CONSTRAINTS ON AUXILIARY DELETION IN COLLOQUIAL WELSH

Florian Breit
CONSTRAINTS ON AUXILIARY DELETION IN COLLOQUIAL WELSH

A dissertation submitted in partial fulfilment of the requirements for the degree of BA (Hons) in Linguistics

By
Florian Breit
School of Linguistics & English Language
Bangor University

Submitted on
21st May 2012
Do not be too timid and squeamish about your actions. All life is an experiment. The more experiments you make the better.

Ralph Waldo Emerson
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>ii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>v</td>
</tr>
<tr>
<td>List of Program Code Listings</td>
<td>vi</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>ix</td>
</tr>
<tr>
<td>Declaration</td>
<td>xi</td>
</tr>
<tr>
<td>1  Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2  Existing Literature on Auxiliary Deletion</td>
<td>4</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>4</td>
</tr>
<tr>
<td>2.2 Auxiliaries and Auxiliary-initial Clauses in Colloquial Welsh</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Previous Descriptions of Auxiliary Deletion in Welsh</td>
<td>10</td>
</tr>
<tr>
<td>2.4 Auxiliary Deletion and Grammatical Person</td>
<td>13</td>
</tr>
<tr>
<td>2.5 Summary</td>
<td>18</td>
</tr>
<tr>
<td>3  Auxiliary Deletion in the <em>Siarad</em> Corpus</td>
<td>21</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>21</td>
</tr>
<tr>
<td>3.2 Methodology</td>
<td>22</td>
</tr>
<tr>
<td>3.3 Results &amp; Discussion</td>
<td>25</td>
</tr>
<tr>
<td>3.4 Summary</td>
<td>29</td>
</tr>
<tr>
<td>4  Sentence Judgements on Auxiliary Deletion</td>
<td>30</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>30</td>
</tr>
<tr>
<td>4.2 Methodology</td>
<td>35</td>
</tr>
<tr>
<td>4.3 Results &amp; Discussion</td>
<td>40</td>
</tr>
</tbody>
</table>
CONTENTS

4.4 Summary ................................................. 48

5 Discussion: Possible Constraints on Auxiliary Deletion 50

6 Conclusions 55

References 58

Appendix 61
   A Program Code Listings for Corpus Study ................. 61
   B Stimuli for Judgement Experiment ........................ 80
   C Program Code Listings for Judgement Experiment ........ 86
   D Instructions for Judgement Experiment ................. 108
      Instructions for Online Task .......................... 108
      Instructions for Offline Task .......................... 109
   E Sample of Offline Judgement Questionnaire ............. 110
   F Poster for Advertising Judgement Experiment .......... 111
   G Consent Form for Judgement Experiment ............... 112
List of Tables

2.1 Paradigm of *bod* in a northern variety of colloquial Welsh. . . . . 7
2.2 Paradigm of *gwneud* in a northern variety of colloquial Welsh. . 8
2.3 Comparison of AD grammaticality by grammatical person between
    Borsley et al. (2007) and Jones (2004). . . . . . . . . . . . . . . 16
3.1 Summary of instances of AD extract and how many speakers
    produced them by grammatical person and number. . . . . . . . . 26
3.2 Results on AD and grammatical person/number in Siarad Cor-
    pus compared to predictions from Borsley et al. (2007) and Jones
    (2004). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28
4.1 Means and standard deviations for constructions which tested
    AD acceptability with different direct subjects . . . . . . . . . 44
4.2 Means and standard deviations for AD acceptability with tense
    and aspect . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
4.3 Means and standard deviations for AD acceptability dependent
    on mood . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
4.4 Means and standard deviations for AD in different constructions
    with non-default surface structure . . . . . . . . . . . . . . . . 47
4.5 Means and standard deviations for AD in subordinate clauses . . 47
4.6 Means and standard deviations for AD in Wh-questions . . . . 48
B.1 Training Stimuli for Judgement Experiment . . . . . . . . . . . . 80
B.2 Test Stimuli for Judgement Experiment . . . . . . . . . . . . . . 80
List of Figures

3.1 Distribution of AD by Grammatical Person and Number . . . . . 27
4.1 Mean Online and Offline Responses Compared . . . . . . . . . . . 43
List of Program Code Listings

1 Script for Autoglossing entire Siarad corpus . . . . . . . . . 61
2 Script for finding AD in autoglossed corpus . . . . . . . . . 62
3 OpenSesame script for judgement experiment . . . . . . . . 86
4 Script for calculating duration of waveform files . . . . . . 91
5 Script for generating offline questionnaire . . . . . . . . . . 93
6 Sample format of stimuli.php file . . . . . . . . . . . . . . . 98
7 Script for entering background data . . . . . . . . . . . . . . 99
8 Script for entering questionnaire results . . . . . . . . . . . 100
9 Script for merging results from online and offline tasks . . . 102
List of Abbreviations

1P  First person plural
1S  First person singular
2P  Second person plural
2S  Second person singular
3P  Third person plural
3S  Third person singular
3Sm/3Sf  Third person singular male/female
AAVE  African American Vernacular English
AD  Auxiliary Deletion
AFF  Affirmative mood
Aux  Auxiliary verb
CSV  Comma separated values (file format)
D/Det  Determiner
FUT  Future tense
HTML  Hypertext Markup Language (file format)
IMPFV  Imperfective aspect
INFL  Inflection
INT  Interrogative mood
NEG  Negation, Negative mood
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>Noun Phrase</td>
</tr>
<tr>
<td>O</td>
<td>Object</td>
</tr>
<tr>
<td>PAST</td>
<td>Past tense</td>
</tr>
<tr>
<td>PD</td>
<td>Particle Deletion</td>
</tr>
<tr>
<td>PFV</td>
<td>Perfective aspect</td>
</tr>
<tr>
<td>PHP</td>
<td>PHP: Hypertext Preprocessor (scripting language)</td>
</tr>
<tr>
<td>PN</td>
<td>Proper Noun</td>
</tr>
<tr>
<td>POS</td>
<td>Positive</td>
</tr>
<tr>
<td>Prep</td>
<td>Preposition</td>
</tr>
<tr>
<td>PRES</td>
<td>Present tense</td>
</tr>
<tr>
<td>PRT</td>
<td>Particle</td>
</tr>
<tr>
<td>Q</td>
<td>Question particle</td>
</tr>
<tr>
<td>S</td>
<td>Subject</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>V</td>
<td>Verb</td>
</tr>
<tr>
<td>V1</td>
<td>Verb first</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to express my thankfulness to the staff at the School of Linguistics & English Language, who at large have offered me many an open ear and ample platform for discussion often far beyond the curriculum. No doubt the changes the department went through while I studied for this degree have been dramatic, and indeed traumatic at times, but they have also been a constant provider of opportunity to see things from different perspectives, which at this early stage in a hopeful linguist’s career has been most useful. Special thanks here foremost go to my personal tutor, Peredur Davies, who has been outstanding and inspiring in every aspect and who is ultimately responsible for getting me hooked on the topic discussed in this dissertation (diolch am yr holl bysgod!), but also Dirk Bury, Marco Tamburelli and Paul Carter for their open doors, insightful discussion of theory and inspiration.

Specifically in creation of this work I would also like to thank my supervisor, Margaret Deuchar, who has provided many useful resources in the initial stages of research for my dissertation and to the dissertation co-ordinator, Vicky Chondrogianni, who has readily offered help with my dissertation when it was needed. Further thanks go to my fellow aspiring linguist Rhian Davies, who has been very patient with my attempts at learning her native language and always lent her expertise as a native informant, for this dissertation and other work prepared during the degree course. Diolch yn fawr!

I have to also extend my gratefulness to all my other Bangor friends
and acquaintances, my Welsh teachers at Lifelong Learning, and to Bangor Linguistics Society, who have all helped pass time and worries all along the way and made these last three years a great experience. Thank you!

Last but not least I have to thank my parents, without whose support – both emotionally and financially – all this would not have been possible and for whom my respect and admiration has grown in proportion to distance from home: Merci viilmols Mama un Papa!
Declaration

I hereby declare that this dissertation is my own work in partial fulfilment of requirements for *BA (Hons) Linguistics*. 

Florian Breit

Bangor, Gwynedd

21st May 2012
1 Introduction

Some recent research on the grammatical properties of the spoken varieties of modern colloquial Welsh have described a phenomenon known as Auxiliary Deletion (AD), through which some clause-initial auxiliaries in these varieties can be omitted. This phenomenon has as of yet not received much individual attention apart from Jones (2004) and Davies (2010), where AD is the subject of a large part of their overall studies; Davies and Deuchar (in preparation) are further currently preparing a somewhat more elaborated version of the investigation into AD presented in Davies (2010).

Welsh being a verb-initial language, auxiliary constructions in Welsh are traditionally associated with periphrastic constructions of the type AuxSVO, which stand in contrast to so called synthetic constructions of the type VSO, with a finite initial verb. This type of construction is the one primarily employed in colloquial varieties of Welsh and stands in relation to colloquial Welsh exhibiting only limited final verb morphology for verbs other than auxiliaries, especially so with the present tense (for more on periphrastic and synthetic constructions see Borsley et al., 2007). Further, Davies (2010, p. 285) has shown that in his corpus analysis of AD with the second person singular pronoun ti, 92.75% of the sentences analysed featured AD, suggest-
CHAPTER 1. INTRODUCTION

ing that AD in these constructions is the norm in spoken colloquial Welsh and it can consequently be argued that this phenomenon clearly deserves some more in-depth treatment than it has so-far received.

While Jones (2004) limits himself to offering his observations and intuitions on what grammatical person AD may occur with, Davies (2010) focuses only on the second person singular, for which he has performed a corpus analysis as part of his PhD study. In the second chapter of this dissertation I will look at the previous literature on AD, focusing on the studies by Jones (2004), Davies (2010) and Davies and Deuchar (in preparation). Following this, in the second chapter I undertake to test Jones’ (2004) and Borsley et al.’s (2007) proposed limitations on grammatical person and AD by carrying out a corpus study on the conversational Welsh Siarad corpus that has been collected at Bangor University over the last years (Deuchar et al., 2009; see also Davies, 2010, pp. 150–178 for a description of the methodology applied in creating the corpus). On the basis of this I will then present an experimental study testing the acceptability of AD in some of the grammatical configurations in which Welsh auxiliaries can be found (e.g. clause-initial, Wh-questions with movement and in-situ, with a proposed subject/object, in coordinated clauses, etc.) in the third chapter. From this it is hoped to gain some insight into the possible constraints that may apply to AD beyond that of grammatical person or number proposed by Jones (2004), which I will discuss in chapter four. This will be followed by my conclusion in the final chapter. It is believed that identification of such constraints can form the basis of further investigation offering greater insights on the processes underlying AD in Welsh. As such the aim of this
dissertations is exploration: to establish some basic information about the syntactic conditions under which AD may occur, in order to lay the ground work for some further theoretical work tackling this phenomenon.
2 Existing Literature on Auxiliary Deletion

2.1 Introduction

Auxiliary Deletion in Welsh has generally not been much discussed before. As mentioned already in the introduction, the two principal sources of description until now are from Jones (2004) and Davies (2010), though the phenomenon itself has previously and since been variously acknowledged (e.g. King, 1996; Borsley et al., 2007).

Davies' (2010) study primarily looks at AD as an indicator of contact-induced language-change. In doing so, he also makes reference to a previous explicit description of AD from Phillips (2007), written in Welsh, in which Phillips stipulates that AD may indicate a shift in word order from VSO/AuxSVO to SVO, an analysis which Davies (2010) agreed with, arguing that this is due to convergence towards English SVO word order stemming from the extensive bilingualism present in Wales\(^1\). They posit

---

\(^1\)There are normally believed to be no monolingual Welsh speakers in Wales above the age of 3 years any more. This can be partially attributed to the nature of bilingualism and presence of English in the media, which exposes speakers to at least some English
that AD occurs due to language contact between Welsh and English and reflects language change in Welsh, where Welsh gradually tends to assume a surface structure similar to that of English SVO, thus leading to a preference in AuxSVO structures with deleted auxiliaries which resemble these. As the relations of AD to language contact and language change are not of much interest apart from the fact and extent to which AD appears to presently occur in Welsh, I will not discuss this matter much further in this chapter, though I will draw on Davies (2010) and Davies and Deuchar (in preparation) in section 2.3, where following an overview of the Welsh auxiliary system in general in section 2.2, I further discuss the extent to which AD is known to currently occur in Welsh and give some real examples of AD in Welsh.

Other than Davies (2010) and Phillips (2007), Jones (2004) and Borsley et al. (2007) do not present any analysis of real data containing AD but instead limit themselves to stating for which grammatical persons AD may occur. This will be further discussed in section 2.4 and later forms the basis for my corpus study in chapter 3, in which I test their intuitions on the correlation of AD and grammatical number.

Similar phenomena to AD in Welsh are also known to occur in other languages, for instance in African American Vernacular English (AAVE) and in Central Salish as already referred to by Davies (2010). However anywhere in Wales, but is especially due to English being part of the National Curriculum for Wales (see DfES, 2008) (note however that Key Stage 1 in Welsh-medium schools is exempt from this requirement, see ACCAC, 2000, p. 2). However, there is no actual comprehensive data on this available, and while the UK Census in Wales features a question on Welsh language ability, it does not contain any questions to record an ability to speak English, so that there is no potential to draw data on monolingual Welsh speakers from this.
CHAPTER 2. EXISTING LITERATURE ON AUXILIARY DELETION

while noting that work on AD in these languages is also largely of a broad descriptive type, I will not discuss these phenomena in detail here due to space constraints.

Section 2.5 presents a summary of the literature reviewed in this chapter.

2.2 Auxiliaries and Auxiliary-initial Clauses in Colloquial Welsh

As a verb-initial language Welsh sentence structures are often divided into two basic constructions: the synthetic construction and the periphrastic construction. Synthetic constructions are of the type VSO with an initial finite verb. Periphrastic constructions are of the basic type AuxSVO with an initial auxiliary and an infinitive verb after the subject, although auxiliaries derived from bod ‘to be’ also require use of an aspectual particle which is placed between the subject and its following infinitive verb (Borsley et al., 2007, pp. 38–47). While literary Welsh makes extensive use of the synthetic construction in all tenses, in colloquial Welsh constructions of the periphrastic type are much more common. One reason for this is that in either variety of Welsh there is no specific inflectional paradigm for the present tense apart from that for bod ‘to be’. Instead there is one inflectional paradigm which in formal Welsh represents both the present and the future tense, but which is only used for the future tense in colloquial Welsh. In colloquial Welsh the periphrastic construction with a present tense paradigm for the auxiliary bod (in conjunction with an aspectual particle such as the
CHAPTER 2. EXISTING LITERATURE ON AUXILIARY DELETION

Table 2.1: Paradigm of *bod* in a northern variety of colloquial Welsh.

<table>
<thead>
<tr>
<th></th>
<th>future</th>
<th>present</th>
<th>past</th>
</tr>
</thead>
<tbody>
<tr>
<td>1S</td>
<td>bydda</td>
<td>dw</td>
<td>ôn</td>
</tr>
<tr>
<td>2S</td>
<td>byddi</td>
<td>wyt</td>
<td>oeddet</td>
</tr>
<tr>
<td>3S</td>
<td>bydd</td>
<td>ydy/mae</td>
<td>oedd</td>
</tr>
<tr>
<td>1P</td>
<td>byddwn</td>
<td>dan</td>
<td>oeddwn</td>
</tr>
<tr>
<td>2P</td>
<td>byddwch</td>
<td>dach</td>
<td>oeddych</td>
</tr>
<tr>
<td>3P</td>
<td>byddwn</td>
<td>ydyn/maen</td>
<td>oedden</td>
</tr>
</tbody>
</table>

imperfective *yn* or the perfective *wedi*) is used to mark the present tense instead (cf. Borsley et al., 2007, 9–12). Additionally to the present tense\(^2\), in colloquial Welsh the periphrastic construction is very common for past tense statements, often with the auxiliary *gwneud* ‘to do’.

Borsley et al. (2007, p. 38) themselves define auxiliaries as “certain verbal elements which appear with a verbal complement of some kind and allow the expression of a meaning which would be expressed by a single verb in some languages.” As such their definition of auxiliaries itself relies heavily on the above distinction between synthetic and periphrastic constructions, and it is probable that as such any initial tensed verb in a periphrastic construction is an auxiliary (see also Borsley et al., 2007, pp. 44–47 for a further discussion of possible syntactic tests for auxiliary status). Following from this, Borsley et al. (2007, p. 38) describe three types of auxiliary-initial clauses: aspectual clauses (i.e. those headed by the auxiliary *bod* and containing an aspect marker), *gwneud*-clauses and *ddaru*-clauses.

\(^2\)Note that with the use of aspect this covers both what equates in English to the present tense and the simple past. For instance both “I eat” and “I am eating” can be expressed as the imperfective present tense statement “Dw i’n bwyta” and “I ate” can
Table 2.2: Paradigm of *gwneud* in a northern variety of colloquial Welsh.

