (01248) 383878 or 382559 (leaving a message if your reach our answering machine) or by email at siarad@bangor.ac.uk. We will then get back to you as soon as we can. (continued...)

If you have any further questions, please contact us. Our address for letters is: Professor M. Deuchar, Dept of Linguistics, University of Wales, Bangor, Bangor, Gwynedd LL29 7ED.

We look forward to hearing from you soon.

Yours sincerely,

Professor Margaret Deuchar
Head of Linguistics

(Please complete and return this slip to the school if you're interested.)

I would be interested in taking part in the bilingualism project. Please put my details on your records for your consideration. I understand that I may be contacted in the near future with further details.

Name:
.....
Signature:
.....
Address:.....
.....
..... Telephone
number:.....
E-mail:.....

Below is a copy of one of the posters used to attract potential participants.

YDYCH CHI EISIAU £10?

Os ydych chi'n gallu siarad Cymraeg a Saesneg, helpwch ni ym Mhrifysgol Cymru, Bangor gyda'n prosiect am ddwyieithrwydd

Sut? Drwy gael eich recordio yn cael sgwrs am unrhyw beth yn y byd efo ffrind neu aelod eich teulu (tua 30-45 munud), a llenwi holiadur byr wedyn

Pryd/Ble? Pryd a ble bynnag sy'n gyfleus i chi – dros y ffôn hyd yn oed!

Eisiau rhagor o fanylion?

Ffoniwch 01248 383878

Anfonwch eich manylion mewn e-bost at siarad@bangor.ac.uk

DO YOU WANT £10?

If you can speak Welsh and English, help us at the University of Wales, Bangor with our project about bilingualism

How? By being recorded having a chat with a friend or family member about anything at all (for about 30-45 minutes), and filling in a short questionnaire afterwards

When/Where? Whenever and wherever's convenient for you – even over the phone!

Want more details?

Phone 01248 383878

Send your details by e-mail to siarad@bangor.ac.uk

APPENDIX B

This is a copy of the English version of the speaker questionnaire administered to each participant. A Welsh version was also produced and supplied back-to-back with this version.

Questionnaire

We would be grateful if you could give us the following background information to help us with our studies.

1. Are you: Male ☐ Female ☐ ?
2. Date of birth:.....

3. What is your present occupation (or if retired or unemployed, what was your last occupation before retiring or becoming unemployed)?

.....
.....

4. Please indicate the areas where you have lived for significant periods of your life:

e.g.: Place: *Llandegfan, Ynys Môn* Dates: *1975-93*
Place: *Liverpool* Dates: *1993-99*
Place: *Melbourne, Australia* Dates: *1999-2002*
Place: *Bethesda, Gwynedd* Dates: *2002-05*

Place: ..... Dates: .....
Place: ..... Dates: .....
Place: ..... Dates: .....
Place: ..... Dates: .....
Place: ..... Dates: .....
Place: ..... Dates: .....

5. What is the highest level of formal education you have completed?

☐ GCSE, O-level/CSE, School Certificate, NVQ level 1 or 2 or equivalent
☐ A/AS level, Higher School Certificate, GNVQ, BTEC National Diploma,
NVQ level 3 or equivalent
☐ Bachelor's Degree, Diploma of Higher/Further Education, PGCE, HND,
NVQ level 4 or equivalent
☐ Master's Degree, Doctorate, NVQ level 5 or equivalent
☐ None of the above

6. Since when have you been able to speak Welsh?
☐ Since I was 2 years old or younger
☐ Since I was 4 years old or younger
☐ Since primary school
☐ Since secondary school

☐ I learned Welsh as an adult

7. Since when have you been able to speak English?
☐ Since I was 2 years old or younger
☐ Since I was 4 years old or younger
☐ Since primary school
☐ Since secondary school
☐ I learned English as an adult

8. On a scale of 1 to 4, how well do you feel you can speak Welsh?
☐ **1** Only know some words and expressions
☐ **2** Confident in basic conversations
☐ **3** Fairly confident in extended conversations
☐ **4** Confident in extended conversations

9. On a scale of 1 to 4, how well do you feel you can speak English?
☐ **1** Only know some words and expressions
☐ **2** Confident in basic conversations
☐ **3** Fairly confident in extended conversations
☐ **4** Confident in extended conversations

10. Which language(s) did your mother speak to you while you were growing up (if applicable)?
☐ Welsh
☐ English
☐ Welsh & English
☐ Other (Please specify).....
☐ N/A

11. Which language(s) did your father speak to you while you were growing up (if applicable)?
☐ Welsh
☐ English
☐ Welsh & English
☐ Other (Please specify).....
☐ N/A

12. Which language(s) did any other guardian or caregiver speak to you while you were growing up (if applicable)?

☐ Welsh
☐ English
☐ Welsh & English
☐ Other (Please specify).....
☐ N/A

13. Through which language(s) were you predominantly taught at primary school?

☐ Welsh
☐ English
☐ Welsh & English

☐ Other (Please specify).....

14. Through which language(s) were you predominantly taught at secondary school?

☐ Welsh
☐ English
☐ Welsh & English
☐ Other (Please specify).....

15. Make a list below of five of the people you speak to most in your everyday life, either in person or on the phone, e.g. your partner, your child, a friend, a workmate etc. Then note which language(s) you mostly speak with that person, as shown in the sample table:

Name of person, or relationship	**Language mostly spoken with that person: (place a tick in one cell below for each line)**			
	Welsh	**English**	**Equally Welsh & English**	**Another language**
1. *Sian*	✓			
2. *Mother*		✓		
3. *Boss*			✓	
4. *Jacques*				✓
5. *Sister*		✓		

Please fill in table below

Name of person, or relationship (use fictitious names if you prefer)	**Language mostly spoken with that person: (place a tick in one cell below for each line)**			
	Welsh	**English**	**Equally Welsh & English**	**Another language**
1.				
2.				
3.				
4.				
5.				

16. How would you rate the Welsh language on a scale of 1 to 5 regarding the following properties? Circle one number in each line.

	←				→	
old-fashioned	1	2	3	4	5	modern
unfriendly	1	2	3	4	5	friendly
uninfluential influential	1	2	3	4	5	
uninspiring inspiring	1	2	3	4	5	

useless	1	2	3	4	5	useful
ugly	1	2	3	4	5	
beautiful						

17. How would you rate the English language on a scale of 1 to 5 regarding the following properties? Circle one number in each line.

old-fashioned	1	2	3	4	5	modern
unfriendly	1	2	3	4	5	friendly
uninfluential	1	2	3	4	5	
influential						
uninspiring	1	2	3	4	5	
inspiring						
useless	1	2	3	4	5	useful
ugly	1	2	3	4	5	
beautiful						

18. Do you consider yourself to be mainly.....?

☐ Welsh
☐ English
☐ Scottish
☐ Irish
☐ British
☐ Other (please specify):.....

19. To what extent do you agree with the following statement:
"In everyday conversation, I keep the Welsh and English languages separate."

☐ **1** Strongly disagree
☐ **2** Disagree
☐ **3** Neither agree nor disagree
☐ **4** Agree
☐ **5** Strongly agree

20. To what extent do you agree with the following statement:
"People should avoid mixing Welsh and English in the same conversation."

☐ **1** Strongly disagree
☐ **2** Disagree
☐ **3** Neither agree nor disagree
☐ **4** Agree
☐ **5** Strongly agree

Speaker's Consent

I hereby give my permission for the information I have given on the above questionnaire to be used for research and/or teaching purposes only (including research publications and/or reports) subject to strict preservation of my anonymity.

I also hereby give my permission for my recording (sound and transcript) to be contributed to linguistic archives, including the international archive *LIDES (http://talkbank.org/data/LIDES)* on the internet, provided that the names of speakers and other persons named during the conversation are replaced by fictitious names in the transcript.

I also hereby agree to permit full access to these data to researchers provided they subscribe to the relevant code of ethics (for the *Talkbank Code of Ethics* see *http://talkbank.org/share/ethics.html*). I also understand that, by signing this consent form, I give the aforementioned researchers permission to present excerpts of these data as part of their work in written and/or in oral form, without further permission from me.

I hereby assign the copyright in my contribution to the Project Director, Professor Margaret Deuchar.

Name.....
.....

Address.....
.....
.....
..... Post
Code.....
.....
Signature
.....
Date
.....

APPENDIX C

Sample of the spreadsheet used to apply the MLF model to the data.

A	B	C	D	T	U	V	W	Y
I	Speaker	Clause	Complex or Sim	Verb ty	Inflected verb is..	Verb agreement points to	# of criteria available	ML
33	LIS	achos@1 (y)na@1 fi@1 (we)di@1 tyfu@1 lan@1 .	Simple	fin	n/a	Welsh	1	Welsh
34	MAB	+< yeah@0 .	Simple	none	n/a	n/a	0	UI
35	LIS	in@2 the@2 country@2 &=laugh .	Simple	none	n/a	n/a	0	UI
37	MAB	country@2 bumpkin@2 &=laugh .	Simple	none	n/a	n/a	0	UI
38	LIS	dw@1 i@1 (y)n@1 love_o@21 (y)r@1 gwlad@1 though@2 .	Simple	fin	dw	Welsh	2	Welsh
39	LIS	mae@1 [?] fe@1 mor@1 nice@0 .	Simple	fin	mae	Welsh	2	Welsh
40	MAB	yeah@0 dw@1 i@1 [/] dw@1 i@1 (y)n@1 hoffi@1 fo@1 &hev +/.	Simple	fin	dw	Welsh	2	Welsh
41	MAB	dw@1 i@1 licio@1 y@1 ffaith@1 ~~bod@1 ti@1 (y)n@1 gallu@1 g~~	Complex	fin	dw	Welsh	2	Welsh
41	MAB	bod@1 ti@1 (y)n@1 gallu@1 ~~gweld@1 y@1 sêr@1 yn@1 nos@~~	Complex	nonfin	n/a	n/a	0	UI
41	MAB	gweld@1 y@1 sêr@1 yn@1 nos@1 .	Complex	nonfin	n/a	n/a	0	UI
42	LIS	yeah@0 .	Simple	none	n/a	n/a	0	UI
43	MAB	mae@1 hynna@1 (y)n@1 great@0 .	Simple	fin	mae	Welsh	2	Welsh
44	LIS	dyma@1 [?] ni@1 mynd@1 mas@1 gyda@1 sleeping@2 bag@(	Simple	nonfin	n/a	n/a	0	UI
44	LIS	mae@1 fi@1 a@1 ffrind@1 xxx [///]	Simple	fin	mae	Welsh	2	Welsh
45	LIS	a@1 cysgu@1 mas@1 tu@1 fas@1 .	Simple	nonfin	n/a	n/a	0	UI
46	MAB	+< yeah@0 [=! laugh] .	Simple	none	n/a	n/a	0	UI
47	LIS	a@1 fel@1 # gwylio@1 (y)r@1 sêr@1 um@0 +...	Simple	nonfin	n/a	n/a	0	UI
48	LIS	ond@1 # mae@1 fe@1 (we)di@1 bwrw@1 glaw@1 arnon@1 ni(	Simple	fin	mae	Welsh	2	Welsh
49	MAB	+< aww@0 .	Simple	none	n/a	n/a	0	UI
50	LIS	unwaith@1 fel@1 # um@0 blwyddyn@1 [/] # blwyddyn@1 cyn@1	Complex	fin	aethon	Welsh	2	Welsh
50	LIS	a@1 surf_io@21 fel@1 .	Complex	nonfin	n/a	n/a	0	UI
50	LIS	dw@1 i@1 meddwl@1 .	Complex	fin	dw	Welsh	2	Welsh
51	MAB	+< yeah@0 .	Simple	none	n/a	n/a	0	UI
52	LIS	achos@1 um@0 penwythnos@1 Pasg@1 neu@1 rywbeth@1 oe	Simple	fin	oedd	Welsh	2	Welsh
53	LIS	oh@0 na@1 .	Simple	none	n/a	n/a	0	UI
54	LIS	y@1 blwyddyn@1 cyn@1 (hyn)ny@1 oedd@1 wythnos@1 Pasg(	Simple	fin	oedd	Welsh	2	Welsh
55	MAB	mm@0 .	Simple	none	n/a	n/a	0	UI
56	LIS	y@1 flwyddyn@1 hon@1 fel@1 Mehefin@1 oedd@1 o@1 neu@	Simple	fin	oedd	Welsh	2	Welsh
57	MAB	+< yeah@0 .	Simple	none	n/a	n/a	0	UI

fusser27 | davies6 | fusser6 | Analysis of ML | Individual speakers | Figures for Thesis

APPENDIX D

Sample of the spreadsheet used to analyse auxiliary deletion in the data.

APPENDIX E

The following is a sample of a transcription of a recording from the Siarad corpus. It is the beginning of Davies6.