<table>
<thead>
<tr>
<th></th>
<th>future</th>
<th>past</th>
</tr>
</thead>
<tbody>
<tr>
<td>1S</td>
<td>wna</td>
<td>wnes</td>
</tr>
<tr>
<td>2S</td>
<td>wnei</td>
<td>wnest</td>
</tr>
<tr>
<td>3S</td>
<td>wneith</td>
<td>wnaeth</td>
</tr>
<tr>
<td>1P</td>
<td>wnawn</td>
<td>wnaethon</td>
</tr>
<tr>
<td>2P</td>
<td>wnewch</td>
<td>wnaethoch</td>
</tr>
<tr>
<td>3P</td>
<td>wnân</td>
<td>wnaethon</td>
</tr>
</tbody>
</table>

Tables 2.1 and 2.2 illustrate northern colloquial Welsh paradigms of *bod* ‘to be’ and *gwneud* ‘to do’ respectively. Note that the different forms *ydy/mae* and *ydyn/maen* are due to mood, and not due to tense. The other auxiliary discussed by Borsley et al. (2007), *ddaru*, does in colloquial Welsh not have any other paradigm than its past tense “ddaru” itself and accordingly is used solely as a past tense marker (and usually transcribed as such), it also does not show agreement for number and appears to be confined to northern varieties of Welsh (Borsley et al., 2007, p. 38).

Auxiliary-initial clauses then can be formed with any of these three auxiliaries, following their respective paradigms and restrictions. Examples are given in (2–1) to (2–3) below:

(2–1) *Dw i’n licio chwarae sboncen.*

be.1S.PRES I IMPFV like play squash

‘I like playing squash.’

be expressed using the perfective statement “Dw i wedi bwyta”, in all three cases “dw” is the first person present tense inflection of *bod* ‘to be’. For a brief discussion of tense in colloquial Welsh see Borsley et al. (2007, pp. 9–10).

Similarly, forms of *bod* beginning in a vowel may be prefixed with *d-* to mark them for negative mood and *r-* to mark them for affirmative mood, *mae/maen* are the affirmative mood equivalents of *ydy/ydyn.*
CHAPTER 2. EXISTING LITERATURE ON AUXILIARY DELETION

(2–2) Wneith y parot dweud ‘ta ra’ yn fuan.
do.3S.FUT the parrot say bye bye PRT soon
‘The parrot will say “bye bye” soon.’

(2–3) Ddaru Rhian sgwennu ’r aseiniad ddoe.
PAST Rhian write the assignment yesterday
‘Rhian wrote the assignment yesterday.’

As can be seen, the bod-derived example in (2–1) includes the imperfective aspect marker yn, while neither the gwneud-clause in (2–2) nor the ddaru-clause in (2–3) include such a particle. It can also be seen how the following verbs, chwarae, dweud and sgwennu are all present in their infinitive form, tense having been marked by the initial auxiliary.

Simple interrogatives using periphrastic constructions have the same surface structure, but differ mainly in intonation, while wh-questions make use of preposed question particles. Similarly, almost any item in such a clause can be fronted in order to focus it. An example of a simple interrogative is given in (2–4), an example of a wh-question is presented in (2–5) and (2–6) shows a construction with a focused subject.

(2–4) Wyt ti ’n licio chwarae sboncen?
be.2S.PRES you IMPFV like play squash
‘Do you like playing squash?’

(2–5) Be’ wyt ti ’n neud?
What be.2S you IMPFV do
‘What are you doing?’

(2–6) Rhian ddaru sgwennu ’r aseiniad.
Rhian.FOC PAST write the assignment
‘It was Rhian who wrote the assignment.’
Further, with negative statements there is an additional intervening negation particle *ddim* present, which follows the subject and in case of aspectual clauses precedes the aspect marker, as is illustrated in (2–7) and (2–8) below:

(2–7)  
\[ Wna \ i \ ddim \ dy \ helpu. \]
\[ \text{do.1S.FUT I NEG 2S help} \]
\[ ‘I won’t help you.’ \]

(2–8)  
\[ Dydy \ o \ ddim \ yn \ gweithio \ ddoe. \]
\[ \text{be.3S.NEG he NEG IMPFV work yesterday} \]
\[ ‘He didn’t work yesterday.’ \]

In this section I have illustrated some of the basic characteristics of auxiliaries in Welsh and how they are used within sentences. In the next section I will discuss what auxiliary deletion is and how this has been previously described in the literature.

### 2.3 Previous Descriptions of Auxiliary Deletion in Welsh

Auxiliary deletion as a specific point of focus has so far been discussed very little, and Davies (2010) is the only source specifically targeting the phenomenon, though Jones (2004) and Phillips (2007) also provide partial descriptions of the phenomenon. While Jones (2004) focuses on a general description in which he limits himself to discussing as factors geographic distribution and grammatical person, Davies (2010) and Phillips (2007) describe actual data containing AD which they have collected. Davies (2010,
p. 264) describes AD simply as sentences without an overt initial finite verb, where he interprets this to have been replaced by a null form of the appropriate auxiliary. He refers to these sentences as –A, while he refers to sentences with overt auxiliaries as +A. Jones (2004) does not give a specific name to the phenomenon but notes that for some speakers, depending on their regional variety, some auxiliaries, depending on grammatical number, can be omitted. An example of AD from Davies (2010) is given in 2–9 below:

(2–9) ti ddim yn licio dreifio?
you NEG IMPFV like drive
‘You don’t like driving?’ (Davies, 2010, p. 267, my gloss)

From this the relation to clauses such as 2–1 can easily be seen, the only element missing being the auxiliary. It can also be assumed that this auxiliary would be one derived from bod, as the imperfective particle clearly indicates that this is an aspectual clause. In fact, most of the AD examples given by Davies (2010) have such an aspectual particle, though some clauses appear to also show what Davies (2010, pp. 316–322) calls particle deletion (PD), where the aspectual particle is deleted. This may happen either in conjunction with AD or not, but where it does, it of course poses a problem in inferring which auxiliary might have been deleted, and Davies (2010, p. 320) shows that a large proportion of PD clauses also feature AD (nearly 90%, though the reverse is not true).

Because of this presence mainly in aspectual clauses, Davies (2010) proposes that these sentences essentially feature a null variant of the bod-auxiliary, which are moreover by default interpreted as present tense. This
CHAPTER 2. EXISTING LITERATURE ON AUXILIARY DELETION

raises the question in the first instance, whether AD is restricted to *bod* or can occur with other forms of auxiliaries too. Davies’ (2010) study also looks only at the second person singular pronoun *ti* in conjunction with AD, so that it remains open whether his findings also hold for different grammatical persons, as for instance those proposed by Jones (2004) – an issue which will be further discussed in section 2.4. However, Davies (2010) also cites some data from Roberts (1988), which contains a wh-questions with wh-fronting and Davies (2010) himself finds a construction where AD follows a conjunction. Both examples are given in 2–10 and 2–11 below:

(2–10) lle *ti* ’n *mynd* *i*?
   where you IMPFV go to
   ‘Where are you going to?’ (from Roberts 1988, p. 199, reported in Davies 2010, p. 276, my gloss)

(2–11) *ond ti* *ddim wedi sylwi hynny*
   but you NEG PFV realise that
   ‘But you haven’t realised that.’ (Davies, 2010, p. 288, my gloss)

These show that AD does also appear to not be restricted just to simple auxiliary initial clauses, but also occurs in other constructions including other overt material before them or even involving movement of a constituent around them such as in 2–10.

Davies (2010) also shows that AD appears to be the norm in current spoken colloquial Welsh, as is represented by the *Siarad* corpus. He found that 92.75% of clauses analysed featured AD (Davies, 2010, p. 285). He also gives further analyses and breakdowns for individual speakers, age groups, geographical distribution and type of construction. From this it should mainly be noted that every speaker he analysed showed AD, with 9 doing
CHAPTER 2. EXISTING LITERATURE ON AUXILIARY DELETION

so 100% of the time, a further 15 at least 83.33% of the time and only four speakers below that, with a single minimum of 50% (Davies, 2010, pp. 291–292). His analysis by age shows that there is statistically significant age variation, with older speakers appearing to produce slightly less AD than other groups, but the whole range of variability is only about 10% (Davies, 2010, p. 297). Regarding regional origin of speakers, he concluded that his data indicated this had “no effect on frequency of [AD]” (Davies, 2010, p. 295). Taken together this shows that AD as a phenomenon itself is both widely distributed and highly common across the speaker population and should be seen as an essential part of the language’s present grammar.

While these corpus studies can show that AD is a wide-ranging phenomenon, which appears to occur in many different situations and with all speakers of modern colloquial Welsh, neither Davies (2010) nor Phillips (2007) make any specific claims about purported constraints on AD. However, some other linguists have given such constraints at least for grammatical person, presumably based on their own observations and intuitions, which I will discuss in the next section.

2.4 Auxiliary Deletion and Grammatical Person

Both Jones (2004) and Borsley et al. (2007) make brief mentions of AD and in doing so mainly focus on describing with which pronouns this occurs. As such Borsley et al. (2007, p. 260) describes the phenomenon as follows: “A
further notable feature of *bod* is that finite forms are sometimes omitted in clause-initial position in colloquial Welsh with certain pronominal subjects.”

This statement actually contains several constraints on AD: First, it may only occur in sentences that feature a clause-initial finite form of *bod*, i.e. with what was described as aspsectual clauses in section 2.2; secondly, AD can only occur with pronominal subjects, i.e. not with proper names or “things”; and thirdly, this only applies to a subset of pronominal subjects, i.e. there is a restriction in the grammatical person and number of the pronominal subject the auxiliary agrees with.

Borsley et al. (2007) qualify their restriction to *bod* by establishing that these sentences may contain a *bod*-derived auxiliary in tag-questions, regardless of whether there was an overt auxiliary or not in the main clause. To exemplify this, Borsley et al. (2007) give the two examples repeated as 2–12a and 2–12b below:

(2–12) (a) *Rwyt ti ‘n mynd, ynd wyt?*

‘You are going, aren’t you?’ (Borsley et al., 2007, p. 261, my gloss)

(b) *Ti ‘n mynd, ynd wyt?*

‘You are going, aren’t you?’ (Borsley et al., 2007, p. 261, my gloss)

They demonstrate by this analogous construction that an initial auxiliary can be assumed, as with the tests on auxiliary status they discuss in Borsley et al. (2007, pp. 44–47). Additionally this supports their assumption that
clauses such as (2–12b) are not actually without an auxiliary but rather that this is a null form of \textit{bod}.

Regarding the restriction on pronominal subjects, Borsley et al. (2007) provide little further discussion, but they state that this “omission is particularly common with \textit{ti} ‘you.S’ but it also occurs with \textit{ni} ‘we’ and \textit{chi} ‘you.PL’ [... and] with \textit{fi} ‘I’ and \textit{nhw} ‘they’ in the speech of some speakers of southern dialects” (Borsley et al., 2007, p. 261). However, this is discussed a little more at length in Jones (2004), who initially notes it as a specific feature of \textit{wyt} that it can also be omitted but then notes that this would also work with the second person plural inflection (e.g. \textit{dach}) and (at least in southern dialects) the first person plural inflection (e.g. \textit{maen/ydyn}) (Jones, 2004, p. 101). He then goes on to state that in actuality, some speakers even show AD with the first person singular, where they would however use the form \textit{fi} rather than the usual clitic \textit{i} (Jones, 2004, p. 101). Note also that Jones (2004) makes no explicit mention of a condition which requires deleted auxiliaries to occur with pronominal subjects, though he notes that the form of the verb permitted to undergo AD is that in the copular construction, which potentially limits the scope of AD to omission of forms of \textit{bod} ‘to be’ and not the general Welsh auxiliary.

When it is assumed that the verb-noun in auxiliary clauses in Welsh forms the predicate of the sentence which is linked to by the auxiliary however, this would be the case with any type of the auxiliary clauses discussed in 2.2, not only in cases where there is no additional non-finite verb present (which is the case with the examples given by Jones, 2004).
Table 2.3: Comparison of AD grammaticality by grammatical person between Borsley et al. (2007) and Jones (2004).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1S</td>
<td>limited</td>
<td>limited</td>
</tr>
<tr>
<td>2S</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>3S</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>1P</td>
<td>yes</td>
<td>limited</td>
</tr>
<tr>
<td>2P</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>3P</td>
<td>limited</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 2.3 shows a summary and comparison of the conditions given by Borsley et al. (2007) and Jones (2004). From this it can be seen that while they largely agree on which grammatical person AD can occur with, they make different predictions for 1P and 3P, where Jones (2004) seems to apply greater restriction in that where Borsley et al. (2007) say that AD with 1P is acceptable, Jones (2004) says this is only the case for some speakers, and where Borsley et al. (2007) say that AD is acceptable with 3P only for some speakers, Jones (2004) does not state that this is the case, leaving the assumption that this would from his account be unacceptable. This has also been assumed with 3S in both cases, which is not mentioned in either accounts as permissible. It should also be noted that 2S and 2P are then the only cases where both accounts would lead to the prediction that these are generally acceptable.

Despite the question over which account, if either, makes more accurate predictions in an actual experimental study, Jones (2004) also makes no mention of a condition whereby AD would be grammatical only together
with a pronominal subject. However, Welsh finite verbs do not show agreement in person with non-pronominal subjects but instead default to the third person inflection of the finite verb, so that for instance in both “the child is” (singular non-pronominal subject) and “the children are” (plural non-pronominal subject) the form of bod ‘to be’ used would be the third person inflection, e.g. mae in a positive statement. Since both Borsley et al. (2007) and Jones (2004) have ruled out AD with 3S, one could then predict that AD with non-pronominal subjects is equally unacceptable. Jones’ (2004) version does however not make explicit any requirement for pronominal subjects, so that, should this prediction of 3S’s unacceptability not show to be borne out, and if he is otherwise correct, non-pronominal subjects may actually also be acceptable.

In the previous section it has been shown that AD is not only restricted to strict surface V1 position of the auxiliary that is deleted, but may also be possible in conjunction with wh-movement and conjunctions. In this section it was shown that some constraints on AD have however been posited in regards to what type of subject it may occur with and also which grammatical person it may occur with. In the next section, I will give a short summary of the literature that was discussed in this chapter and the implications that follow for the studies presented in this dissertation.

---

4Note that Davies (2010) does in fact find some AD clauses with non-pronominal subjects, such as “Lily’n byw ’da Kristen” (Davies, 2010, p. 270).
2.5 Summary

In this chapter I have first given a brief outline of auxiliaries and auxiliary clauses in colloquial Welsh. In this it was illustrated how auxiliaries can be derived from different verb stems and how those derived from *bod* feature additional aspeccual content. It was also noted that these constructions are fairly flexible and as such while their default word order is AuxSVO, most constituents may be fronted for focus.

Section 2.3 then gave a broader outline and example of AD in Welsh and discussed some corpus based studies of AD which demonstrated how widely spread the phenomenon appears to be and that at least for the second person singular it was so common that it can quite possibly even be considered unmarked in comparison to its overt counterpart. It was also argued that the data presented through these studies shows that AD appears to occur together with some other basic configurations where the auxiliary is not the first overt item, such as with wh-fronting and after conjunctions.

In the following section, 2.4, it was then shown that some other linguists have in their outlines of AD posited different constraints to do with the type of subject AD may occur with, principally pronominal subjects, and the grammatical person of the auxiliary/subject present. It was mainly concluded that AD should be highly acceptable with 2S and 2P, not with 3S and that it is of questionable status with other grammatical persons, depending on the account followed.

Following these descriptions of AD, it is clear that grammatical con-
CHAPTER 2. EXISTING LITERATURE ON AUXILIARY DELETION

Constraints on AD have not been empirically investigated and all that is present are descriptive corpus studies giving positive examples of AD and short mentions of AD in other places which posit some constraints but are probably based on the author’s intuitions and are not in agreement across authors, so that for a better understanding of the phenomenon, grammatical constraints on AD need to be further investigated.

In relating these constraints with corpus studies such as those conducted by Davies (2010) and Phillips (2007) however, it must be noted that corpus studies are unlikely to shed much light on actual grammatical constraints on the use of AD. In the case of deciding which auxiliary may have been deleted, this is simply because of the complication with AD discussed in section 2.3, which at least for clear present tense interpretable sentences does not offer any reliable source for determining which auxiliary would have been used had it been overt. It also has to be noted that while corpus studies can clearly show that some constructions are used regularly and should thus be considered part of the grammar, it is as a method never able to exclude constructions from the grammar of its language as it will never contain negative examples. This is in analogy to some of the consequences of Zipf’s law, which states that there is a linear distribution of word frequencies for all the words in a given corpus along a constant defined by the relationship of a word’s frequency and it’s ranking in relation to other frequencies in the corpus. This leads to the prediction that the majority of words in a corpus are relatively rare or infrequent, with some occurring only once (Manning and Schütze, 1999, pp. 23–29). These non-recurring
CHAPTER 2. EXISTING LITERATURE ON AUXILIARY DELETION

words are known as hapax legomena\(^5\) and when it is assumed that Zipf's law holds not only for words but also collocations of these words, it is predicted that there will also be constructions which only occur marginally in a corpus or not at all: if there are predicted to be hapax legomena for lexical items, it can be assumed that there are also hapax legomena on the basis of constructions. Occurrences of any type of construction in a corpus can then only be taken as a positive indicator for the grammatical possibility of the construction, but never as a defining corpus of all positive examples of their range\(^6\). It is for this reason that in order to determine the boundaries of acceptability surrounding a phenomenon such as AD, the collection of acceptability judgements, which is further discussed in chapter 4 where I conduct such a judgement experiment, appears to be one of the most appropriate methods for establishing further constraints on AD in Welsh.