	A	B	C	D	E	F	G	H	K	T	U	V
1	File	Speaker	ID (al	Clause	-A?	Declarative	Interrogative	WH or YN interrogati	Negative	Negation system t	Lingualit	Particle retention?
92	Fusser14	AWE	1428	ti cael llun.	y	y	n	x	n		m	n
93	Fusser14	AWE	1707	ti cofio hwnna?	y	n	y	yn	n		m	n
94	Fusser14	AWE	1831	lle ti isio [?]...?	y	n	y	wh	n		m	x
95	Fusser14	AWE	1858	ti licio anifeiliaid dwyt.	y	y	n	x	n		m	n
96	Fusser14	AWE	1861	pa anifail ti licio orau?	y	n	y	wh	n		m	n
97	Fusser14	AWE	2143	banana ti fwyta xx.	y	y	n	x	n		m	n
98	Fusser14	AWE	2304	< ti licio bownsio dwyt.	y	y	n	x	n		m	n
99	Fusser14	AWE	2391	ti gallu deud xxxx?	y	n	y	yn	n		m	n
100	Fusser14	AWE	2514	ti licio mynd i pictures?	y	n	y	yn	n		m	n
101	Fusser14	AWE	2523	< be arall ti 'di weld?	y	n	y	wh	n		m	y
102	Fusser14	AWE	2644	< ti nabod Elfyn?	y	n	y	yn	n		m	n
103	Fusser14	AWE	2724	ti fod i fynd â # xx [#] het blisman?	y	n	y	yn	n		m	n
104	Fusser14	AWE	2808	< wna i sticio fo ar dy wal di os ti isio yeah # yn y llofft?	y	y	n	x	n		m	x
105	Fusser14	BEL	47	ti gallu dilyn o?	y	n	y	yn	n		m	n
106	Fusser14	BEL	292	a ti cofio Saint_Malo?	y	n	y	yn	n		m	n
107	Fusser14	BEL	324	ti gallu cael xx a geith mam a dad yn siarad am y gwyliau # yeah'	y	y	n	x	n		m	n
108	Fusser14	BEL	345	ti isio siarad efo fi rwan?	y	n	y	yn	n		m	n
109	Fusser14	BEL	356	ti cofio?	y	n	y	yn	n		m	n
110	Fusser14	BEL	370	a ti'n cofio chwarae?	y	n	y	yn	n		m	y
111	Fusser14	BEL	422	ond ti ddim wedi sylwi hynny.	y	y	n	x	y	dim	m	y
112	Fusser14	BEL	435	ond ti lciio [/] # ti licio teithio hyd yn oed mwy na fi.	y	y	n	x	n		m	n
113	Fusser14	BEL	441	uh # a ti licio # llefydd newydd.	y	y	n	x	n		m	n
114	Fusser14	BEL	718	ti [?] 'n cofnodi be dan ni'n gael # bwyta [#] be dan ni'n bwyta bob	y	y	n	x	n		m	y
115	Fusser14	BEL	744	< ti [/] ti 'di &s sgwennu fo lawr.	y	y	n	x	n		m	y
116	Fusser14	BEL	779	ti cofio # ni # bwyt(a) [/] uh bwyta allan?	y	n	y	yn	n		m	n
117	Fusser14	BEL	914	ti 'di roi crempog a(r) [?] # fy mhen yeah?	y	n	y	yn	n		m	y
118	Fusser14	BEL	920	mae gen [#] &m ti'n gallu defnyddio dy cwsg &d [#] # dy uh # clwt	y	y	n	x	n		m	y
119	Fusser14	BEL	1072	oh a ti'n cofio...?	y	n	y	yn	n		m	y
120	Fusser14	BEL	1312	ti cofio dynes # yn y llun?	y	n	y	yn	n		m	n
121	Fusser14	BEL	1386	ti cofio pla(nt) [/] plant yn canu?	y	n	y	yn	n		m	n

ALL DATA | General Analysis | Stammers5 | Davies6 | Fusser14 | Robert6 | Robert8 | Fu

@Begin
@Languages: cy, en, sp
@Participants: DAN Daniel Adult, HEC Hector Adult
@ID: cy|siarad|DAN|25;|male|||Adult||
@ID: cy|siarad|HEC|23;|male|||Adult||
@Situation: Informal conversation in a University meeting room between two close friends
@Comment: DAN was brought up in north-west Wales and has lived there all his life
@Comment: HEC was brought up in north-west Wales and has lived there all his life
@Date: 02-AUG-2006
@Comment: Researcher: Peredur Davies, Bangor University
@Coder: Peredur Davies, Bangor University
@Time Duration: 00:34:51
@Comment: Filename: davies6.cha; Soundfile: davies6.wav.
@Comment: Language markers: @1 = Welsh, @2 = English, @0 = Undetermined, @21 = word with first morpheme(s) English, second morpheme(s) Welsh, @12 = word with first morpheme(s) Welsh, second morpheme(s) English.
@Comment: speakers had bought along tabloid newspapers which they read during the interview, and often quote headlines/stories from in the transcription.
@Warning: HEC's radio microphone cuts off (due to exhausted batteries) repeatedly during the first halt of the recording.
*DAN: fod@1 i@1 gael@1 un@1 iddyn@1 nhw@1 ddoe@1 tra@1 o'n@1 i@1 (y)n@1
fod@1 yn@1 Builth@2 . •%snd:"davies6"_0_2032•
%gls: be.NONFIN PRT get.NONFIN one to.3PL PRON.3PL yesterday whilst be.1S.IMP PRON.1S PRT be.NONFIN in Builth
%eng: supposed to get one for them yesterday whilst I was in Builth.
*HEC: oe(dd)@1 [?] . •%snd:"davies6"_2188_3019•
%gls: be.3S.IMP
%eng: really?
*HEC: mam@0 a@1 dad@0 a@1 bobl@1 ddiarth@1 <ydy@1 o@1 wedyn@1> [?] xx . •%snd:"davies6"_3053_5027•
%gls: mother and dad and people unfamiliar be.3S.PRES PRON.3SM after xx
%eng: it's mum and dad and strangers afterwards.
*HEC: (oe)s@1 (dd)im@1 # mynedd@1 bod@1 o_gwmpas@1 tŷ@1 de@1 . •%snd:"davies6"_5348_7802•
%gls: be.3S.PRES NEG patience be.NONFIN around house TAG