\(^5\)Crystal (2008, p. 224) gives as a definition for *hapax legomenon* "a word which occurs only once in a text, author, or extant corpus of a language". This concept is here extended beyond the word level to that of an entire construction that only occurs once in a corpus of text.

\(^6\)This is similar to the argument of the poverty of the stimulus, where in the absence of negative evidence the insufficient range of positive evidence presents the problem for the grammatical acquisition process (cf. Chomsky, 1988).
3 Auxiliary Deletion in the 

*Siarad* Corpus

3.1 Introduction

As has been outlined in section 2.4 above, both Jones (2004) and Borsley et al. (2007) state that the range of pronouns AD occurs with is limited by the grammatical person of their accompanying pronouns. However, they make slightly different predictions as to which grammatical persons these would be, as has been shown in table 2.3, which compares both their predictions. Notably they differed in that Borsley et al. (2007) states that AD occurs with the first person plural and in a limited range also with the third person plural, while Jones (2004) states that occurrence with the first person plural is limited and that AD does not occur with the third person plural. They both agree that AD is limited with the first person singular and does not occur with third person singular, while it does occur with all other pronouns.

In this chapter I describe a corpus analysis that was carried out to validate these predictions, and to see which of the two predictions, if any, reflects
CHAPTER 3. AUXILIARY DELETION IN THE SIARAD CORPUS

the data in the corpus more accurately. For this an automated analysis of the Siarad corpus (Deuchar et al., 2009), a 40-hour conversational corpus of colloquial Welsh, was carried out, in which examples of sentences featuring AD were extracted. How this was achieved is further described in section 3.2 below, and section 3.3 presents and discusses the results of this corpus study, while I briefly summarise them in section 3.4.

3.2 Methodology

In order to extract instances of AD from the Siarad corpus, the corpus was first glossed using the Bangor AutoGlosser (Donnelly and Deuchar, 2011), a constraint-grammar based program that provides automatic tagging for corpora in the CHAT file format (MacWhinney, 2000). The version of the AutoGlosser used was cloned via Git\(^7\) from its official repository\(^8\) on Monday, 27th February 2012. Some minor changes and bug fixes\(^9\) were made to the AutoGlosser’s source code in order to make it run on a Microsoft Windows NT platform with PHP Version 5.3. While these changes should not alter the behaviour of the AutoGlosser, the author plans on committing the changes to the repository and so these should be available in later versions of the AutoGlosser, provided they are to be accepted by the repository’s maintainer. Although the Siarad corpus was originally glossed manually, using the additional glossing tier produced by the AutoGlosser

---

\(^7\)A version control system; see http://git-scm.com/ for more information.

\(^8\)http://thinkopen.co.uk/git/autoglosser

\(^9\)These were instances where files were linked in a Unix file hierarchy which were changed to their Windows equivalents and a number of instances where the source code resulted in E\_NOTICE and E\_WARNING php errors due to the use of outdated syntax or missing declarations for variables and indices.
was advantageous because the glosses it provides are more consistent, richer in detail and also because it does not gloss items that are structurally irrelevant for the purpose of this analysis, such as filled pauses, thus providing a tier for analysis that only contains the constituents that are indeed relevant to identifying AD. In order to gloss the entire corpus, the script given in program listing 1 (appendix A) was used, which runs the AutoGlosser on every CHAT file in the corpus and then extracts the glossed CHAT files from the AutoGlosser’s output.

Following this, a program was written to extract instances of AD from this autoglossed corpus, the program code of which is given in program listing 2 (appendix A). This program includes a parser for the CHAT file format (lines 449 and following), which reads a CHAT file into an object representation (referred to as a \textit{ChatDocument}) and also creates dependent objects for all the tiers and lines in the ChatDocument (which in turn are represented by the \textit{ChatLine} classes) as well as the header data such as speaker information. With this the ChatDocument has a complete, hierarchical representation of the CHAT file it is associated with in which every line is associated with a parent object up to the ChatDocument itself. In case of the dependent tier \texttt{%aut} created by the AutoGlosser this was especially useful since it allows to trace back to the original input line that was glossed from and in turn from this the associated speaker data provided in the files headers. This was used to not only count relevant structures but extract information and the relevant data from the corpus about each instance of AD. In order to extract instances of AD, the program contains the function \texttt{find\_ad()} (lines 387 and following). This function reads a given
CHAPTER 3. AUXILIARY DELETION IN THE SIARAD CORPUS

file into a ChatDocument object and then goes through all the dependent tiers created by the AutoGlosser. Each of these lines was tested on the first item that was overtly glossed (i.e. items the AutoGlosser did not gloss but marked instead as empty, such as filled pauses, were ignored). The condition applied to this item was that it be a pronoun and also that it is only marked for number and person, i.e. it should match the regular expression 

/\bPRON\.[0-9]\[S|P\]/\10. If these conditions are met for the line under observation, this is then counted as an instance of AD and the function then collects and returns data about it, such as the speaker, what pronoun was used, which file and where within it it occurred as well as an extract of 50 characters from the beginning of the parent ChatLine object (i.e. the line that the dependent AutoGlosser tier belongs to). The procedural part of the program ran this function and another function extract_speaker_data() (lines 319 and following) on every CHAT file in the autoglossed corpus. The data thus collected by the program is then written into a relational SQLite 3 database as well as three tabulator separated CSV files. The database can after easily be used to analyse the data and extract subsets depending on several conditions and relations, while the CSV files offer an easy way of importing the same set of data into spreadsheet and statistical software such as Microsoft Excel or SPSS. In terms of content both output formats are equal. The database contains three tables, files, speakers, and ad_instances, which are equivalent to the similarly named CSV files. The

\10Note that the code does not actually use the perl compatible regular expression module (PREG) supplied with PHP, but instead chunks the item and tests for these conditions to match. This approach was adopted purely for performance reasons and is otherwise equivalent to the given expression.
files table contains the filename of the CHAT file and a unique ID for this, which is referenced in the speakers and ad_instances. The speakers table contains all the information extracted about speakers from the CHAT file, such as their age, sex, role, etc. and also gives every speaker a unique ID per file (since the same name could plausibly occur in several files in the corpus but not necessarily refer to the same speaker), which is used for reference in the ad_instances table. The ad_instances table contains all the actual instances of AD that were extracted by the program, for each giving the file ID, the speaker ID, the line number it occurred at in the CHAT file, which person the pronoun had, which number the pronoun had, a combination of these two (e.g. 1S, 2P, ...) and a 50 character extract of the relevant passage in the corpus.

3.3 Results & Discussion

A total of 1662 instances of AD were extracted from the corpus, distributed over 143 out of the total number of 156 speakers in the corpus (this count excludes speakers in the corpus whose role was given as “Investigator”).

Table 3.1 gives a count of the number of instances of AD that were found for each of the grammatical persons in Welsh and how many unique speakers have produced at least one of these instances of AD. This immediately shows that AD with the second person singular is the most common, with 129 speakers producing at least one such utterance. This is followed by AD with the first person singular, which though significantly lower, was still produced by 64 speakers, almost exactly half of the number of speakers
producing AD with the second person singular. The first, second and third person plurals were all relatively rare compared to this, with 25, 27 and 24 speakers producing them respectively. Notably, there was not a single instance of AD together with the third person singular in the extracted data.

This suggests that the predictions of both Borsley et al. (2007) and Jones (2004) held true in that AD with the second person singular was very common, while it was more limited with the first person singular and never occurred with the third singular. However it appears that when it comes to the plural their predictions are less accurate, where Borsley et al. (2007) suggested that AD with the first person plural is acceptable and Jones (2004) predicted that AD with the third person plural is unacceptable, the collected data would suggest that in both cases these are rather limited, and where they both predicted that AD with the second person plural would be acceptable, the data also shows this to be more likely of limited nature. It should especially be noted that this impression is confirmed when not only
the extent but also the distribution of AD instances across the speakers who produce them is taken into account. From the scatter plot in figure 3.1, which for each speaker shows how many instances of AD they produced with the given pronoun, it can be easily seen that there is very large variance in how many such utterances individual speakers produce and that this variance is greater with the more common first and second person singular. All the plural forms show very little variance, suggesting that even among

Figure 3.1: Scatter plot of the distribution of AD instances by grammatical person and number. Every small bar represents one or more speakers who produced $y$ amount of AD instances. The dashed line shows the average number of AD instances per grammatical person and number.
Table 3.2: Results on AD and grammatical person/number in Siarad Corpus compared to predictions from Borsley et al. (2007) and Jones (2004).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1S</td>
<td>limited</td>
<td>limited</td>
<td>yes</td>
</tr>
<tr>
<td>2S</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>3S</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>1P</td>
<td>yes</td>
<td>limited</td>
<td>limited</td>
</tr>
<tr>
<td>2P</td>
<td>yes</td>
<td>yes</td>
<td>limited</td>
</tr>
<tr>
<td>3P</td>
<td>limited</td>
<td>no</td>
<td>limited</td>
</tr>
</tbody>
</table>

those who produced them they may not necessarily be prevalent with the exception of one outlier in the first person plural. This speaker however was also the only child in the corpus who had a conversation with its mother, and it may be well possible that the high rate of AD with the first person plural can be attributed both to the situation and ongoing acquisition process.

This figure also gives averages (dashed lines), which are much higher in the cases of first and second person singular than in the three plural cases, where they actually approach single instances. In comparison it would then be reasonable to say that while the plural forms are limited, the first and second person singular appear to be quite widespread and common\textsuperscript{11}. How this compares to the previous accounts from Borsley et al. (2007) and Jones (2004) is illustrated in table 3.2.

\textsuperscript{11}Though note that this does not take into account geographical distribution, so that AD with the first person singular may indeed be more appropriately classified as limited in some parts of Wales. In terms of acceptability however one would expect that this is still given in other parts of the country when the phenomenon is so widespread.
3.4 Summary

In this chapter I have described a corpus study that looked for instances of AD in direct collocation with personal pronouns in the Welsh conversational Siarad corpus. I have explained how a program was used to extract these and compared the results with the accounts of AD presented by Borsley et al. (2007) and Jones (2004), showing that while they make good predictions overall, a more simple account would be that AD is common with the first and second person singular and very limited with any of the three plural pronouns, while it never occurs with the third person singular.

In the next chapter I will describe a sentence judgement experiment that was carried out after this corpus study to explore some of the wider grammatical patterns in which auxiliaries are used in Welsh and how these relate to AD. This also includes a further look at AD and grammatical person, this time not from a production point of view but explicitly from that of grammatical acceptability.
4 Sentence Judgements on Auxiliary Deletion

4.1 Introduction

In the previous chapter a corpus study was described which tested the predictions made by Jones (2004) and Borsley et al. (2007) and concluded with a slightly different set of predictions for the grammaticality of AD in relation to grammatical person and number (see table 3.2). It was noted that while a corpus study can confirm the existence of predicted constructions, it is due to its limitedness and absence of negative evidence ultimately unsuitable to identify grammatical constraints on AD.

In this chapter an acceptability judgement experiment is described which seeks to test a broad variety of the type of constructions that commonly feature auxiliary use in colloquial Welsh. From this it is hoped to gain a relatively theory-neutral overview of the acceptability of AD in a number of grammatical conditions, such as different agreement patterns, tense, mood, focus sentences, &c. The purpose of this experiment shall be to identify some of the constructions in which AD may be unacceptable. Such data it
is believed is necessary in order to be able to identify areas where further
investigation is needed and also to form the basis of any theoretically mo-
tivated explanatory hypotheses for AD from an objective set of data in the
future.

In order to achieve this set of acceptability judgements, a two-part ex-
periment was constructed. With the assumption that AD is a highly verbal
phenomenon not typical in written discourse (perhaps with the exception
of extremely informal writing such as on social networks, e.g. twitter or
facebook), it was thought that an accurate test of acceptability is best also
contained in this modality, and (1997, Ch. 6) notes that mode of stimulus
presentation indeed affects judgements in this way. For this reason an audit-
ory judgement experiment was designed, in which stimuli were presented to
participants over headphones who then responded on a computer keyboard.
This was followed by a written task in which the same stimuli were presen-
ted as written sentences. This functioned both as a control for the auditory
acceptability experiment and to gain a more finely graded response from
participants through the use of Likert scales. Such a control was thought to
be required not only as it will increase reliability of results through duplica-
tion of judgements, but especially because as Cowart (1997) notes, there
is not much literature on the use of auditory stimuli in syntactic judgement
experiments as of yet\textsuperscript{12}. The precise procedure employed for these two tasks
is further described in section 4.2 below.

\textsuperscript{12}This can probably be attributed to the relative complicatedness of designing and
executing such experiments compared with written tasks, which can for instance be
distributed to a whole class attending a lecture at one time, as also noted by Cowart
(1997, Ch. 6).
CHAPTER 4. SENTENCE JUDGEMENTS ON AD

The constructions tested in this experiment can broadly be classified into the following six groups:

1. Subject agreement, which further tests the results already obtained through the corpus study in chapter 3, but extends this to test gender on the third person singular and non-agreeing non-pronominal subjects which default to the 3S inflection of the auxiliary discussed in section 3.4.

2. Tense and aspect, which tests some different auxiliaries that occur with different tenses and either with or without aspect, such as bod ‘to be’ plus aspectual particle, gwneud ‘to do’ without such a particle and ddaru ‘PAST’. The auxiliary bod was tested for past, future and present tense with the imperfective particle yn and in the present tense with the perfective wedi. The auxiliary gwneud was tested for past and future tense, as there is no explicit present tense for this auxiliary, as already described in section 2.2. Notably, ddaru was not explicitly tested here because in an AD clause there would not be any specific surface structure that shows difference to gwneud sentences (they show different agreement, i.e. ddaru does not agree with its subject, but this is of course lost along with the auxiliary) nor does ddaru have any specific functionality that would prime a listener to assuming that the presented sentence is indeed a ddaru-clause.

3. Mood, this included the auxiliary bod in positive, negative and interrogative mood, which (depending on dialect) is reflected in the initial
morphology of the applicable forms of bod, e.g. rwyt ‘be.PRES.POS’, dwyt ‘be.PRES.NEG’ and wyt ‘be.PRES.INT’.

4. Focus and subject/object movement. This addresses clauses in which either the subject or the object has been fronted for focus, so that differing from the default AuxSVO word order, the word orders SAuxVO, OAuxSV and VOAuxS (focused verb with pied-piped object) are tested for compatibility with AD.

5. This group aims at testing bod-introduced subordinate clauses, which are similar in function to English that-clauses. This is of interest because of the slightly different role of bod in highlighting clause structure and also in that bod here does not show inflection (though arguably it shows agreement in the form of initial consonant mutation to agree with its direct subject\(^{13}\)).

6. Simple wh-questions and wh-questions involving a preposition. Wh-questions generally also show a different surface structure to the default question but are much more common than fronting for focus, so that they make for a valid separate category to be tested. In

---

\(^{13}\)This is because the subject of bod in these clauses occurs as a possessive construction, which in the full form shows as a so-called sandwich construction as shown in the example below:

\[ \text{Roedd Sion yn meddwl fy mod i ar y tren i Loegr.} \]

‘Sion thought that I was on the train to England.’

However, in colloquial speech the possessive adjective (fy in the example above) is commonly dropped, so that in effect information dependent on agreement such as subject and also gender in the third person is solely indicated by the kind of mutation on bod in these utterances.
wh-questions with prepositions, such as “Who are you going out with?” Welsh traditionally also requires an initial preposition with pied-piped complement if the prepositional phrase is to be the element questioned, so that the surface word order is PrepOAuxSV, however Borsley et al. (2007, pp. 114–116) also discuss preposition stranding similar to English interrogatives with prepositions\(^{14}\), where the preposition remains at the end of the sentence, leading to the surface structure OAuxSVPrep. Crucially Borsley et al. (2007) hypothesise that the mechanisms involved in the two types of constructions are different, so that it does not automatically follow that acceptability of AD with one of these constructions also licenses the other.

Each of these four groups was tested using a control condition with an overt auxiliary and comparing these judgements to a condition where the auxiliary was deleted. For each test case four utterances per condition were used, except in the first group, where only two utterances per condition were tested (which was thought sufficient because of the existing data on AD and subject agreement). This procedure is also described in more detail in section 4.2 below.

In this section I have described why a corpus study alone is not sufficient to obtain the data necessary to identify constraints on AD in Welsh and

\(^{14}\)Besides Borsley et al. (2007), such constructions are also mentioned in Davies (2010, p. 276), who takes them from Roberts (1988), and Hendrick (1988, p. 180) who gives as his source Jones and Thomas (1977). A common attribution here seems to be that this is a recent phenomenon, where speakers apparently model the construction on the surface structure of the English equivalent. That the sources for this go back at least 25 years at the time of writing suggest however that, at least for some group of speakers, this must be a reasonably common and recurring pattern and may thus be a construction internal to these speakers’ Welsh, and not just be the mirror of its English equivalent.
argued that such data can be collected via the experiment described in this chapter. In the next section I will give more detail on the methodology that was employed to do this. This is followed by section 4.3 which presents the results from the experiment and a short discussion of them, as well as section 4.4 which presents a summary of this experiment and its results.

4.2 Methodology

In the last section I have described the matter this experiment sets out to investigate and given a rough outline of the approach taken to do this, including a discussion of the kinds of stimuli and conditions that were tested in the experiment. In this section I will describe in more detail the methodology that was adopted in constructing both the auditory ‘online’ judgement task and the written ‘offline’ task which was administered to participants after completion of the online task.

Initially a list of stimuli to be used in accordance with the groups to be tested as outlined in section 4.1 was devised. This included four sentences per construction to be tested, except for the constructions in group one, which tested agreement in person and number with the direct subject, where it was thought sufficient to only test two sentences per construction due to the already existing data from the corpus study as described in chapter 3. Each sentence was then adapted to the two conditions +A (with overt auxiliary) and –A (without an over auxiliary, i.e. an AD sentence), so that all in all there were 8 stimuli per condition (4 for those in group one). Where information was lost due to AD that was important for the
meaning of the sentence, such as tense, sentences were constructed to force an interpretation in that sense. For instance when a past tense sentence was tested, a phrase such as “last year” was added to force a past tense interpretation even in the absence of the tense-carrying auxiliary. A further list of 10 stimuli were constructed for a training task (detailed below), which did not necessarily test the +A and –A conditions but also included other manipulations, which served partially to distract the participant from the +A/–A condition and also to make it easier to see from the data whether the task was understood by looking at some clearly ungrammatical stimuli in the training set. A full list of the stimuli used in the training task and in the test task is contained in appendix B, split into the same groups outlined in section 4.1. It should be noted that some items, if they fit more than one construction/group were re-used to avoid unnecessary duplication which would also highlight the condition tested to the participant. These are indexed by a reference in the form #NN to the stimuli used in their stead in the table in appendix B.

The experiment in which these stimuli were to be tested was designed to be in two parts. First, an online task in which participants would be played recorded versions of the stimuli and had to respond via key presses on a keyboard in a limited amount of time and second an offline task in which participants were given the same stimuli as a printed list but with the ability to go through them in their own time and to rate them in more detail on a Likert scale.

For the online task, all the stimuli were recorded into uncompressed 44.1kHz stereo Waveform Audio files with a Zoom H1 digital audio re-
corder and then edited in Audacity to remove any pauses before and after the stimuli. The speaker used for the recordings was a 20 year old female native speaker from east-mid Wales. In the experiment, participants were first presented with a set of instructions, which explained the procedure of the task and that they should decide whether a sentence they heard sounded natural to them or not (see appendix D for the exact text of the instructions) and that they should press the key Z if they felt the sentence was unnatural or M if they felt it sounded natural. This was followed by a practice task using the training stimuli and then the array of test stimuli (both of which were pseudo-randomised for each instantiation of the experiment) in four blocks of 38 stimuli each, with self-timed breaks in-between\textsuperscript{15}. Presentation of each stimuli was preceded by the appearance of a fixation cross 300ms before the stimulus was played, a short beep (a 400kHz sine wave, amplitude 0.6, duration 100ms, generated with Audacity) 200ms before the stimulus was played and then a delay of 100ms immediately after which the stimuli audio file was played. This procedure was adopted to give participants both a visual and an auditory clue as to when they would hear the next sentence but not directly overlap or adjoin to that stimulus’ beginning, as that was probably the most important part of most stimuli. Participants could then respond immediately from the beginning of the stimulus’ playback up until 1500ms after it had completed playing. During the same time they were shown a reminder of their response options by displaying a red cross above the letter Z in the bottom left and a green tick with the letter M in the

\textsuperscript{15}Participants were informed by an on-screen message how far they were through the task and could resume the task by pressing any key.
bottom right of the screen, reflecting the response options of Z for an unnatural and M for a natural sentence. This procedure was implemented and run using the open-source behavioural experiment software OpenSesame (Mathôt et al., in press; Mathôt and Theeuwes, 2011), the script for which is repeated as program code listing 3 in appendix C, on an ASUS Eee PC 1215N with Windows 7 SP1 and using Behringer HPM1000 headphones for audio playback. Participants’ responses and their reaction times were logged together with the stimulus’ ID (see list in appendix B) and the stimulus’ playback duration, which was measured automatically using the script repeated in program code listing 4 in appendix C.

After completing the auditory online task participants were asked to complete the above described offline task. This involved a printed questionnaire containing a set of instructions followed by a list of the training stimuli and then a list of all the test stimuli, split into blocks of 30 stimuli each. Next to each of these stimuli was a Likert scale which covers the range one to five, whereas participants were told in the included set of instructions (see appendix D) that they should use the box labelled one to indicate that a sentence feels completely unnatural to them, and the box labelled five if it felt completely natural. As with the previous online task, every individual questionnaire was pseudo-randomised for each individual participant. This was achieved by generating the questionnaires automatically via the script repeated in program code listing 5, appendix C, which generates a printable HTML document. A facsimile of the one of the pages of one such document is included as a sample in appendix E. In order to be able to associate answers with their stimuli later on, a reference code
CHAPTER 4. SENTENCE JUDGEMENTS ON AD

was included on every line which masked the stimulus’ ID and condition behind a partially pseudo-generated number\textsuperscript{16}. The masking was applied so that no order could be derived from these numbers by participants and a four-digit number was chosen to make it unlikely that any patterns would be visible from this. After completion by the participants, their answers were typed up manually and stored in a CSV file (cf. section 3.2). To ease this task the script in program code listing 8 was used\textsuperscript{17}.

Participants were additionally asked to complete a short questionnaire on their demographic background, which asked for their age, gender, level of education and the rough area where they grew up. Level of education was multiple-choice at either GCSEs/A-Levels, (Some) Higher Education or (Some) Postgraduate Education. The area they gave was used to later on code them to be from south, north or mid Wales, which should then serve as a rough assessor for dialectal variation\textsuperscript{18}. Like the offline judgement questionnaire described above, this data was then typed up manually and stored in CSV files with the help of the script repeated in program code listing 7, appendix C. After all data was collected, the script repeated in program code listing 9 in appendix C was used to combine the results from the online task, the offline task and the demographic background data into a relational SQLite database, from which data can be easily extracted in different ways using the Structured Query Language (SQL) for analysis in

\textsuperscript{16}See lines 264 and following in program code listing 5 for the precise algorithm that was used to do this.

\textsuperscript{17}The script is intended to be run as an interactive website which displays an HTML form to insert the data and processes them before storage in a CSV file.

\textsuperscript{18}Though note that dialectal distribution and variation in Wales is much more complex than this and so the actual place names they gave were also retained for more detailed analysis if desired.
CHAPTER 4. SENTENCE JUDGEMENTS ON AD

statistical software.

Criteria for participant recruitment were that they are native Welsh speakers and that they are over 18 years of age. Ethics approval for the experiment was obtained from the College of Arts and Humanities Research Ethics Committee on 12 April 2012 and participants were subsequently recruited by advertising via bilingual posters hung up on notice boards around Bangor University, advertisement on the Bangor University online forum\textsuperscript{19} and via the social networks \textit{facebook} and \textit{twitter}. A copy of the message/poster that was used for advertising the study is included in appendix F. Additionally other native Welsh speakers known to me were e-mailed directly about the experiment and also asked to e-mail any of their friends who were native Welsh speakers. At the end of the experiment, which took on average around 30 minutes, participants were reimbursed £5 for their time. Consent was obtained prior to commencement of the experiment and participants were asked to sign the consent form replicated in appendix G.

In this section I have described in detail the methodology that was used to collect judgement data on some of the structures highlighted in section 4.1. In the next section I will present and discuss some of the results obtained from this.

4.3 Results & Discussion

In this section I will discuss the results obtained from the judgement experiment described in the previous sections. For this I will first give some

\textsuperscript{19}http://forum.bangor.ac.uk
brief data about the participants who were recruited and the nature of the overall data collected. This will be followed by a group by group discussion of judgement results based on the six groups of constructions tested that were outlined in section 4.1, in which I will comment on any patterns visible in the data for the relevant constructions.

20 participants between the ages of 19 and 58 ($M = 31.85$) were recruited, of which 17 were from north Wales, 1 from mid Wales and 2 from south Wales. Male to female ratio was balanced at 10:10; 5 participants had GCSEs or A-levels, 8 had at least some higher education and 7 had at least some postgraduate education.

A total of 85 single judgement results from the offline task were excluded and this included all results for stimulus #8 in the test set due to a programming error which omitted these in generating the form handed out to participants. The other exclusions here were where participants missed out single questions in the offline task, which in the case of one participant resulted in 30 missing responses due to failure to complete one page of the form. In the online task, there were 176 timed out responses, though no single stimulus was affected over 4 times from this, and this was distributed over a range of 100 out of the 152 stimuli that were tested.

With concern to the overall judgements and the methodology used, one important consideration must be the reliability of the judgements obtained in the online task, as there is not much research on this method of collecting acceptability judgements. As Cowart (1997, p. 63) notes there are likely to be some differences between results for judgement experiments presented in different ways but they should still be highly correlated at large. This was
confirmed by a chi-square test on the correlation between the online task’s yes/no judgements and the one-to-five scale ratings collected from the offline task for every individual participant and stimulus, which showed a highly significant relation ($\chi^2(8, N = 2956) = 645.59, p < .001$). An additional test on correlation was carried out comparing the mean online response ($1 \leq x \leq 2$) with the mean offline response ($1 \leq x \leq 5$) for every stimulus, which also showed a highly significant relationship between the two measurements ($r(150) = .79, p < .001$) and this relationship can also be seen together with some expected outliers in the scatter plot in figure 4.1. This shows that the data collected shows good overall reliability across the two methods of stimuli presentation used and I will on this basis propose that the differences in judgements found across the two methods are meaningful on the basis of my previous stipulation that the online judgements are closer to the criteria relevant in colloquial spoken language. In the following I will examine the results obtained on the different constructions that were tested.

The first group of constructions tested for AD acceptability with different pronominal and non-pronominal subjects. Table 4.1 contains computations of the means and standard deviations for the combined responses by construction and condition. This shows that AD is very acceptable with 2S, where it may even be preferred over an overt auxiliary by some speakers as indicated by the slightly higher mean score in the online task, though in the written task the overt auxiliary seemed to be slightly preferred. This is followed by 1P, 2P and 1S which also seem to be quite acceptable with mean ratings of acceptability well over the median of the respective scales. Notably with 3S there is a difference in acceptability between the male
CHAPTER 4. SENTENCE JUDGEMENTS ON AD

Figure 4.1: Scatter plot and trend \((R^2 = .62)\) in comparison of mean responses for online and offline task per stimulus

(3Sm) and the female (3Sf) pronoun constructions, with the former being accepted much more readily in the online task, though again in the written offline task both seem to be quite unacceptable. The difference in the online responses between 3Sm and 3Sf was then shown to be significant \((t(73) = 2.76, p < .01)\) by running an independent samples t-test. Judgements on AD sentences with noun phrases (NP) and proper nouns (PN) appear to be very unstable and varied, as indicated by their ratings close to median and standard deviations, and so possibly justify further investigation.

The second group tested some different auxiliaries (\textit{bod} and \textit{gwneud}) and the interaction between their tense and (with \textit{bod}) aspect. Again means and
### Table 4.1: Means and standard deviations for constructions which tested AD acceptability with different direct subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Condition</th>
<th>Online Response</th>
<th>Offline Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>1S</td>
<td>+A</td>
<td>2.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.68</td>
<td>.48</td>
</tr>
<tr>
<td>2S</td>
<td>+A</td>
<td>1.92</td>
<td>.28</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.97</td>
<td>.16</td>
</tr>
<tr>
<td>3Sf</td>
<td>+A</td>
<td>2.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.63</td>
<td>.49</td>
</tr>
<tr>
<td>3Sm</td>
<td>+A</td>
<td>1.95</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.32</td>
<td>.48</td>
</tr>
<tr>
<td>1P</td>
<td>+A</td>
<td>1.97</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.79</td>
<td>.41</td>
</tr>
<tr>
<td>2P</td>
<td>+A</td>
<td>1.79</td>
<td>.41</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.82</td>
<td>.39</td>
</tr>
<tr>
<td>3P</td>
<td>+A</td>
<td>2.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.46</td>
<td>.51</td>
</tr>
<tr>
<td>NP</td>
<td>+A</td>
<td>2.00</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.53</td>
<td>.50</td>
</tr>
<tr>
<td>PN</td>
<td>+A</td>
<td>1.90</td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.43</td>
<td>.50</td>
</tr>
</tbody>
</table>

Standard deviations were computed for every construction tested, which are given in Table 4.2. In relation to tense with bod this shows that while the present tense constructions are highly acceptable, constructions that imply past tense are rated to be fairly unacceptable. Constructions that imply future tense appear to be more acceptable, though not as acceptable as those clearly falling within present tense. AD with gwneud appears to be generally relatively unacceptable, regardless of whether the sentences implied future or past tense. As the bod.PRES constructions were also
Table 4.2: Means and standard deviations for AD acceptability with bod and gwneud, PRES/PAST/FUT and PFV/IMPFV

<table>
<thead>
<tr>
<th>Construction</th>
<th>Condition</th>
<th>Online Response</th>
<th>Offline Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>bod.PAST</td>
<td>+A</td>
<td>1.68</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>–A</td>
<td>1.39</td>
<td>.49</td>
</tr>
<tr>
<td>bod.PRES</td>
<td>+A</td>
<td>1.93</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>–A</td>
<td>1.96</td>
<td>.20</td>
</tr>
<tr>
<td>bod.FUT</td>
<td>+A</td>
<td>1.84</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>–A</td>
<td>1.73</td>
<td>.45</td>
</tr>
<tr>
<td>gwneud.PAST</td>
<td>+A</td>
<td>1.93</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>–A</td>
<td>1.30</td>
<td>.46</td>
</tr>
<tr>
<td>gwneud.FUT</td>
<td>+A</td>
<td>1.95</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>–A</td>
<td>1.47</td>
<td>.50</td>
</tr>
<tr>
<td>bod + PFV</td>
<td>+A</td>
<td>1.86</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>–A</td>
<td>1.97</td>
<td>.16</td>
</tr>
</tbody>
</table>

followed by an imperfective particle (yn), it is already known that this is highly acceptable at this point, and bod with the perfective particle (wedi) appears to be equally acceptable.

In the third group, the acceptability of AD in affirmative, interrogative and negative mood constructions was tested, for which means are shown in table 4.3. Here all three groups showed very high acceptability over both groups, regardless of mood, so that it can be concluded that mood is not a significant factor in AD acceptability.

The fourth group looked at AD in constructions with subject or object fronting for focus, leading to a variety of surface structures. Again a range of means and standard deviations has been computed for these constructions, which is given in table 4.4, though this does not include the
CHAPTER 4. SENTENCE JUDGEMENTS ON AD

Table 4.3: Means and standard deviations for AD acceptability dependent on mood

<table>
<thead>
<tr>
<th>Construction</th>
<th>Condition</th>
<th>Online Response</th>
<th>Offline Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>AFF</td>
<td>+A</td>
<td>1.93</td>
<td>.25</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.96</td>
<td>.20</td>
</tr>
<tr>
<td>INT</td>
<td>+A</td>
<td>1.95</td>
<td>.22</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.97</td>
<td>.16</td>
</tr>
<tr>
<td>NEG</td>
<td>+A</td>
<td>1.97</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.95</td>
<td>.21</td>
</tr>
</tbody>
</table>

default AuxSVO structures, as the previous constructions in group three have already pertinently demonstrated these to be acceptable. This data shows that while all of these constructions are less acceptable overall even with an over auxiliary, those featuring AD are only slightly less acceptable than their counter-parts. Notably however, constructions with an SAuxVO surface structure\(^{20}\), seemed to be much more acceptable when they were presented as spoken sentences in the online task than they were in the written offline task.

In the fifth group simple subordinate constructions roughly equatable to English *that*-clauses where tested, means and standard deviations for which are given in table 4.5. These results show that while the subordinates overall received slightly lower acceptability ratings those featuring AD are only slight lower in acceptability than those with an overt auxiliary. This is possibly indicative of AD being acceptable in these constructions, but some

\(^{20}\)NB: This featured a subject in the form \([\text{DP}]\{\text{Y}\}[[\text{Nti}]]\) ‘the you’, and the verb *bod* in the form *sy* rather than the 2S agreement inflection *rwyt*. 
Table 4.4: Means and standard deviations for AD in different constructions with non-default surface structure

<table>
<thead>
<tr>
<th>Construction</th>
<th>Condition</th>
<th>Online Response</th>
<th>Offline Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>SAuxVO</td>
<td>+A</td>
<td>1.87</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.70</td>
<td>.46</td>
</tr>
<tr>
<td>VOAuxS</td>
<td>+A</td>
<td>1.66</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.56</td>
<td>.50</td>
</tr>
<tr>
<td>OAuxSV</td>
<td>+A</td>
<td>1.61</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.53</td>
<td>.50</td>
</tr>
</tbody>
</table>

Table 4.5: Means and standard deviations for AD in subordinate clauses

<table>
<thead>
<tr>
<th>Construction</th>
<th>Condition</th>
<th>Online Response</th>
<th>Offline Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Subordinate</td>
<td>+A</td>
<td>1.85</td>
<td>.36</td>
</tr>
<tr>
<td></td>
<td>−A</td>
<td>1.71</td>
<td>.46</td>
</tr>
</tbody>
</table>

Further testing would be appropriate to confirm this.