%eng: I can't be bothered being around the house, eh.
*HEC: <wneu(d)@1 fi@1> [?] bloody@0 torri@1 grass@2 ddoe@1 # achos@1 bod@1
pobl@1 ddiarth@1 yn@1 dod@1 . •%snd:"davies6"_8522_11433•
%gls: make.NONFIN PRON.1S bloody cut.NONFIN grass yesterday because be.NONFIN people unfamiliar PRT come.NONFIN
%eng: made me cut the bloody grass yesterday because strangers were coming.
*HEC: ddeudais@1 i@1 wrth@1 Dad@0 [=! laugh] <bloody@0 hell@2> ["] . •%snd:"davies6"_12126_14170•
%gls: say.1S.PAST PRON.1S to Dad bloody hell
%eng: I said to Dad, "bloody hell".
*HEC: +" (dy)dyn@1 nhw@1 (ddi)m@1 mynd@1 [?] i@1 # sylwi@1 pa@1 mor@1
hir@1 (y)dy@1 dy@1 fucking@2 grass@2 di@1 na(g)@1 (y)dy@1 ! •%snd:"davies6"_14173_17485•
%gls: be.3PL.PRES.NEG PRON.3PL NEG go.NONFIN PRT notice.NONFIN which so
long be.3S.PRES POSS.2S fucking grass PRON.2S NEG be.3S.PRES
%eng: they're not going to notice how long your fucking grass is, eh!
*HEC: (y)r@1 Arglwydd@1 mawr@1 . •%snd:"davies6"_17782_18558•
%gls: DET Lord great
%eng: Good Lord.
*HEC: +" <gwair@1 Dean_Andrews@0> [?] <wedi@1 tyfu@1 yn@1 ddiawledig@1>
[=! laugh] . •%snd:"davies6"_18749_21832•
%gls: grass Dean_Andrews PRT.PAST grow.NONFIN PRT devilish
%eng: "Dean Andrews's grass has grown terribly!"
*DAN: +< &=laugh . •%snd:"davies6"_20567_21821•
*DAN: &=groan . •%snd:"davies6"_23169_23843•
*HEC: callia@1 de@1 . •%snd:"davies6"_23773_25816•
%gls: become_wise.2S.IMPER TAG
%eng: Get real, eh.
*DAN: Arglwydd@1 mawr@1 ! •%snd:"davies6"_26027_28359•
%gls: Lord great
%eng: Good Lord!
*DAN: p(wy)@1 uffar@1 (y)dy@1 hwnna@1 ? •%snd:"davies6"_28842_31603•
%gls: who hell be.3S.PRES that_one
%eng: Who the hell is that?
*DAN: &=laugh <James_Blunt@0 (y)dy@1 hwnna@1 yeah@0> [=! laugh] ? •%snd:"davies6"_33672_34924•

%gls: James_Blunt be.3S.PRES that_one yeah
%eng: that's James Blunt, yeah?
*HEC: dw@1 (ddi)m@1 (gwy)bod@1 . •%snd:"davies6"_35042_35551•
%gls: be.1S.PRES NEG know.NONFIN
%eng: Dunno.
*DAN: &=laugh . •%snd:"davies6"_35493_36723•
*HEC: &=yawn . •%snd:"davies6"_37049_38848•
*DAN: +< dw@1 (ddi)m@1 (gwy)bod@1 chwaith@1 [=! laugh] . •%snd:"davies6"_38283_39092•
%gls: be.1S.PRES NEG know.NONFIN either
%eng: I dunno either.
*DAN: <edrych@1 fath@1 â@1 Mick_Jagger@0> [=! laugh] . •%snd:"davies6"_39094_40427•
%gls: look.NONFIN kind with Mick_Jagger
%eng: looks like Mick Jagger.
*HEC: xxx two@2 hundred@2 million@0 mae@1 o@1 werth@1 xxx ? •%snd:"davies6"_41268_43318•
%gls: xxx two hundred million be.3S.PRES PRON.3SM worth xxx
%eng: [...] he's worth 200 million [...]?
*DAN: ia@1 . •%snd:"davies6"_43587_44433•
%gls: yes
%eng: yes.
*HEC: mae@1 boy@0 # Richards@0 (y)na@1 werth@1 hundred@2 and@2 eighty@2
yndy@1 ? •%snd:"davies6"_45032_47463•
%gls: be.3S.PRES boy Richards there worth hundred and eighty be.3S.PRES
%eng: that Richards bloke is worth a hundred and eighty, isn't he?
*DAN: Keith_Richards@0 yndy@1 . •%snd:"davies6"_47682_48568•
%gls: Keith_Richards be.3S.PRES
%eng: Keith Richards, yes.
*HEC: +< Keith_Richards@0 . •%snd:"davies6"_47853_48457•
*HEC: hwnna@1 ddisgyn@1 allan@1 o@1 (y)r@1 # coconut@0 tree@2 . •%snd:"davies6"_48568_50581•
%gls: that_one fall.NONFIN out of DET coconut tree
%eng: he fell out of a coconut tree.
*HEC: &=burp oh@0 excuse@2 me@2 . •%snd:"davies6"_51106_52479•
%gls: IM excuse me
%eng: oh, excuse me.
*DAN: &=sigh yndy@1 . •%snd:"davies6"_52525_53811•
%gls: be.3S.PRES
%eng: Yes.

*DAN: cofia@1 di@1 mae@1 Charlie_Watt@0 werth@1 # eighty_five@2 . •%snd:"davies6"_54020_57631•
%gls: remember.2S.IMPER PRON.2S be.3S.PRES Charlie_Watt worth eighty_five
%eng: remember, Charlie Watt is worth eighty-five.
*HEC: mmm@0 be@1 (y)dy@1 o@1 ? •%snd:"davies6"_57844_58606•
%gls: IM what be.3S.PRES PRON.3SM
%eng: mmm, what's he?
*HEC: drummer@2 ? •%snd:"davies6"_58656_59059•
*DAN: ia@1 . •%snd:"davies6"_59076_60272•
%gls: yes
%eng: yes.
@Comment: sounds of drinks being slurped.
*DAN: Arglwydd@1 sbïa@1 . •%snd:"davies6"_64596_65722•
%gls: Lord look.2S.IMPER
%eng: my Lord, look!
*DAN: earnings@2 flwyddyn@1 dwytha@1 . •%snd:"davies6"_65884_67011•
%gls: earnings year last
%eng: last year's earnings.
*DAN: eighty@2 mil@2 . •%snd:"davies6"_67011_67765•
*HEC: tax@0 one@2 point@0 rywbeth@1 ia@1 . •%snd:"davies6"_67906_69635•
%gls: tax one point something yes
%eng: tax of one point something, yes.
*DAN: +< <point@0 two@2 [//] <xx nine@2 two@2 five@2> [=! laugh]> [?] . •%snd:"davies6"_68850_70947•
*HEC: +< sut@1 ? •%snd:"davies6"_69577_70366•
%gls: how
%eng: how?
*HEC: sut@1 ? •%snd:"davies6"_70645_70981•
%gls: how
%eng: how?
*HEC: ches@1 i@1 (ddi)m@1 chance@0 i@1 ddarllen@1 o@1 . •%snd:"davies6"_70987_72177•
%gls: get.1S.PAST PRON.1S NEG chance PRT read.NONFIN PRON.3SM
%eng: I didn't get a chance to read it.
*DAN: <sure@0 bod@1 nhw@1 (we)di@1> [///] # gweithio@1 ffwrdd@1 o@1 adra@1 oedden@1 nhw@1 sure@0 de@1 . •%snd:"davies6"_72126_75446•
%gls: sure be.NONFIN PRON.3PL PRT.PAST work.NONFIN away from home be.3PL.IMP PRON.3PL sure TAG
%eng: they probably worked away from home, didn't they.