The sixth and final group tested Wh-questions, both with and without prepositions and with pied-piped and stranded prepositions. Means and standard deviations for all these three construction types are given in Table 4.6. While it illustrates that constructions with stranded prepositions (Wh + Prep) are much less acceptable than those with pied-piped prepositions, there does not appear to be any significant difference between −A and +A conditions. In fact the results in the online responses are remarkably consistent across conditions. This suggests that wh-question formation does not impact on AD acceptability, and that pied-piping vs preposition stranding has no effect on AD acceptability.
Table 4.6: Means and standard deviations for AD in Wh-questions

<table>
<thead>
<tr>
<th>Construction</th>
<th>Condition</th>
<th>Online Response</th>
<th>Offline Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Wh + A</td>
<td>1.91</td>
<td>.29</td>
<td>4.41</td>
</tr>
<tr>
<td>–A</td>
<td>1.91</td>
<td>.29</td>
<td>4.35</td>
</tr>
<tr>
<td>Wh + Prep + A</td>
<td>1.62</td>
<td>.49</td>
<td>3.74</td>
</tr>
<tr>
<td>–A</td>
<td>1.63</td>
<td>.49</td>
<td>3.08</td>
</tr>
<tr>
<td>Prep + Wh + A</td>
<td>1.99</td>
<td>.11</td>
<td>4.77</td>
</tr>
<tr>
<td>–A</td>
<td>1.95</td>
<td>.22</td>
<td>4.83</td>
</tr>
</tbody>
</table>

4.4 Summary

In this chapter I have presented a judgement experiment in which six groups of auxiliary carrying constructions were tested for their compatibility with AD. In the previous section I have presented and discussed some of the main results from this experiment, which showed that as was already presumed from the previous studies on AD discussed and presented in this work, the type of subject to an auxiliary is of great importance in AD acceptability and notably that non-agreeing non-pronominal subjects (which fall back to the 3S inflection of the auxiliary) appear to have been judged quite inconsistently and need further investigation. It was also noted that gender in constructions with 3S appeared to be a significant factor, which has not been previously described in the literature on AD in Welsh. Tense and type of auxiliary used (e.g. gwneud ‘to do’) were also shown to be an important factor, notably AD was largely unacceptable with auxiliaries other than bod ‘to be’ and tense other than present tense, though future tense appeared to
be slightly more acceptable than past tense. On the other hand it was found that mood, aspect, movement for focus and wh-question formation had no significant impact on AD acceptability, whether they involved a pied-piping strategy or not.

In the next chapter I will briefly discuss how these findings integrate with the previous literature discussed in chapter 2 and the findings from the corpus study in presented in chapter 3 and what they implicate for further research on the constraints involved in Welsh auxiliary deletion.
5 Discussion: Possible Constraints on Auxiliary Deletion

In this dissertation I have so-far described some of the previous research that has been carried out on auxiliary deletion in Welsh and proposed that in order to gain a better understanding of the extent of the phenomenon and its implications for the grammar of Welsh, exploratory work is needed to find some of the basic constraints that apply to AD. I have followed this by describing two exploratory studies, a corpus study that looked at AD and the kinds of pronominal subjects it co-occurred with, and a judgement experiment that tested a wide array of typical Welsh constructions involving auxiliaries for their grammatical acceptability. In this chapter I will discuss what these two studies can tell us about the possible constraints that underlie AD in colloquial Welsh and how this relates to the previous literature described in chapter 2.

One of the central issues in previous descriptions was the subject of an AD clause, specifically the grammatical number and person of pronom-
CHAPTER 5.  SENTENCE JUDGEMENTS ON AD

inal subjects. Different accounts were given by Borsley et al. (2007) and Jones (2004) and so one of the aims of the first study was to see which of their predictions fitted better with real data found in a corpus of colloquial Welsh. The corpus study showed that while both their predictions were quite accurate for singular pronominal subjects, with the plural pronouns there was no precise agreement with either account and it was found that AD would occur with all the plural pronouns, but very limited. Based on the contrast in number of instances between AD with 1S or 2S and with 1P, 2P or 3P, it was also proposed that AD with 1S, which both previous accounts described as limited, must probably be quite acceptable regardless of the speakers dialect\(^{21}\). This was however not confirmed in the judgement experiment, which suggested that while it was still acceptable with most northern speakers, it was significantly less acceptable than 2S and 1P and 2P, which is actually more in agreement with the predictions made by Borsley et al. (2007) (cf. also table 3.2). An interesting additional discovery in the judgement experiment was that 3S, which was previously described as ungrammatical and also not found in the corpus analysis, was of roughly the same acceptability as 1S with a female pronoun, while only male pronouns were mainly unacceptable with AD. Noun phrases and proper nouns were also found to be mostly unacceptable as subjects of AD clauses, though they were for some speakers. This high variance and actually higher acceptability than production (corpus v judgements) may indeed be a sign of language change in progress, where it could be expected that if gram-

\(^{21}\)One of the main assumptions for the limitedness of 1S is that it is found mainly in the southern dialect of Welsh.
maticality of AD widened this would first show in acceptability and were only then possibly followed by wider adoption. Interesting here would be a further corpus analysis focusing on data from very young speakers which could further confirm this\(^\text{22}\).

Another major question was whether AD is limited to the auxiliary *bod* ‘to be’ or whether, as the term suggests, it also occurs with other Welsh auxiliaries such as *gwneud* ‘to do’ or *ddaru* ‘PAST’. The judgement experiment’s results suggested that deletion of *gwneud* is unacceptable, however there are some complications in concluding from this that AD is limited to *bod*. One of these is that *gwneud* and *ddaru* are both limited to tenses other than present tense, and this was also shown to be a constraint on AD in that past tense constructions were largely judged unacceptable and future tense constructions significantly less acceptable than present tense constructions. Additionally a problem in testing these (and indeed in their interpretability with AD) arises from what Davies (2010, pp. 326–323) describes as particle deletion, where the aspectual particle is omitted. This leads to ambiguity in sentences such as 5–2 which could have been derived from either 5–1a or 5–1 b.

\[(5–1) \quad \text{(a)} \quad Wyt\quad ti\quad 'n\quad siarad\quad efo\quad Sion?\]
\[\text{be.2S.PRES}\quad \text{you}\quad \text{IMPFV}\quad \text{speak}\quad \text{with}\quad \text{Sion}\]
\[\text{‘Were you speaking to Sion?’}\]

\[(b)\quad Wnest\quad ti\quad siarad\quad efo\quad Sion?\]
\[\text{do.2S.PAST}\quad \text{you}\quad \text{speak}\quad \text{with}\quad \text{Sion}\]
\[\text{‘Did you speak to Sion?’}\]

\(^{22}\text{Davies (2010, pp. 295–302) already argues that the 2S pronoun itself shows higher adoption in younger speakers, which he suggest could be suggestive of language change in progress.}\]
CHAPTER 5. SENTENCE JUDGEMENTS ON AD

(5–2) *Ti siarad efo Sion?*

you talk with Sion

‘Were you speaking to Sion?’
‘Did you speak to Sion?’

This is of course easily resolved if AD remains limited to *bod*, but I suggest that further evidence is required to answer this question. A possible way here may be collecting acceptability judgements on items where the subject would show initial consonant mutation in a future tense *gwneud* clause but not in a similar *bod* clause, provided that particle drop can be shown to not lead to initial consonant mutation.

The acceptability judgements further showed that neither word order (cf. e.g. focus clauses or Wh-questions v the default AuxSVO) nor subordination constrain AD in any obvious way. Further neither mood nor aspect appeared to constrain AD either. However, as discussed above already AD is constrained by the types of subject it can occur with and presumably the agreement it shows with them as well as tense. A notable difference in the relation these two factors have to the auxiliary is that they affect its inflectional morphology, as opposed to mood which results in prefixation (if it is overt at all, which depends on dialect) and aspect exhibits no overt effect on the auxiliary. An argument here may be that the relation between the auxiliary and these constraining factors is stronger than that between the auxiliary and other factors. While this does not directly explain any bias for gender in the third person, I suggest that third person gender is in itself an important factor in that other relations depend on it, for instance in inhibiting different patterns of initial consonant mutation in adjectives; while
inflections for 3Sf and 3Sm are homophonous on the surface then, they may be different internally and the way in which speakers constrain AD may best be explained via the inflectional paradigm applied to the auxiliary.
6 Conclusions

This dissertation set out to explore some of the constraints that apply to auxiliary deletion in colloquial Welsh. It started from the viewpoint that AD had been studied very little in the wider context of the kind of constructions (specifically periphrastic constructions) in which it could potentially occur and that most descriptions of it until now focused on the pronouns it co-occurred with and whether it is motivated by language change due to the influence of English. The question followed what other factors may constrain the occurrence of AD, such as the type of auxiliary used, the constructions it occurred in, word order and movement, and the kinds of relations that affect the auxiliary in these sentences, such as mood and agreement.

Data was collected through two studies, a corpus analysis on the *Siarad* corpus of informal Welsh speech and a subsequent judgement experiment. This showed that while individual patterns of AD are highly variable, there are some clear factors that play a role in whether AD may occur in a sentence or not. It was argued that a common feature of factors that were found to constrain AD was that they had an important relation with the auxiliary, and usually one that is determinative for the inflectional morpho-
logy of the auxiliary. The two elements that showed to be vital in this were the subject AD occurred with and the tense of the clause. Further, these studies provided the first objective account on both the relative acceptability and spoken distribution of AD with pronominal subjects other than 2S, where it was shown that while previous predictions were generally quite good at predicting the data, slight differences are present. Additionally it was shown that there is a clear gap between the occurrence of these items in the speech of the Siarad corpus and the acceptability that speakers attach to them when they are exposed to these constructions, and in a minor fashion even that these depend on the modality through which they are experienced (i.e. auditory vs visually).

In light of the existing debate over whether this phenomenon reflects language change due to the influence of English (e.g. Davies, 2010; Davies and Deuchar, in preparation), it was also noted that the data would support an analysis where the phenomenon was initially introduced in the realm of the 2S present tense paradigm of bod, which one would expect to also be most common in colloquial speech, and is now widening onto other parts across the inflectional paradigm of the auxiliary.

This work also highlighted some further areas of uncertainty that would warrant further experimental investigation, such as whether auxiliaries other than bod ‘to be’ can be deleted and given the acceptability ratings of 3Sf whether AD does really never occur with 3S in spontaneous speech.

Some limitations of the study and experiment were that they only looked at some very broad structures to identify the major factors that play a role, and further investigation in the areas shown to be relevant here may well
CHAPTER 6. CONCLUSIONS

highlight some finer important details. The corpus study was also limited due to the corpus’s focus on adult Welsh-English bilinguals. Re-running the corpus analysis on a corpus of children’s speech and non Welsh-English bilingual speakers, such as those in the *Patagonia* corpus could give further insights and have implications for the analysis of AD as language change in progress.
References

ACCAC (2000). *English in the National Curriculum in Wales*. Cardiff: Qualifications, Curriculum and Assessment Authority for Wales, on behalf of the National Assembly for Wales.


Davies, P. and Deuchar, M. (in preparation). Auxiliary deletion in bilingual Welsh-English speech: internal change or the influence of English?
REFERENCES


REFERENCES


Appendix

A Program Code Listings for Corpus Study

Listing 1: Script for Autoglossing entire Siarad corpus

```php
<?php

// Set up PHP to report all errors
error_reporting(E_ALL);
ini_set("display_errors", 1);
ini_set("log_errors", 1);
ini_set("error_log", ".errors.log");

// Check user arguments...
```
Listing 2: Script for finding AD in autoglossed corpus

```php
if ($argc != 2) {
    die('This script takes exactly one argument (the path to the directory to
    .be autoglossed). "(-$argc)." arguments given.');
}

//Read directory and run AutoGlosser on it's CHAT files...
$dirname = $argv[1];
$dir = @dir($dirname) or die('The specified directory could not be found.');
@mkdir('outputs/.' . basename($dir->path).'_autoglossed');
while (false !== $file = $dir->read()) {
    if (substr($file, -4) == '.cha') {
        //Only chat files
        $exec_file = $dir->path . '/' . $file;
        //Now run do_everything for each..
        print '\'n***'n* Autoglossing file: '$file
        passthru('php do_everything.php "$exec_file"');
        copy('outputs/.' . basename($file, '.cha') . '/' . basename($file, '.cha')
            . '_autoglossed.txt';
        'outputs/.' . basename($dir->path).'_autoglossed/' . $file);
    }
}
?>
```
//Some PHP stuff
error_reporting(E_ALL);
ini_set('display_errors', 1);
define('UTF8_BOM', chr(0xEF).chr(0xBB).chr(0xBF));

// Where to find the chat files for analysis
$original_dir = './Siarad';
$autoglossed_dir = './Siarad_autoglossed';
$out_dir = './';

// MAIN SCRIPT FOR FINDING AD IN SIARAD

// Prepare database...
echo "Preparing database...\t\t"
$fh = @fopen($out_dir . '/ad_data.sqlite', 'w'); // This will 'empty' the db..
if ($fh === false)
    die('\nError: Could not open file ' . $out_dir . '/ad_data.sqlite' . ' for writing.');
fclose($fh);
$db = new SQLite3($out_dir . '/ad_data.sqlite', SQLITE3_OPEN_READWRITE);
$result = $db->exec('CREATE TABLE files
  (f_id INTEGER PRIMARY KEY,
f_filename TEXT);
  CREATE TABLE speakers
  (s_id INTEGER PRIMARY KEY,
f_id INTEGER,
s_name_code TEXT,
s_name TEXT,
s_role TEXT,
s_language TEXT,
s_corpus TEXT,
s_age TEXT,
s_sex TEXT,
s_group TEXT,
s_SES TEXT,
s_education TEXT);
  CREATE TABLE ad_instances
  (ad_id INTEGER PRIMARY KEY,
s_id INTEGER,
f_id INTEGER,
ad_line_no INTEGER,
ad_person INTEGER,
ad_number TEXT,
ad_persnum TEXT,
ad_extract TEXT)");
if(!$result) {
    die('SQL Error at line __LINE__: ' . $db->lastErrorMsg());
}
"Done\n";
APPENDIX

94  // Extract all file names to search for ad...
95  echo "Creating file index... \\
96  $filelist = array();
97  $dir = @dir($autoglossed_dir) or die("\nThe directory with the autoglossing \\
98  data could not be found.");
99  $stmt = $db->prepare('INSERT INTO files 
100  (f_filename) 
101  VALUES 
102  (:f_filename);');
103  while(false !== $file = $dir->read()) {
104    if(substr($file, -4) == " . cha") { // Only chat files
105      $stmt->reset();
106      $stmt->bindValue(':f_filename', $file);
107      $stmt->execute();
108      $f_id = $db->lastInsertRowID();
109      $filelist[] = array($file, $f_id);
110    }
111  }
112  echo "Done.\n";
113
114  // Write file list...
115  echo "Writing file list... \\
116  $fh = @fopen($out_dir." / file_list.csv", 'w');
117  if($fh === false) {
118    die("\nError: Could not open file '" . $out_dir." / file_list.csv' for writing.");
119  }
120  fwrite($fh, UTF8_BOM);
121  fwrite($fh, 'f_id\tf_filename\n');
122  foreach($filelist as $file) {
123    fwrite($fh, implode(" \t", $file)."\n");
124  }
125  fclose($fh);
126  echo "Done.\n";
127
128  // Extract speaker data..
129  echo "Extracting speaker data... \\
130  $last_count = 0;
131  $speaker_index = array();
132  $stmt = $db->prepare(' INSERT INTO speakers 
133  (f_id, 
134    s_name_code, 
135    s_name, 
136    s_role, 
137    s_language, 
138    s_corpus, 
139    s_age, 
140    s_sex, 
141    s_group, 
142    s_SES, 
143    s_education 
144  ) VALUES 
145  ( 
146    :f_id, 
147    :s_name_code, 
148    :s_name, 
149    :s_role, 
150  )
151
152  `
for($i=0; $i<count($filelist); $i++) { 
    list($filename, $f_id) = $filelist[$i]; 
    shell_del_chrs($last_count); 
    $out_str = '(' . ($i+1) . '/' . count($filelist) . ')'; 
    echo $out_str; 
    $last_count = strlen($out_str); 
    $speakers = extract_speaker_data($original_dir.'/'.$filename); 
    foreach($speakers as $speaker) { 
        $stmt->reset(); 
        $stmt->bindValue(':f_id', $f_id); 
        $stmt->bindValue(':s_name_code', $speaker['name_code']); 
        $stmt->bindValue(':s_name', $speaker['name']); 
        $stmt->bindValue(':s_role', $speaker['role']); 
        $stmt->bindValue(':s_language', $speaker['language']); 
        $stmt->bindValue(':s_corpus', $speaker['corpus']); 
        $stmt->bindValue(':s_age', $speaker['age']); 
        $stmt->bindValue(':s_sex', $speaker['sex']); 
        $stmt->bindValue(':s_group', $speaker['group']); 
        $stmt->bindValue(':s_SES', $speaker['SES']); 
        $stmt->bindValue(':s_education', $speaker['education']); 
        $stmt->execute(); 
        $s_id = $db->lastInsertRowID(); 
        $speaker_index[] = array_merge(array('s_id' => $s_id, 'f_id' => $f_id), $speaker); 
    } 
    shell_del_chrs($last_count); 
    unset($last_count); 
    echo "Done.\n"; 
}