*HEC: +< ia@1 # neu@1 mynd@1 i@1 ryw@1 tax@0 haven@2 neu@1 (ry)wbeth@1 . •%snd:"davies6"_74886_76932•
%gls: yes or go.NONFIN to some tax haven or something
%eng: yes, or went to some tax haven or something.
*DAN: &=groan . •%snd:"davies6"_76918_77441•
*DAN: Arglwydd@1 mawr@1 . •%snd:"davies6"_77532_80818•
%gls: Lord great
%eng: good Lord.
*DAN: mae@1 peth@1 <Glyn@0 to@2 win@2> ["] yna@1 dal@1 (y)n@1 dop@0 . •%snd:"davies6"_80796_82513•
%gls: be.3S.PRES thing Glyn to win there still in top
%eng: that Glyn-to-win thing is still top.
%com: Referring to a series of Big Brother, the reality TV programme.
*DAN: <yfo@1 a@1 &m &m> [/] # fo@1 a@1 Mickey@0 sy@1 (y)n@1 dop@0 wan@1 # yn_ôl@1 hwn@1 . •%snd:"davies6"_82534_86552•
%gls: PRON.3SM and PRON.3SM and Mickey be.PRES.REL PRT top now after DET
%eng: it's him and Mickey who are top now, according to this.
*HEC: Pete@0 (y)na@1 sy@1 (y)n@1 mynd@1 i@1 ennill@1 efo@1 [?] bookies@0 de@1 . •%snd:"davies6"_86607_89900•
%gls: Pete that REL PRT go.NONFIN PRT win.NONFIN with bookies TAG
%eng: that Pete's going to win according to [?] the bookies, right.
*HEC: mae@1 o@1 (ry)wbeth@1 # stupid@0 (fa)th@1 â@1 ten@2 to@2 one@2 on@2 # i@1 guro@1 . •%snd:"davies6"_89937_92544•
%gls: be.3S.PRES PRON.3SM something stupid like with ten to one on to beat.NONFIN
%eng: he's something stupid like ten to one on to win
*DAN: oh@0 yndy@1 ? •%snd:"davies6"_92102_92544•
%gls: IM be.3S.PRES
%eng: oh, is it?
*HEC: yndy@1 . •%snd:"davies6"_92544_93043•
%gls: be.3S.PRES
%eng: yes
*DAN: ti@1 dal@1 watsiad@1 o@1 ? •%snd:"davies6"_93656_94753•
%gls: PRON.2S continue.NONFIN watch.NONFIN PRON.3SM
%eng: you still watching it?
*HEC: na(g)@1 (y)dw@1 . •%snd:"davies6"_94660_95462•
%gls: NEG be.3S.PRES
%eng: no.

*HEC: mynd@1 off@0 o@1 ddiawledig@1 <deud@1 gwir@1> [?] wrtha@1 chdi@1 . •%snd:"davies6"_95411_97064•
%gls: go.NONFIN off PRON.3SM hellish say.NONFIN truth to.2S PRON.2S
%eng: gone off it a lot to be honest with you.
*HEC: wnes@1 l@1 watsiad@1 rywfaint@1 ohono@1 fo@1 # ti(b)o(d)@1 pan@1
oedden@1 nhw@1 yn@1 y@1 prison@2 task@0 yna@1 . •%snd:"davies6"_96865_101358•
%gls: do.1S.PAST PRON.1S watch.NONFIN some of.3SM PRON.3SM know.2S when
be.3PL.IMP PRON.3PL in DET prison task there
%eng: I watched a bit of it, you know, when they were in that prison task.

(et cetera)

References

Appel, René and Muysken, Peter. 1987. *Language contact and bilingualism*. New York: Edward Arnold.

Aikhenvald, Alexandra Y. 2006. "Grammars in contact: a cross-linguistic perspective" in *Grammars in contact: a cross-linguistic typology*, A.Y. Aikhenvald and R.M.W. Dixon (eds.), 1–66. Oxford: Oxford University Press.

Auer, Peter. 2007. Style and social identities: Alternative approaches to linguistic heterogeneity. Berlin: Walter de Gruyter.

Backus, Ad. 2004. "Convergence as a mechanism of language change". *Bilingualism: Language and Cognition* 7: 179–181.

Backus, Ad. 2005. "Codeswitching and language change: one thing leads to another?". *International Journal of Bilingualism* 9: 307–40.

Barbour, Stephen and Stevenson, Patrick. 1990. *Variation in German: A critical approach to German sociolinguistics*. Cambridge: Cambridge University Press.

Bentahila, Abdelali and Davies, Eirlys. 1983. "The syntax of Arabic-French code-switching". *Lingua* 59: 301–30.

Berk-Seligson, Susan. 1986. "Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism". *Language in Society* 15: 313–48.

Biberauer, Teresa, Newton, Glenda and Sheehan, Michelle. In press. "Impossible changes and impossible borrowings: the Final-over-Final Constraint" in *Continuity and Change in Grammar,* A. Breitbarth, C. Lucas, S. Watts and D. Willis (eds.). Amsterdam: John Benjamins.

Boberg, Charles. 2004. "Real and apparent time in language change: Late adoption of changes in Montreal English". *American Speech* 79, 250–69.

Bolonyai, Agnes. 1998. "In-between languages: language shift/maintenance in childhood bilingualism". *International Journal of Bilingualism* 2: 21–43.

Bolonyai, Agnes. 1999. *The hidden dimensions of language contact: The case of Hungarian-English bilingual children.* University of South Carolina Ph.D. dissertation.

Borsley, Robert. 2005. "A new approach to Welsh negation". Unpublished paper, retrieved from http://privatewww.essex.ac.uk/~rborsley/new-negation.pdf.

Borsley, Robert and Jones, Bob Morris. 2005. *Welsh negation and grammatical theory*. Cardiff: University of Wales Press.

Borsley, Robert, Tallerman, Maggie and Willis, David. 2007. *The syntax of Welsh*. Cambridge: Cambridge University Press.

Bowern, Claire. 2008. "Syntactic change and syntactic borrowing in generative grammar" in *Principles of syntactic reconstruction*, G. Ferraresi and M. Goldbach (eds.), 187–216. Philadelphia: John Benjamins.

Bullock, Barbara E. and Gerfen, Chip. 2004. "Phonological convergence in a contracting language variety". *Bilingualism: Language and Cognition* 7: 95–104.

Bullock, Barbara E. and Toribio, Almeida J. 2004. "Convergence as an emergent property in bilingual speech". *Bilingualism: Language and Cognition* 7: 91–93.

Chambers, Jack K. 1995. *Sociolinguistic theory*. Oxford: Blackwell.

Chambers, Jack K., Trudgill, Peter and Schilling-Estes, Natalie. 2004. *The handbook of language variation and change*. Malden, MA: Blackwell.

Chan, Brian H-S. 2008. "Code-switching, word order and the lexical/functional category distinction". *Lingua* 118: 777-809.

Cheshire, Jenny. 1998. "Linguistic variation and social function" in *Language and gender: a reader*, J. Coates (ed.), 29–41. Oxford: Blackwell.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Clyne, Michael G. 1967. *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia*. The Hague: Nijhoff.

Clyne, Michael G. 1987. "Constraints on code-switching: How universal are they?". *Linguistics* 25: 739–64. Page numbers used in this thesis pertain to the reprinted article in *The bilingualism reader*, L. Wei (ed.), 257–80. New York: Routledge.

Coetsem, Frans van. 1988. *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.

Commissioners of Inquiry into the state of education in Wales. 1847. *Reports of the commissioners of inquiry into the state of education in Wales, &c.* London: W. Clowes and Sons.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. Harlow, Essex: Longman.

Davies, Peredur and Deuchar, Margaret. In press. "Using the Matrix Language Frame model to measure the extent of word-order convergence in Welsh-English bilingual speech" in *Continuity and change in grammar,* A. Breitbarth, C. Lucas, S. Watts and D. Willis (eds.). Amsterdam: John Benjamins.

Davies, John. 1994. *A History of Wales*. London: Penguin.

Declerck, Renaat. 1991. *Tense in English: Its structure and use in discourse*. London: Routledge.

Deuchar, Margaret and Davies, Peredur. 2009. "Code switching and the future of Welsh". *International Journal of the Sociology of Language* 195:15–38.

Deuchar, Margaret. 2005a. "Congruence and Welsh-English code-switching". *Bilingualism: Language and Cognition* 8: 255–269.

Deuchar, Margaret. 2005b. "Minority language survival in northwest Wales: an introduction" in *ISB4: Proceedings of the 4th International Symposium on Bilingualism*, J. Cohen, K.T. McAlister, K. Rolstad and J. MacSwan (eds.), 621–24. Somerville, MA: Cascadilla Press.

Deuchar, Margaret. 2006. "Welsh-English code-switching and the Matrix Language Frame model". *Lingua* 116: 1986–2011.

Deuchar, Margaret, Muysken, Pieter and Wang, Sung-Ian. 2007. "Structured variation in codeswitching: Towards an empirically based typology of bilingual speech patterns". *International Journal of Bilingual Education and Bilingualism* 10: 298–340.

Dorian, Nancy. 1993. "Internally and externally motivated change in language contact settings: Doubts about dichotomy" in *Historical linguistics. Problems and perspectives,* C. Jones (ed.), 131–54. London: Longman.

Dorian, Nancy. 2006. "Negative borrowing in an indigenous language shift to the dominant national language". *International Journal of Bilingual Education and Bilingualism* 9: 557–77.

Duden. 1995. *Grammatik der deutschen Gegenwartssprache* (vol. 4), G. Drosdowski, K. Razum and P. Eisenberg (eds.). Mannheim: Bibliographisches Institut.

Dutkova-Cope, Lida 2001. "The language of Czech Moravians in Texas: Do you know what *Párknu Káru u Hauza* means?". *Southwest Journal of Linguistics* 20: 51–84.

Eppler, Eva. 2004. "...'because dem Computer brauchst' es ja nicht zeigen.': because + German main clause word order". *International Journal of Bilingualism* 8: 127–144.

Field, Fredric. 2005. "Long term effects of CS: Clues to structural borrowing". *International Journal of Bilingualism* 9: 341–60.

Galloway, Brent D. 1977 [1993]. *A grammar of Upriver Halkomelem*. Berkeley: University of California Press.

Gelderen, Elly van. 2009. "Language change and Survive: Feature economy in the numeration" in *Towards a derivational syntax: Survive-Minimalism*, M. Putnam (ed.), 257–266. Amsterdam: John Benjamins.

Grosjean, François. 1998. "Studying bilinguals: Methodological and conceptual issues". *Bilingualism: Language and cognition* 1: 131–49.

Grosjean, François. 2000. "Processing mixed language: Issues, findings and models" in *The Bilingualism Reader*, L. Wei (ed.), 443–69. London: Routledge.

Grosjean, François. 2001. "The bilingual's language modes" in *One mind, two languages*, J. Nicol (ed.), 1–22. Malden, MA: Blackwell.

Gumperz, John J., and Wilson, Robert. 1971. "Convergence and creolization: A case from the Indo-Aryan/Dravidian border" in *Pidginization and creolization of languages*, D. Hymes (ed.), 151–68. Cambridge: Cambridge University Press.

Heine, Bernd and Kuteva Tania. 2005. *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.

Herring, Jon, Deuchar, Margaret, Parafita Couto, M. Carmen and Moro Quintanilla, Mónica. In press. "'I saw the madre': evaluating predictions about codeswitched determiner-noun sequences using Spanish-English and Welsh-English data".

Hutz, Matthias. 2004. "Is there a natural process of decay? A longitudinal study of language attrition" in *First language attrition: Interdisciplinary perspectives on methodological issues*, M. Schmid, B.

Köpke, M. Keijzer and L. Weilemar (eds.),, 189-206. Amsterdam/Philadelphia: John Benjamins.

Jake, Janice L., Myers-Scotton, Carol and Gross, Steven. 2002. "Making a Minimalist approach to codeswitching work: Adding the Matrix Language". *Bilingualism: Language and Cognition* 5: 69–91.

Johanson, Lars. 2002a. *Structural factors in Turkic language contacts*. London: Routledge.

Johanson, Lars. 2002b. "Do languages die of 'structuritis'?: On the role of code-copying in language endangerment". *Rivista di Linguistica* 14, 249–270.

Jones, Bedwyr L. 1981. "Welsh: Linguistic conservation and shifting bilingualism" in *Minority languages today*, E. Haugen, J.D. McClure and D. Thomson (eds.), 40–51. Edinburgh: Edinburgh University Press.

Jones, Morris and Thomas, Alan R. 1977. *The Welsh language: Studies in its syntax and semantics*. Cardiff: University of Wales Press for the Schools Council.

Jones, Robert O. 1993. "The sociolinguistics of Welsh" in *The Celtic Languages*, M. Ball with J. Fife (eds.), 536–605. London: Routledge.