// Write speaker data... 
// Writing speaker data...\nt\nt; 
$fh = @fopen($out_dir.'/speaker_data.csv', 'w'); 
if($fh === false) { 
    die('\nError: Could not open file `$out_dir/speaker_data.csv`\n\nfor writing.\n'); 
} 
fwrite($fh, UTF8_BOM); 
fwrite($fh, "s_id\tf_id\ts_name_code\ts_name\ts_role\ts_language\ts_corpus\ts_age\ts_sex\ts_group\ts_SES\ts_education\n"); 
foreach($speaker_index as $speaker) { 
    fwrite($fh, implode("\t", $speaker).'\n'); 
} 
fclose($fh); 
echo "Done.\n"; 

// Find ad lines for every file... 
// Parsing files for ad lines...\nt; 
$last_count = 0; 
$ad_index = array();
$stmt1 = $db->prepare('SELECT s_id, s_name_code FROM speakers WHERE f_id = :f_id; ');
$stmt2 = $db->prepare('INSERT INTO ad_instances ('.
    s_id, f_id, ad_line_no, ad_person, ad_number, ad_persnum, ad_extract
) VALUES ('.
    :s_id, :f_id, :ad_line_no, :ad_person, :ad_number, :ad_persnum, :ad_extract
); ');

for($i=0; $i<count($filelist); $i++) {
    list($filename, $f_id) = $filelist[$i];
    shell_del_chrs($last_count);
    $out_str = ' ( ' . ($i+1). ' / ' . count($filelist) . ' ) ' ;
    echo $out_str ;
    $last_count = strlen($out_str) ;
    $stmt1->reset();
    $stmt1->bindValue(':f_id', $f_id);
    $results = $stmt1->execute();
    $speakers = array();
    while($result = $results->fetchArray()) {
        $speakers[$result['s_name_code']] = $result['s_id'];
    }
    $ad_lines = find_ad($autoglossed_dir . '/' . $filename);
    foreach($ad_lines as $ad_line) {
        $stmt2->reset();
        $stmt2->bindValue(':s_id', $speakers[$ad_line['name_code']] );
        $stmt2->bindValue(':f_id', $f_id );
        $stmt2->bindValue(':ad_line_no', $ad_line['line_no'] );
        $stmt2->bindValue(':ad_person', (int) $ad_line['g_person'] );
        $stmt2->bindValue(':ad_number', $ad_line['g_number'] );
        $stmt2->bindValue(':ad_persnum', $ad_line['g_persnum'] );
        $stmt2->bindValue(':ad_extract', $ad_line['extract'] );
        $stmt2->execute();
        $ad_id = $db->lastInsertRowID();
        $ad_index[] = array('ad_id' => $ad_id, 's_id' => $s_id , 'f_id' => $f_id , 'ad_line_no' => $ad_line['line_no'], 'ad_person' => $ad_line['g_person'], 'ad_number' => $ad_line['g_number'], 'ad_persnum' => $ad_line['g_persnum'], 'ad_extract' => $ad_line['extract']);
    }
}
APPENDIX

```plaintext
shell_del_chrs($last_count);
unset($last_count);
echo "Done.\n";

// Write ad instances...

echo "Writing list of AD instances...\t";
$fh = @fopen($out_dir . '/ad_list.csv', 'w');
if ($fh === false) {
die("\nError: Could not open file '" . $out_dir . '/ad_list.csv' . ' for writing.';)
}
fwrite($fh, UTF8_BOM);
fwrite($fh, "ad_id	s_id	f_id
tad_line_no	ad_person	ad_number	ad_person_num	ad_extract\n");
foreach($ad_index as $ad_line) {
    fwrite($fh, implode("\t", $ad_line) . '\n');
}
fclose($fh);

echo "Done.\n";

// Script execution is complete.\n";

// CLASSES AND FUNCTIONS

//∗∗∗
∗Delete Characters from Shell STDOUT
∗
∗This function overwrites the last n characters on STDOUT with whitespace and
∗then sets the cursor to the beginning of that whitespace. This only works in
∗a shell environment when backspaces can override the current line and does
∗not work across linebreaks.
∗
∗@param int $count How many characters to overwrite
∗@return void
∗/

function shell_del_chrs($count) {
    for($i=0;$i<$count;$i++) {
        echo chr(8); // return to left
    }
    for($i=0;$i<$count;$i++) {
        echo ' '; // overwrite with ws
    }
    for($i=0;$i<$count;$i++) {
        echo chr(8); // return to left
    }
}

//∗∗∗
∗Extract speaker data from CHAT files
∗
∗This function extracts all available data about participants from the given
∗CHAT file.
∗
∗@param string $filename The CHAT file from which the data should be extracted
∗@return array Returns a numeric array of the speaker information
∗/

function extract_speaker_data($filename) {
    // Open and parse file
```

67
$cf = new ChatDocument($filename);
$cf->parseFile();

// Get all header lines
$speaker_data = array();
$header_lines = $cf->getHeaderLines();
foreach($header_lines as $header_line) {
    switch(strtolower($header_line->getIdentity())) {
    case 'participants':
        // data will be of the format XXX Name Role, XXX Name Role, ...
        $parts_header = $header_line->getData();
        $parts_header = explode(' ', $parts_header);
        foreach($parts_header as $parts_item) {
            $parts_item = explode(' ', trim($parts_item), 3);
            $id = $parts_item[0];
            $name = $parts_item[1];
            $role = $parts_item[2];
        }
        $filename = $header_line->getParent()->getFilename();
        $filename = basename($filename);
        $speaker_data[$id] = array('name_code' => $id,
                                   'name' => $name,
                                   'role' => $role);
        break;
    case 'id':
        // Format is: lang|corpus|code|age|sex|group|SES|role|edu
        // Index: 0 1 2 3 4 5 6 7
        $id_header = $header_line->getData();
        $id_header = explode(' | ', $id_header);
        $speaker_data[$id_header[2]] += array('language' => $id_header[0],
                                               'corpus' => $id_header[1],
                                               'age' => $id_header[3],
                                               'sex' => $id_header[4],
                                               'group' => $id_header[5],
                                               'SES' => $id_header[6],
                                               'education' => $id_header[8]);
        break;
    }
}

// replace array keys with numbered index
$new_speaker_data = array();
foreach($speaker_data as $item) {
    $new_speaker_data[] = $item;
}
return $new_speaker_data;

/***
APPENDIX

388  * Find instances of Auxiliary Deletion in an AutoGlosser CHAT file
389  * This function searches the given CHAT file's dependent tier %aut line
390  * generated by the Bangor AutoGlosser for instances where the first over item
391  * is a personal pronoun and returns an array
392  *
393  * @param string $filename The CHAT file which should be parsed for AD instances
394  * @return array Returns an array of AD instances in the specified CHAT file
395  */
396
397  function find_ad($filename) {
398     //Open and parse file
399     $cf = new ChatDocument($filename);
400     $cf->parseFile();
401
402     //Get all autoglosser lines
403     $aut_lines = array();
404     $part_lines = $cf->getPartLines();
405     foreach($part_lines as $part_line) {
406         $dependent_lines = $part_line->getDependentLines();
407         foreach($dependent_lines as $dependent_line) {
408             if($dependent_line->getIdentifer() == 'aut') {
409                 $aut_lines[] = $dependent_line;
410             }
411         }
412     }
413     unset($part_line, $part_lines, $dependent_line, $dependent_lines);
414
415     //Find lines that begin with pronouns
416     $ad_lines = array();
417     foreach($aut_lines as $aut_line) {
418         $first_item = trim($aut_line->getData()); //rm any empty glosses
419         $first_item = substr($first_item, 0, strpos($first_item, ' ')); //1st ws
420         if(!empty($first_item)) {
421             $first_item = explode('.', $first_item);
422             if( count($first_item) == 3 //match for xxx.xxx.xxx
423                 && $first_item[1] == 'PRON' //match for xxx.PRON.xxx
424                 && is_numeric($first_item[2][0]) //match for xxx.xxx.(0-9)x
425             ) {
426                 //This is probably a AD clause! It starts with a pronoun..
427                 //Now gather data about it...
428                 $filename = $aut_line->getParent()->getParent()->getFilename();
429                 $filename = basename($filename);
430                 $line_no = $aut_line->getOrigLineNo();
431                 $speaker = $aut_line->getParent()->getIdentifer();
432                 $g_person = $first_item[2][0];
433                 $g_number = $first_item[2][1];
434                 $extract = substr($aut_line->getParent()->getData(), 0, 50);
435                 $ad_lines[] = array('name_code' => $speaker,
436                     'g_person' => $g_person,
437                     'g_number' => $g_number,
438                     'g_persnum' => $g_person.$g_number,
439                     'line_no' => $line_no,
440                     'extract' => $extract
441             )
442         }
443     }
444
445     return $ad_lines;
446
69
APPENDIX

447 }
448
449 /***
450 * Root Class for CHAT Objects
451 *
452 * This is a generic root class from which all other CHAT Objects are derived.
453 * It cannot be directly instantiated.
454 *
455 * @package ChatTools
456 * @abstract
457 */
458 abstract class ChatObject {
459     //dummy class
460 }
461
462 /***
463 * CHAT Document Class
464 *
465 * This class provides functionality for reading, parsing, modifying and writing
466 * CHAT files as used by in the CHILDES project.
467 * If parses the lines in the CHAT file and builds a structure from these so
468 * that every line has a parent showing its relations to other lines in the CHAT
469 * document. Headers and Participant lines are children of the ChatDocument,
470 * while the dependent tier lines are children of their headin Participant line.
471 *
472 * @package ChatTools
473 * @link http://childes.psy.cmu.edu/manuals/chat.pdf The manual for CHAT files
474 */
475 class ChatDocument extends ChatObject {
476     /****
477     * Filename of the CHAT file the ChatDocument operates on
478     * This should be set and retrieved using the setFilename() and
479     * getFilename() methods, which ensure that the file exists and is
480     * writable.
481     *
482     * @access protected
483     */
484     protected $filename;
485     /****
486     * Array of the header lines in the CHAT document
487     * This is an array of all the header lines in the CHAT document. It may be
488     * retrieved or modified using the setHeaderLines() and getHeaderLines()
489     * methods.
490     *
491     * @access protected
492     */
493     protected $header_lines;
494     /****
495     * Array of the participant lines in the CHAT document
496     * This is an array of all the participant lines in the CHAT document. These
497     * have the dependend tier lines as children. It may be retrieved or
498     * modified using the setPartLines() and getPartLines() methods.
499     *
500     * @access protected
501     */
502     protected $part_lines;
503 }
504
505
protected $part_lines;

Class Constructor

This is the class constructor. It takes one argument, which is the filename of the CHAT file that shall be manipulated. If you want to create a new CHAT file, you must first create an empty file which you can then manipulate with the class. The file must exist and be writable.

Note that the class does not automatically parse the file upon creation, so if it is not a new file you must still call the parseFile() method to parse it.

@access public

public function __construct($filename) {
    $this->setFilename($filename);
}

Set filename of CHAT document

This sets the filename of the CHAT document. It is automatically called when the object is created and may later be used to modify the filename, e.g. when you want to save the file under a different name after having manipulated it. The given file must both exist and be writable, if it is intended to be a new file, you must first create it.

@access public

public function setFilename($filename) {
    $this->checkFile($filename);
    $this->filename = $filename;
}

Get filename of CHAT document

This returns the filename currently used by the ChatDocument.

@access public

public function getFilename() {
    return $this->filename;
}

Set CHAT Header Lines for the CHAT document

This function lets you replace the complete set of header lines used by the CHAT document. It must be given as an indexed array, each item of which is a valid ChatHeaderLine object with this instance of the ChatDocument as its parent.
APPENDIX

```php
565  *
566  * @param $lines The array of ChatHeaderLine objects to be used
567  * @access public
568  * @return void
569  */
570  public function setHeaderLines(array $lines) {
571      foreach ($lines as $line) {
572          if (!is_a($line, 'ChatHeaderLine')) {
573              throw new InvalidArgumentException('The given array of ChatLine
574                  * headers contains members *
575                  * that are not valid ChatHeaderLine objects. ');
576          }
577      }
578      $this->header_lines = $lines;
579  }
580
581  /**
582   * Get CHAT Header Lines for the CHAT document
583   *
584   * This returns a numeric array of all the header lines of the CHAT document
585   *
586   * @return array An array of all the headers in the document
587   * @access public
588   */
589  public function getHeaderLines() {
590      return $this->header_lines;
591  }
592
593  /**
594   * Set CHAT Participant Lines for the CHAT document
595   *
596   * This function lets you replace the complete set of participant lines used
597   * by the CHAT document. It must be given as an indexed array, each item of
598   * which is a valid ChatPartLine object with this instance of the
599   * ChatDocument as its parent.
600   *
601   * @param $lines The array of ChatPartLine objects to be used
602   * @access public
603   * @return void
604   */
605  public function setPartLines(array $lines) {
606      foreach ($lines as $line) {
607          if (!is_a($line, 'ChatPartLine')) {
608              throw new InvalidArgumentException('The given array of *
609                  * participant ChatLines *
610                  * contains members that are *
611                  * not valid ChatPartLine *
612                  * objects. ');
613          }
614      }
615      $this->part_lines = $lines;
616  }
617
618  /**
619   * Get CHAT Participant Lines for the CHAT document
620   *
621   * This returns a numeric array of all the participant lines of the CHAT
622   * document
623   */
624  ```
public function getPartLines() {
    return $this->part_lines;
}

//∗∗∗
∗ Check whether the specified file exists and is writable
∗
∗ This method checks whether the file specified by $filename exists and is
∗ writable. The $filename argument is optional and if not given the
∗ current filename of the ChatDocument will be used instead.
∗
∗ @param string $filename The path to the file to check
∗ @return bool Returns true if the filename is valid, otherwise throws
∗ an InvalidArgumentException.
∗ @access protected
∗/
protected function checkFile($filename=null) {
    if ($filename == null) {
        $filename = $this->filename;
    }
    if (!file_exists($filename)) {
        throw new InvalidArgumentException('The specified file "' . $filename . '" does not exist. ');
    }
    if (!is_writable($filename)) {
        throw new InvalidArgumentException('The specified file "' . $filename . '" is not writeable. ');
    }
    return true;
}

//∗∗∗
∗ Parse the associated CHAT file
∗
∗ This method will parse the associated CHAT file (see $filename) and
∗ overwrite any current header and participant lines with those from the
∗ file. Note that the dependent tier lines are accessible through their
∗ parent ChatPartLine objects.
∗
∗ @return void
∗ @access public
∗/
public function parseFile() {
    $this->checkFile();
    $lines = file($this->filename);
    $last_part_line = false;
    for($i=0; $i<count($lines); $i++) {
        $lines[$i] = rtrim($lines[$i], '\r\n');
        switch($lines[$i][0]) {
        case '@':
            $x = new ChatHeaderLine($this, $lines[$i]);
            $x->setOrigLineNo($i+1);
            $this->header_lines[] = $x;
            break;
        case '*':
            break;
        }
    }

 APPENDIX

73
$last_part_line = new ChatPartLine($this, $lines[$i]);
$last_part_line->setOrigLineNo($i+1);
$this->part_lines[] = $last_part_line;
break;
case '%':
$x = new ChatDependentLine($last_part_line, $lines[$i], $last_part_line);
$x->setOrigLineNo($i+1);
break;
}
unset($x, $last_part_line);

/**
 * Base class for CHAT lines
 * This is an abstract class that provides some base functionality for all types
 * of CHAT lines: header lines, participant lines, and dependent tier lines.
 * These three different types of lines have their own respective classes
 * derived from this class: ChatHeaderLine, ChatPartLine and ChatDependentLine.
 * ChatLine cannot be used directly, but you can use it to check whether a given
 * object is any type of CHAT line with the is_a() function.
 */
abstract class ChatLine extends ChatObject {

/**
 * Reference to Parent ChatObject
 * This is a reference to the line's parent ChatObject. This may be another
 * ChatLine or a ChatDocument. The parent can only be set on construction
 * and thereafter not be modified. You can use the method getParent() to
 * obtain a reference to the parent item of a ChatLine.
 */
protected $parent;

/**
 * Identifier of the CHAT line
 * Every CHAT line has an identifier, usually inbetween one of @, *, % and
 * a colon :, these are three letters long for participant lines and
 * dependent tier lines and can be of varying length for header lines. A
 * special case are the Begin and End identifiers, which are not followed by
 * a colon in the CHAT file. Note that the identifier maintained here does
 * not have any of the @, *, % and : characters, since they are predictable
 * from the other properties of the object. So for "*EXA:" this would
 * contain the string "EXA", for "@Comment:" it would be 'Comment', etc.
 * This can be modified using the setIdentifier() and getIdentifier() method.
 */
protected $identifier;
APPENDIX

* Line data of the CHAT line
* This contains the actual data of the given CHAT line, i.e. what normally
* follows the identifier and a tab character. This would be things such as
* the actual transcription or the gloss text, depending on the type of CHAT
* line.
* You should use the setData() and getData() methods to modify this.
* @access protected
*
protected $data;

/****
* Original Line Number
*
* If the line was parsed from an existing CHAT file, then this contains the
* line number at which the line was originally positioned in the file when
* parsed. This can be useful for finding it in the raw file data if needed.
* If the ChatLine was not originally parsed from a file this is 0.
* Otherwise it will be any number of 1 or above.
* You may optionally use the setOrigLineNo() and getOrigLineNo() methods to
* modify this value.
* @access public
*/
public $orig_line_no = 0;

/****
* Constructor for ChatLine objects
*
* This is a generic constructor function for ChatLine objects. It takes the
* parent document and the raw line data (not the line data contained in the
* ChatLine object) as its argument.
*
* @param ChatDocument $parent The parent ChatDocument for the line
* @param string $data The raw, unparsed, line from the CHAT file
* @access public
* @return void
*/
public function __construct(ChatDocument $parent, $data) {
    $this->parent = $parent;
    $this->parseLine($data);
}

/****
* Set the Original Line Number
*
* This sets the original line number of the ChatLine object. This should be
* a reference to where the line was originally positioned in the CHAT file
* before parsing.
*
* @param int $line_no The line number of the line in the CHAT file
* @return void
* @access public
*/
public function setOrigLineNo($line_no) {
    $this->orig_line_no = (int) $line_no;
}

/****
* Get the Original Line Number

This returns the original line number of the ChatLine object. This is a reference to where the line was originally positioned in the CHAT file before parsing. This may be useful for looking up the raw data in the CHAT file.

@return int Returns the original line number
@access public
*/

public function getOrigLineNo() {
    return (int) $this->orig_line_no;
}

/**
 * Parse a line from raw data
 * This method parses the given data into the identifier and the line data and stores these in the present ChatLine object.
 * @param $data The unparsed, raw data from the CHAT file
 * @return void
 * @access protected
 */

protected function parseLine($data) {
    // Identifier: *XXX: -> XXX; %xxx: -> xxx; @x...x; -> x...x
    $this->identifier = substr($data, 1, strpos($data, ':')-1); //Remaining data on line
    $tier = substr($data, strpos($data, '\t')+1); //Eventually one could break down the individual items on the tier...
    $this->data = $tier;
}

/**
 * Get Parent ChatObject
 * This returns a reference to the parent ChatObject of the present ChatLine.
 * @return ChatObject Returns the parent ChatObject
 * @access public
 */

public function getParent() {
    return $this->parent;
}

/**
 * Get Identifier
 * This returns the identifier of the ChatLine object. See the description of the $identifier variable for more information on what this is.
 * @return string Returns the identifier of the CHAT line
 * @access public
 */

public function getIdentifier() {
    return $this->identifier;
}
Get Line Data

This method returns the line data for the present ChatLine. This is usually what comes behind the line identifier (e.g. "xyz in *EXA: xyz"). See the description of the $data variable for further information.

@return Returns the line data for the CHAT line
@access public

/public function getData()
    return $this->data;
}

CHAT Header Line Class

This class implements representations of header lines (lines beginning with @ in the CHAT file format). At present, it's functionality is identical to that of ChatLine and so its main use is for type hinting purposes.

/public class ChatHeaderLine extends ChatLine {
    // this does not provide any extra functionality to other ChatLines
}

CHAT Participant Line Class

This class implements representations of participant lines (lines that begin with an asterisk * in CHAT files). It extends the ChatLine object for some functionality relating to its ability to have subordained dependent tier lines.

/public class ChatPartLine extends ChatLine {
    // The Line's Dependent Tier
    *
    * This contains an array of all the dependent tier lines which are dependent on the participant line.
    *
    * @access protected
    */
    protected $dependent_lines = array();

    /**
    * Add a dependent line
    *
    * This adds a dependent tier line to the participant line. If the lines is already dependent on the participant line it the method call will be ignored as references are unique.
    *
    * @param ChatDependentLine $line The dependent tier line to be added
    * @return void
    * @access public
    */
    public function addDependentLine(ChatDependentLine $line) {
        if(!in_array($line, $this->dependent_lines) ) {
            // Add the dependent line
            $this->dependent_lines[] = $line;
        }
    }
```php
public function setDependentLines(array $lines) {
    foreach($lines as $line) {
        if(!is_a($line, 'ChatDependentLine')) {
            throw new InfiniteIterator('The given array of dependent ' .
                'ChatLines contains members that ' .
                'are not valid ChatDependentLine ' .
                'objects.');
        }
    }
    $this->dependent_lines = $lines;
}

public function getDependentLines() {
    return $this->dependent_lines;
}

/****
* Set all dependent lines
* This method is similar to addDependentLine() but it allows for the whole
* array of dependent tier lines to be replaced at once.
* @param array $lines An array of ChatDependentLine objects
* @return void
* @access public
*/

ChatDependentLine constructor
* This is the constructor for dependent tier lines. It behaves like the
* constructor for ChatLine but instead of a ChatDocument for the $parent
* parameter it expects a CharPartLine.
* @param ChatPartLine $parent The parent ChatPartLine for the line
* @param string $data The raw, unparsed, line from the CHAT file
```
```php
978     * @access public
979     * @return void
980     */
981     public function __construct(ChatPartLine $parent, $data) {
982         $this->parent = $parent;
983         $this->parent->addDependentLine($this);
984         $this->parseLine($data);
985     }
986 }
987 ?>
```
# B Stimuli for Judgement Experiment

Table B.1: Training Stimuli for Judgement Experiment

<table>
<thead>
<tr>
<th>#</th>
<th>Construction +A</th>
<th>–A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>*Dwyt Wyn ddim isio yfed bara brith o gwbl!</td>
<td>*Wyn ddim isio yfed bara brith o gwbl!</td>
</tr>
<tr>
<td></td>
<td>training</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Be’ oedd y derwydd yn licio mwyaf?</td>
<td>*Be’ yn derwydd y licio mwyaf?</td>
</tr>
<tr>
<td>3</td>
<td>Pryd wyt ti’n gorffen dy arholiad?</td>
<td>Pryd ti’n gorffen dy arholiad?</td>
</tr>
<tr>
<td>4</td>
<td>Mae Sioned yn gwerthu lot o stwff yn y farchnad.</td>
<td>*Sioned yn gwerthu lot o stwff yn y farchnad.</td>
</tr>
<tr>
<td>5</td>
<td>*Wnâth hi ddim cael lle oddi wrth y stafell.</td>
<td>Doedd gynni hi ddim lle yn eu ’stafell.</td>
</tr>
</tbody>
</table>

Table B.2: Test Stimuli for Judgement Experiment

GROUP I (Grammatical Person/Number)

<table>
<thead>
<tr>
<th>#</th>
<th>Construction +A</th>
<th>–A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1S Dw i’n licio hufen iâ. Fi’n licio hufen iâ.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Dw i’n byw ym Mangor.</td>
<td>Fi’n byw ym Mangor.</td>
</tr>
<tr>
<td>3</td>
<td>2S #43</td>
<td>#43</td>
</tr>
<tr>
<td>4</td>
<td>#44</td>
<td>#44</td>
</tr>
<tr>
<td>5</td>
<td>3Sf Mae hi’n byw yng Nghaerdydd.</td>
<td>Hi’n byw yng Nghaerdydd.</td>
</tr>
<tr>
<td></td>
<td>Mae hi’n astudio Seicoleg yn y Brifysgol.</td>
<td>Hi’n astudio Seicoleg yn y Brifysgol.</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3Sm Mae o’n dod o Ryl yn wreiddiol.</td>
<td>O’n dod o Ryl yn wreiddiol.</td>
</tr>
<tr>
<td>8</td>
<td>Mae o’n yfed lot o gwrw.</td>
<td>O’n yfed lot o gwrw.</td>
</tr>
</tbody>
</table>
### APPENDIX

<table>
<thead>
<tr>
<th>#</th>
<th>Construction</th>
<th>+A</th>
<th>-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Dan ni’n neidio o gwmpas ar y gwely.</td>
<td>Ni’n neidio o gwmpas ar y gwely.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dan ni’n licio mynd i Sbaen.</td>
<td>Ni’n licio mynd i Sbaen.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Dach chi’n dadlau â’r bobl drws nesaf bob hyn a hyn.</td>
<td>Chi’n dadlau â’r bobl drws nesaf bob hyn a hyn.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dach chi’n byw’n bell o bob man.</td>
<td>Chi’n byw’n bell o bob man.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Maen nhw’n mynd am dro ar y traeth.</td>
<td>Nhw’n mynd am dro ar y traeth.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maen nhw’n gwisgo yr un crys-t.</td>
<td>Nhw’n gwisgo yr un crys-t.</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Mae’r plant yn chwarae efo ffrind.</td>
<td>Y plant yn chwarae efo ffrind.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mae’r gath yn yfed llefrith.</td>
<td>Y gath yn yfed llefrith.</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Mae Sian yn licio pêl-droed yn fawr.</td>
<td>Sian yn licio pêl-droed yn fawr.</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Bod ti’n siopa am oriawr dwy flynedd yn ól.</td>
<td>Ti’n siopa am oriawr dwy flynedd yn ól.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roeddet ti’n ymweld â dy nain di wythnos dwytha’.</td>
<td>Ti’n ymweld â dy nain di wythnos dwytha’.</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Roeddet ti’n ffonio fi neithiwr.</td>
<td>Ti’n ffonio fi neithiwr.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roeddet ti’n canu mewn côr yn blentyn.</td>
<td>Ti’n canu mewn côr yn blentyn.</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Bod ti’n siopa am oriawr dwy flynedd yn ól.</td>
<td>Ti’n siopa am oriawr dwy flynedd yn ól.</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Roeddet ti’n ymweld â dy nain di wythnos dwytha’.</td>
<td>Ti’n ymweld â dy nain di wythnos dwytha’.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roeddet ti’n ffonio fi neithiwr.</td>
<td>Ti’n ffonio fi neithiwr.</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Roeddet ti’n canu mewn côr yn blentyn.</td>
<td>Ti’n canu mewn côr yn blentyn.</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Byddi di’n galw dy daid di ’fory.</td>
<td>Ti’n galw dy daid di ’fory.</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>#43</td>
<td>#43</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>#44</td>
<td>#44</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>#45</td>
<td>#45</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>#46</td>
<td>#46</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>#43</td>
<td>#43</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>#44</td>
<td>#44</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>#45</td>
<td>#45</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>#46</td>
<td>#46</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX

28  Byddi di’n siarad efo fy athro i wythnos nesaf.
     Ti’n siarad efo fy athro i wythnos nesaf.
29  Byddi di’n chwarae pêl-fasged efo Dylan nes ymlaen.
     Ti’n chwarae pêl-fasged efo Dylan nes ymlaen.
30  Byddi di’n mynd ar wyliau yn y Swistir flwyddyn nesaf.
     Ti’n mynd ar wyliau yn y Swistir flwyddyn nesaf.
31  Wnest ti neud dy waith cartref di yn dda ddoe.
     Ti neud dy waith cartref di yn dda ddoe.
32  Wnest ti fwydo’r planhigion mwy nag wythnos yn ôl.
     Ti fwydo’r planhigion mwy nag wythnos yn ôl.
33  Wnest ti enill y gêm tro dwytha’.
     Ti enill y gêm tro dwytha’.
34  Wnest ti ateb y neges John neithiwr.
     Ti ateb y neges John neithiwr.
35  Wnest ti ’sgubo ’fory.
     Ti ’sgubo ’fory.
36  Wnest ti drwsio’r ffenest ac wna i drwsio’r drws.
     Ti drwsio’r ffenest ac wna i drwsio’r drws.
37  Wnest ti baratoi’r bwyd ar gyfer y parti wythnos nesaf.
     Ti baratoi’r bwyd ar gyfer y parti wythnos nesaf.
38  Wnest ti nôl y plant o’r ysgol yfory.
     Ti nôl y plant o’r ysgol yfory.
39  Rwyt ti ’di gwastrafi amser.
     Ti ’di gwastrafi amser.
40  Rwyt ti ’di golchi’r llestri.
     Ti ’di golchi’r llestri.
41  Rwyt ti ’di darllen yr holl lyfr.
     Ti ’di darllen yr holl lyfr.
42  Rwyt ti ’di enill y cwis tafarn.
     Ti ’di enill y cwis tafarn.

GROUP III (Mood)

<table>
<thead>
<tr>
<th>#</th>
<th>Construction</th>
<th>+A</th>
<th>–A</th>
</tr>
</thead>
<tbody>
<tr>
<td>43</td>
<td>bod+AFF</td>
<td>Rwyt ti’n chwarae tennis yn dda.</td>
<td>Ti’n chwarae tennis yn dda.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rwyt ti’n tecstio at dy ffrindiau yn aml.</td>
<td>Ti’n tecstio at dy ffrindiau yn aml.</td>
</tr>
</tbody>
</table>
## APPENDIX

<table>
<thead>
<tr>
<th>#</th>
<th>Construction +A</th>
<th>-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>Ti’n gwrando at Radio Cymru.</td>
<td>Ti’n gwrando at Radio Cymru.</td>
</tr>
<tr>
<td>46</td>
<td>Ti’n eistedd yn y ’stafell fwy.</td>
<td>Ti’n eistedd yn y ’stafell fwy.</td>
</tr>
<tr>
<td>47</td>
<td>Ti’n bwyta cinio rwan?</td>
<td>Ti’n bwyta cinio rwan?</td>
</tr>
<tr>
<td>48</td>
<td>Ti’n licio mynd am dro?</td>
<td>Ti’n licio mynd am dro?</td>
</tr>
<tr>
<td>49</td>
<td>Ti’n cael cawod heno?</td>
<td>Ti’n cael cawod heno?</td>
</tr>
<tr>
<td>50</td>
<td>Ti’n dilyn Pobol y Cwm ar S4C?</td>
<td>Ti’n dilyn Pobol y Cwm ar S4C?</td>
</tr>
<tr>
<td>51</td>
<td>Ti ddim yn yfed lot fel arfer.</td>
<td>Ti ddim yn yfed lot fel arfer.</td>
</tr>
<tr>
<td>52</td>
<td>Ti ddim yn bwyta sioeled.</td>
<td>Ti ddim yn bwyta sioeled.</td>
</tr>
<tr>
<td>53</td>
<td>Ti ddim yn siarad efo Glyn rhagor.</td>
<td>Ti ddim yn siarad efo Glyn rhagor.</td>
</tr>
<tr>
<td>54</td>
<td>Ti ddim yn cael mynd adra eto.</td>
<td>Ti ddim yn cael mynd adra eto.</td>
</tr>
</tbody>
</table>

### GROUP IV (Focus and Subject/Object-Movement)

<table>
<thead>
<tr>
<th>#</th>
<th>Construction +A</th>
<th>-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>AuxSVO #43 Y ti sy’n coginio cinio heno.</td>
<td>Y ti’n coginio cinio heno.</td>
</tr>
<tr>
<td>56</td>
<td>AuxSVO #44 Y ti sy’n rhoi’r ddarlith.</td>
<td>Y ti’n rhoi’r ddarlith.</td>
</tr>
<tr>
<td>57</td>
<td>AuxSVO #45 Y ti sy’n casglu’r plant o’r ysgol.</td>
<td>Y ti’n casglu’r plant o’r ysgol.</td>
</tr>
<tr>
<td>58</td>
<td>AuxSVO #46 Y ti sy’n dod â photel o win.</td>
<td>Y ti’n dod â photel o win.</td>
</tr>
<tr>
<td>59</td>
<td>VOAuxS Ffonio’r gwasanaeth tân wyt ti.</td>
<td>Ffonio’r gwasanaeth tân ti.</td>
</tr>
<tr>
<td>60</td>
<td>VOAuxS Mynd i’r cigydd wyt ti.</td>
<td>Mynd i’r cigydd ti.</td>
</tr>
<tr>
<td>61</td>
<td>VOAuxS Siarad efo ffrind wyt ti.</td>
<td>Siarad efo ffrind ti.</td>
</tr>
<tr>
<td>63</td>
<td>Dillad wyt ti’n prynu.</td>
<td>Dillad ti’n prynu.</td>
</tr>
</tbody>
</table>
### APPENDIX

<table>
<thead>
<tr>
<th>#</th>
<th>Construction</th>
<th>+A</th>
<th>–A</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>I’r canolfan hamdden wyt ti’n mynd.</td>
<td>I’r canolfan hamdden ti’n mynd.</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>Paned wyt ti’n yfed.</td>
<td>Paned ti’n yfed.</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Yr heddlu wyt ti’n osgoi.</td>
<td>Yr heddlu ti’n osgoi.</td>
<td></td>
</tr>
</tbody>
</table>

**GROUP V (Subordinates)**

<table>
<thead>
<tr>
<th>#</th>
<th>Construction</th>
<th>+A</th>
<th>–A</th>
</tr>
</thead>
<tbody>
<tr>
<td>71</td>
<td>–infl Dw i’n meddwl fod ti’n gyrru yn dda.</td>
<td>Dw i’n meddwl ti’n gyrru yn dda.</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>Mae’n bosib fod ti’n gwestud gormod.</td>
<td>Mae’n bosib ti’n gwestud gormod.</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>Mae Alun yn credu fod ti’n gofnyn am lawer.</td>
<td>Mae Alun yn credu ti’n gofnyn am lawer.</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>Dan ni’n gobeithio fod ti’n medru enill.</td>
<td>Dan ni’n gobeithio ti’n medru enill.</td>
<td></td>
</tr>
</tbody>
</table>

**GROUP VI (Wh-Questions)**

<table>
<thead>
<tr>
<th>#</th>
<th>Construction</th>
<th>+A</th>
<th>–A</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>WH Sut wyt ti’n mynd i Wrecsam?</td>
<td>Sut ti’n mynd i Wrecsam?</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>Be’ wyt ti’n deud wrth Gwen am y ddamwain?</td>
<td>Be’ ti’n deud wrth Gwen am y ddamwain?</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>Pwy wyt ti’n ei warhodd?</td>
<td>Pwy ti’n ei warhodd?</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>Pryd wyt ti’n symd i’r Drenwydd?</td>
<td>Pryd ti’n symd i’r Drenwydd?</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>WH+Prep Lle wyt ti’n mynd i?</td>
<td>Lle ti’n mynd i?</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>Be’ wyt ti’n golchi dy gi efo?</td>
<td>Lle ti’n golchi dy gi efo?</td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>Lle wyt ti’n mynd i yn nes ymlaen?</td>
<td>Lle ti’n mynd i yn nes ymlaen?</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>Pwy wyt ti’n siarad efo?</td>
<td>Pwy ti’n siarad efo?</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>Prep+WH O le wyt ti’n dod yn wreiddiol?</td>
<td>O le ti’n dod yn wreiddiol?</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>Efo pwy wyt ti’n dawnsio?</td>
<td>Efo pwy ti’n dawnsio?</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>Ers pryd wyt ti’n byw yma?</td>
<td>Ers pryd ti’n byw yma?</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>Ar pa’ gwch wyt ti’n mynd?</td>
<td>Ar pa’ gwch ti’n mynd?</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>--------------------------</td>
<td>-----------------------</td>
<td></td>
</tr>
</tbody>
</table>

C Program Code Listings for Judgement Experiment

Listing 3: OpenSesame script for judgement experiment