Joshi, Aravind K. 1985. "Processing of sentences with intrasentential code switching" in *Natural language parsing*, D.R. Dowty, L. Karttunen and A. Zwicky (eds.), 190–205. Cambridge: Cambridge University Press.

King, Gareth. 1996. *Modern Welsh: a comprehensive grammar*. London: Routledge.

Klavans, Judith L. 1983. "The syntax of code switching: Spanish and English" in *Proceedings of the Linguistic Symposium on Romance Languages,* W.W. Cressey and D.J. Napoli (eds.), 213–232. Philadelphia: John Benjamins.

Labov, William. 1966. *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Labov, William. 1990. "The intersection of sex and social class in the course of linguistic change". *Language Variation and Change* 2, 205-254.

Labov, William. 2001. *Principles of Linguistic Change: Social Factors*. Malden, MA: Blackwell.

Labov, William. 2002. "Driving forces in linguistic change". *Proceedings of the 2002 International Conference on Korean Linguistics*. Seoul: Hankwuk Mwunhwasa.

Lewis, M. Paul (ed.). 2009. *Ethnologue: Languages of the world* (16th ed.). Dallas, TX: SIL International online version (http://www.ethnologue.com).

Lewis, Saunders. 1962. *Tynged yr Iaith*. Radio lecture broadcast on the BBC, 13th February 1962.

Luescher, Adrian. 2008. Investigation into the application of the Matrix Language Frame model to German-English speech data. Bangor University M.A. dissertation.

Maclagan, Margaret A., Gordon, Elizabeth and Lewis, Gillian. 1999. "Women and sound change: Conservative and innovative behavior by the same speakers". Language variation and change 11, 19–41.

MacSwan, Jeff. 2000. "The architecture of the bilingual language faculty: evidence from intrasentential code switching." *Bilingualism: Language and Cognition*, 3: 37-54.

MacSwan, Jeff. 2005a. "Codeswitching and generative grammar: A critique of the MLF model and some remarks on 'modified minimalism'". *Bilingualism: Language and Cognition* 8, 1–22.

MacSwan, Jeff. 2005b. "Remarks on Jake, Myers-Scotton and Gross's response: There is no 'Matrix Language'". *Bilingualism: Language and Cognition* 8, 277–84.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*, 3rd edition. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Martin, Victoria C. 1997. Modelling Umgangssprache: The interaction between dialect and standard in Viennese German. University of Cambridge Ph.D. dissertation.

Matras, Yaron. 2000. "Mixed languages: A functional-communicative approach". *Bilingualism: Language and Cognition* 3: 79–99.

Matras, Yaron and Sakel, Jeanine (eds.). 2007. *Grammatical borrowing in cross-linguistic perspective*. Berlin: Mouton de Gruyter.

Maurer, Friedrich. 1926. *Untersuchungen über die Deutsche Verbstellung in ihrer geschichtlichen Entwicklung*. Heidelberg: Winter.

M'Barek, M. Na'it and Sankoff, David. 1988. "Le discours mixte arabe/français: des emprunts ou des alternances de langue?" *Revue canadienne de linguistique* 33: 143–154.

Milroy, Lesley. 1980. *Language and social networks*. Oxford: Blackwell.

Montrul, Silvina. 2004. "Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence". *Bilingualism: Language and Cognition* 7: 125–142.

Muysken, Pieter. 2000. *Bilingual speech: A typology of code mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 1993. *Duelling Languages*. Oxford: Clarendon Press.

Myers-Scotton, Carol. 1998. "A way to dusty death: the Matrix Language turnover hypothesis" in *Endangered languages: Language loss and community response*, L.A. Grenoble and L.J. Whaley (eds.), 289–316. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Myers-Scotton, Carol. 2006. *Multiple voices: An introduction to bilingualism*. Malden, MA: Blackwell Publishing.

Myers-Scotton, Carol. 2008. "Language contact: Why outsider system morphemes resist transfer". *Journal of language contact* Thema 2: 21–41.

Nicol, Janet. 2001. *One mind, two languages*. Malden, MA: Blackwell.

Nishimura, Miwa. 1986. "Intrasentential code switching: The case of language assignment" in *Bilinguals: Psycholinguistic and neuropsychological perspectives*, J. Vaid (ed.), 123–43. Hillsdale, NJ: Erlbaum.

Pakendorf, Brigitte. 2007. Contact in the prehistory of the Sakha (Yakuts): Linguistic and genetic perspectives. University of Leiden PhD dissertation.

Patocka, Franz. 1997. *Satzgliedstellung in den bairischen Dialekten Österreichs.* Frankfurt: Peter Lang.

Paulasto, Heli. 2006. *Welsh English syntax: Contact and variation.* Joensuu: University of Joensuu.

Pavlenko, Aneta. 2004. "L2 influence and L1 attrition in adult bilingualism" in *First language attrition: Interdisciplinary perspectives on methodological issues*, M. Schmid, B. Köpke, M. Keijzer and L. Weilemar (eds.), 47–60. Amsterdam/Philadelphia: John Benjamins.

Penhallurick, Robert. 2004. "Welsh English: Morphology and syntax" in *A handbook of varieties of English 2: Morphology and syntax*, B. Kortmann, K. Burridge, R. Mesthrie, E.W. Schneider and C. Upton (eds.), 102–113. Berlin: Mouton de Gruyter.

Phillips, John D. 2007. "Mae nodweddion hynotaf y Gymraeg ar ddiflannu". Yamaguchi University journal.

Poplack, Shana. 1980. "'Sometimes I'll start a sentence in Spanish y termino en español': toward a typology of code-switching" in *The Bilingualism* Reader, L. Wei (ed.), 221–56. London/New York: Routledge. Page numbers used in this thesis pertain to the reprinted article in *The bilingualism reader*, L. Wei (ed.), 205–40. New York: Routledge.

Poplack, Shana. 1987. "Contrasting patterns of code-switching in two communities" in *Aspects of Multilingualism: Proceedings from the fourth*

Nordic symposium on bilingualism, 1984, E. Wande, J. Anward, B. Nordberg, L. Steensland and M. Thelander (eds.), 51–77. Uppsala: Borgströms.

Poplack, Shana and Walker, James A. 2003. "Bilingual speech: a typology of code-mixing" (book review). *Journal of Linguistics* 39: 678–683.

Poplack, Shana, Walker, James A. and Malcolmson, Rebecca. 2006. "An English 'like no other'?: Language contact and change in Quebec". *Canadian Journal of Linguistics* 51, 185–213.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey and Svartvik, Jan. 1989. *A comprehensive grammar of the English language*. London/New York: Longman.

Rista-Dema, Mimoza. 2008. "Verb-subject word order in Albanian English". *World Englishes* 27: 419–33.

Roberts, Ian. 2005. *Principles and Parameters in a VSO language: A case study in Welsh.* Oxford/New York: Oxford University Press.

Roberts, Anna E. 1988. "Age-related variation in the Welsh dialect of Pwllheli" in *The use of Welsh: A contribution to sociolinguistics*, M.J. Ball (ed.), 104–22. Clevedon: Multilingual Matters.

Rowland, Thomas. 1853 [1876]. *A grammar of the Welsh language: Based on the most approved systems, with copious examples from some of the most correct Welsh writers*. Holywell: W. Morris, 'Cymro' and 'Eglwysydd' Office.

Russ, Charles V.J. (ed.). 1990. *The dialects of modern German*. London: Routledge.

Sanchez, Liliana. 2004. "Functional convergence in the tense, evidentiality and aspectual systems of Quechua Spanish bilinguals". *Bilingualism: Language and Cognition* 7: 147–62.

Sandalo, Filomena. 1995. *A grammar of Kadiwéu*. University of Pittsburgh Ph.D. dissertation.

Sankoff, David. 1988. "Problems of representativeness" in *Sociolinguistics: An International handbook of the Science of Language and Society*, U. Ammon, N. Dittmar and K.J. Mattheier (eds.), 899–903. Berlin: Walter de Gruyter.

Sapp, Christopher D. 2006. Verb order in subordinate clauses from Early New High German to Modern German. Indiana University PhD dissertation.

Sasse, Hans-Jürgen. 1992a. "Theory of language death" in *Language death: Factual and theoretical explorations with special reference to East Africa*, M. Brenzinger (ed.), 7–30. Berlin: Walter de Grutyer.

Sasse, Hans-Jürgen. 1992b. "Language decay and contact-induced change: Similarities and differences" in *Language death: Factual and theoretical explorations with special reference to East Africa*, M. Brenzinger (ed.), 59–80. Berlin: Walter de Grutyer.

Schmid, Monika S. 2002. *First language attrition, use and maintenance: The case of German Jews in Anglophone countries*. Amsterdam/Philadelphia: John Benjamins.

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

Sebba, Mark. 1998. "A congruence approach to the syntax of codeswitching". *International Journal of Bilingualism* 2: 1–19.

Shaw, Patricia A., Campbell, Jill and Grant, Larry. 2008. "Word order change in Central Salish". Paper given at Continuity and Change in Grammar, University of Cambridge, March 20th 2008.

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.

Simpson, John A. and Weiner, Edmund S.C. (eds.). 1989. *Oxford English dictionary*. Oxford: Oxford University Press.

Smith, Daniel J. 2006. "Thresholds leading to shift: Spanish/English codeswitching and convergence in Georgia, U.S.A.". *International Journal of Bilingualism* 10: 207–240.

Spurrell, William. 1870. *Grammadeg o laith y Cymry/A grammar of the Welsh language*. Carmarthen: W. Spurrell.

Tagliamonte, Sali A. 1998. *"'Was/Were' variation across the generations: View from the City of York." Language Variation and Change* 10: 153–92.

Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Thomas, Alan R. 1982. "Change and decay in language". In *Linguistic controversies: Essays in linguistic theory and practice in honour of F.R. Palmer*, D. Crystal (ed.), 209–219. London: Edward Arnold Publishers.

Thomas, Alan R. 1985. "Welsh English: A grammatical conspectus". In *Varieties of English around the world: Focus on: England and Wales*, W. Viereck (ed.), 213–222. Amsterdam/Philadelphia: John Benjamins.

Thomas, R.J. (ed.). 2003. *Geiriadur Prifysgol Cymru: A dictionary of the Welsh language* (2nd ed.). Cardiff: University of Wales Press.

Thomason, Sarah G. 2001. *Language contact: an introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah G. 2003. "Contact as a source of language change" In *Handbook of Historical Linguistics*, B.D. Joseph and R.D. Janda (eds.), 687–712. Malden, MA: Blackwell.

Thomason, Sarah G. and Kaufman, Terence. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.

Thorne, David A. 1996. *Gramadeg Cymraeg*. Llandysul: Gomer Press.

Thorne, David A. 2000. *Gafael mewn gramadeg*. Llandysul: Gomer Press.

Toribio, Almeida J. 2004. "Convergence as an optimization strategy in bilingual speech: Evidence from code-switching". *Bilingualism: Language and Cognition* 7: 165–173.

Treffers-Daller, Jeanine. 1994. *Mixing two languages: French-Dutch contact in a comparative perspective.* Berlin: Walter de Grutyer.

Treffers-Daller, Jeanine. 2005. "Evidence for insertional codemixing: Mixed compounds and French nominal groups in Brussels Dutch". *International Journal of Bilingualism* 9: 477–508.

Treffers-Daller, Jeanine and Mougeon, Raymond. 2005. "The role of transfer in language variation and change: Evidence from contact varieties of French". *Bilingualism: Language and Cognition* 8: 93–8.

Uned Iaith Genedlaethol Cymru [National Welsh Language Unit]. 1976. *Gramadeg Cymraeg gyfoes / Contemporary Welsh grammar*. Cowbridge: D. Brown & Sons.

Waas, Margit. 1996. *Language attrition Downunder: German speakers in Australia*. Frankfurt: Peter Lang.

Weinreich, Uriel. 1953. *Languages in contact: Findings and problems*. New York: Linguistic Circle of New York.

Williams, Stephen J. 1980. *A Welsh grammar*. Cardiff: University of Wales Press.

Williams, Malcolm. 2000. "The pragmatics of preciate fronting in Welsh English" in The Celtic Englishes II, H.L.C. Tristram (ed.), 210–30. Heidelberg: Universitätsverlag C. Winter.

Willis, David. 2008. "English-Welsh language contact in the realm of negation". Paper presented at Transmission and Diffusion Congress, Radbod University Nijmegen/Max Planck Institute for Psycholinguistics, 17th January 2008.

Wiltschko, Martina. 2000. "The syntax of transitivity and its effects: Evidence from Halkomelem Salish" in *WCCFL 20: Proceedings of the 20th West Coast conference on formal linguistics*, K. Megerdoomian and L. Bar-el (eds.). Somerville, MA: Cascadilla Press.

Wiltschko, Martina. 2002. "The syntax of pronouns: evidence from Halkomelem Salish". *Natural Language & Linguistic Theory* 20: 157–195.

Winford, Donald. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

Winford, Donald. 2005. "Contact-induced changes: Classification and processes". *Diachronica* 22: 373–427.

Winford, Donald. 2007. "Some issues in the study of language contact". *Journal of Language Contact* thema 1: 22–39.