```python
# Generated by OpenSesame 0.26 (Earnest Einstein)
# Thu Apr 26 19:59:57 2012 (nt)
# Copyright Sebastiaan Mathot (2010−2011)
# <http://www.cogsci.nl>
#
# NOTE: This script was edited in order to skip several hundred lines which are
# automatically generated by the script waveDur.php. The two points at
# which this happened (the training and experimental loops) have been
# marked by a comment with reference to that script’s output in this
# source file and need to be replaced with the script’s output before the
# experiment can be run.

set foreground "white"
set subject_parity "even"
set description "Default description"
set title "AD Experiment"
set sampler_backend "legacy"
set coordinates "relative"
set height "768"
set mouse_backend "psycho"
set width "1024"
set compensation "0"
set keyboard_backend "psycho"
set background "black"
set subject_nr "0"
set canvas_backend "psycho"
set start "experiment"
set synth_backend "legacy"

define inline_script set_response_timeout
    set __run ""
    ___prepare__
    cue_duration = self.get('cue_duration')
    self.experiment.set('response_timeout', cue_duration+1500)
    ___end__
    set description 'Executes Python code'

define inline_script stop_playback
    ___run__
    import pygame
    pygame.mixer.stop()
    ___end__
    set __prepare ""
    set description 'Executes Python code'

define text_display instructions_1
    set foreground "white"
    set font_size '18'
```
APPENDIX

Welcome to the auditory judgement task!

During this experiment, you will hear a short beep followed by a sentence in Welsh.

Some of these sentences are perfectly fine colloquial Welsh sentences, as you could possibly hear them somewhere in the street. However, some of the sentences were changed and probably don’t sound right to you.

Your task is to listen carefully to all the sentence and decide as quickly as you can whether you think that what you’ve just heard is an acceptable example of a colloquial Welsh sentence or not.

If you think it is okay, you should press the right (M) key — but if you think it doesn’t really feel right to you, press the left (Z) key!

(Press any key for more instructions...)

Don’t worry whether you think the sentences are “proper Welsh” — most of them aren’t, and we don’t really care. What we want to know about is your personal intuition, what you would think if you heard this in real life. So remember that you are the real expert in this experiment!

We will now first give you 10 sentences to practice, as this task takes a little getting used to at first. After this you will have the chance to take a little break (as at several points during the experiment!) before the real thing starts.

Should you have any problems you can ask the researcher for help during the break.

To start the practice session press any key...

We will now first give you 10 sentences to practice, as this task takes a little getting used to at first. After this you will have the chance to take a little break (as at several points during the experiment!) before the real thing starts.

Should you have any problems you can ask the researcher for help during the break.

To start the practice session press any key...
APPENDIX

```
define inline_script experimental_loop_count
   set _run ''
   ___prepare___
   # Loop counter
   if self.has( 'experimental_loop_counter' ):
       loop_counter = self.get( 'experimental_loop_counter' )
       self.experiment.set( 'experimental_loop_counter', loop_counter +1)
   else:
       self.experiment.set( 'experimental_loop_counter', 0)
   ___end___
   set description 'Executes Python code'

define text_display thank_you
   set foreground 'white'
   set font_size '18'
   set description 'Presents a display consisting of text'
   set maxchar '50'
   set align 'center'
   ___content___
   Thank you!
   ___end___
   set background 'black'
   set duration 'keypress'
   set font_family 'mono'

define sequence stimulus_presentation
   run set_response_timeout 'always'
   run fixation_dot 'always'
   run pre_beep_delay 'always'
   run beep 'always'
   run post_beep_delay 'always'
   run show_keys 'always'
   run stimuli 'always'
   run keyboard_response 'always'
   run stop_playback 'always'
   run logger 'always'

define loop training_loop
   set repeat '1'
   set description 'Repeatedly runs another item'
   set skip '0'
   set offset 'no'
   set item 'stimulus_presentation'
   set column_order 'cue_no;cue_condition;cue_file;cue_duration'
   set cycles '10'
   set order 'random'
   # INSERT RESULTS FROM WAVEDUR.PHP SCRIPT FOR ./TRAIN DIRECTORY HERE
   #
   run stimulus_presentation
```

88
define sketchpad take_a_break_2
set duration 'keypress'
set description 'Displays stimuli'
set start_response_interval 'no'
draw textline 0 −96 'Well done! You’ve completed \[
\text{experimental_loop_counter}/152 \text{ items now.}'
center=1 color=white
font_family=mono font_size=18 show_if='always'
draw textline 0 −32 'Time to take a little break...' center=1 color=
white font_family=mono font_size=18 show_if='always'
draw textline 0 32 'Just press any button to continue when you’re ready !'
center=1 color=white font_family=mono font_size=18 show_if='always'
define loop experimental_loop
set item 'sequence'
set cycles '152'
set column_order 'cue_no;cue_condition;cue_file;cue_duration'
# INSER T RESULTS FROM WAVEDUR.PHP SCRIPT FOR ./FINAL DIRECTORY HERE
#
run experimental_sequence

define sampler beep
set volume '0.3'
set description 'Plays a sound file in .wav or .ogg format'
set sample 'beep.wav'
set pitch '1'
set duration 'sound'
set stop_after '0'
set pan '0'
set fade_in '0'
define sequence experiment
run training 'always'
run experimental_loop 'always'
run thank_you 'always'
run exit 'always'
define keyboard_response exit
set allowed_responses 'q'
set description 'Collects keyboard responses'
set timeout 'infinite'
set flush 'yes'
define advanced_delay post_beep_delay
set duration '100'
set jitter '0'
set description 'Waits for a specified duration'
set jitter_mode 'Std. Dev.'
define fixation dot fixation dot
set foreground 'white'
set style 'cross'
set description 'Presents a central fixation dot with a choice of
various styles'
set y '0'
set background 'black'
set duration '500'
APPENDIX

```
set x '0'
set penwidth '3'
define logger logger
  set ignore_missing 'yes'
  set unicode 'no'
  set description 'Logs experimental data'
  set use_quotes 'yes'
  set auto_log 'no'
  log 'cue_duration'
  log 'cue_condition'
  log 'cue_file'
  log 'cue_no'
  log 'response_keyboard_response'
  log 'response_time_keyboard_response'
define sequence experimental_sequence
  run experimental_loop_count 'always'
  run take_a_break ' [experimental_loop_counter] = 38 ' [experimental_loop_counter] = 38
  run take_a_break ' [experimental_loop_counter] = 76 ' [experimental_loop_counter] = 76
  run take_a_break ' [experimental_loop_counter] = 114 ' [experimental_loop_counter] = 114
  run stimulus_presentation 'always'
define advanced_delay pre_beep_delay
  set duration '100'
  set jitter '0'
  set description 'Waits for a specified duration'
  set jitter_mode 'Std. Dev.'
define text_display end_of_training
  set foreground 'white'
  set font_size '18'
  set description 'Presents a display consisting of text'
  set maxchar '50'
  set align 'center'
  __content__

Well done, you've completed the training task.
Feel free to take a little break now!

When you think you are ready just press any key to start the experiment

__end__
  set background 'black'
  set duration 'keypress'
  set font_family 'mono'
define sequence training
  run instructions_1 'always'
  run instructions_2 'always'
  run training_loop 'always'
  run end_of_training 'always'
define text_display take_a_break
  set foreground 'white'
  set font_size '18'
  set description 'Presents a display consisting of text'
  set maxchar '50'
  set align 'center'
  __content__
```

90
APPENDIX

Well done! You’ve completed the first experimental_loop_counter out of 152 items.

Time to take a little break...

Press any button to continue with the experiment when you are ready!

__end__

set background "black"
set flush 'yes'
set duration "keypress"
set font_family "mono"

define sampler stimuli
set sample "./audio/[cue_file].wav"
set description "Plays a sound file in .wav or .ogg format"
set volume '1'
set timeout "0"
set pitch '1'
set duration "0"
set stop_after "0"
set pan '0'
set fade_in '0'

define keyboard_response keyboard_response
set allowed_responses "z;m"
set description "Collects keyboard responses"
set timeout "[response_timeout]"
set flush 'yes'

Listing 4: Script for calculating duration of waveform files
APPENDIX

```php
error_reporting(E_ALL);
ini_set('display_errors', 1);

// Paths
$root_path = "./";

// MAIN SCRIPT

// Check command line argument is okay
if(isset($argv[1]) && ($argv[1] == 'TRAIN' || $argv[1] == 'FINAL')) {
  $type = $argv[1];
} else {
  die('First argument must be either TRAIN or FINAL.');
}

// Scan all files and print info
$dir = scandir($root_path.$type);
$i = 0;
foreach($dir as $file) {
  if(substr($file, -4) == " . wav") {
    $dur = (int) wavDur("./$type/".$file);
    $cond = substr($file, 2, 2);
    $no = substr($file, 0, 2);
    print "setcycle $i cue_no \$no\"n"
    print "setcycle $i cue_condition \$cond\"n"
    print "setcycle $i cue_file \$type/$no$cond\"n"
    print "setcycle $i cue_duration \$dur\"n"
    $i++;
  }
}

// FUNCTIONS

///

/**
 * Read Header and Duration from RIFF Waveform Files
 * This function reads the header information from a RIFF Waveform file and then calculates the file duration (for the audio content) in milliseconds.
 * The function is adapted from an earlier code snippet posted by "valinea" on 07 August 2006 at http://snipplr.com/view/285/.
 * @link http://snipplr.com/view/285/
 * @author valinea (http://snipplr.com/users/valinea/)
 * @param string $file Path to the waveform file to be analysed
 * @return int Returns the duration of the waveform in seconds.
 */

function wavDur($file) {
  $fp = fopen($file, 'r');
  if(fread($fp, 4) == "RIFF") {
    fseek($fp, 20);
    $rawheader = fread($fp, 16);
    $packing = 'vtype/vchannels/Vsamplerate/Vbytespersec/valignment/vbits';
    $header = unpack($packing, $rawheader);
    $pos = ftell($fp);
    while(fread($fp, 4) != "data" && !feof($fp)) {
```

92
Listing 5: Script for generating offline questionnaire

```php
<?php

// SETUP

// Some PHP stuff
error_reporting(E_ALL);
ini_set('display_errors', 1);

// Paths
$stimuli_list = './stimuli.php';
$output_file = './quest.html';

// MAIN SCRIPT

// Fetch stimuli list
require($stimuli_list);
if (!isset($TRAIN) || !isset($FINAL)) {

```

```


APPENDIX

die (' Stimuli set is either missing $TRAIN or $FINAL data structures. ');

// Start output buffering and print HTML header
ob_start ( ) ;
print '<?xml version='1.0' encoding='UTF−8'?>';
print <<<HTML
<html>
<head>
<meta http-equiv='Content-Type' content='text/html; charset=utf−8'>
<title>AD EXPERIMENT &mdash ; OFFLINE JUDGEMENT TASK</title>
<style>
body {
  font−size:12pt;
  font−family:serif;
  padding:0px;
  margin:0px;
}
h1 {
  font−size:150%;
}
h2 {
  font−size:115%;
  page−break−before:auto;
  page−break−after:avoid;
}
h2 .break {
  page−break−before:always;
}
div p {
  display:inline−block;
  margin:2px;
}
div p .id {
  width:auto;
  font−size:60%;
  font−family:monospace;
  color:#777;
}
div p .stimuli {
  width:70%;
}
div p .likert {
  width:auto;
}
div p .likert span {
  display:inline−block;
  width:22px;
  text−align: center;
}
</style>
</head>
<body>
<h1>AD EXPERIMENT &mdash ; OFFLINE JUDGEMENT TASK</h1>
<h2>Instructions</h2>
<p>
In this task you are presented with a number of colloquial Welsh sentences similar to those you’ve heared in the computer−aided task you have just completed. Again some of these sentences will be just
fine and some of them will probably seem rather odd to you.

As opposed to the previous task however, this time we want you to rate how acceptable the given sentences seem to you. For this you will see a five point scale next to every sentence. You should use the number 1 to indicate that you feel the sentence is completely unacceptable and the number 5 to indicate that it feels completely acceptable to you. Use any of the numbers in-between to indicate that you have a tendency to say it is acceptable or unacceptable, or the box in the middle if you cannot decide at all.

Again this is about what you feel is appropriate in the colloquial, spoken language and that this is not about what you may have been taught about Welsh in school. As you will surely know sometimes what people do can be very different from what they teach! So remember that this is about your personal opinion about the language you speak and so you are the real expert!

HTML;

// Print likert scale for training data
print <<<HTML
<h2>Block 1</h2>
<div>
<p class="id"></p>
<p class="stimuli"></p>
<p class="likert">
<span>1</span>
<span>2</span>
<span>3</span>
<span>4</span>
<span>5</span>
</p>
</div>

HTML;

// Randomise and print training stimuli
$training_data = rand_stimuli($TRAIN);
foreach($training_data as $stimulus) {
$id = hide_id($stimulus[0], $stimulus[1]);
$sentence = $stimulus[2];
print <<<HTML
<div>
<p class="id">$id</p>
<p class="stimuli">$sentence</p>
<p class="likert">
<span>&#9744;</span>
<span>&#9744;</span>
<span>&#9744;</span>
<span>&#9744;</span>
<span>&#9744;</span>
</p>
</div>

HTML;
$experimental_data = rand_stimuli($FINAL);
$block_counter = 2; // Block 1 were the Training stimuli

// Print experimental stimuli in blocks of 30
for ($i=0; $i< count($experimental_data); $i++) {
    if ($i%30 == 0) {
        // Print block header with numbers for likert scale
        print <<<HTML
            <h2 class="break">Block $block_counter</h2>
            <div>
                &nbsp;&nbsp;&nbsp;&nbsp;
                &nbsp;&nbsp;&nbsp;&nbsp;
                &nbsp;&nbsp;&nbsp;&nbsp;
                &nbsp;&nbsp;&nbsp;&nbsp;
                &nbsp;&nbsp;&nbsp;&nbsp;
            </div>
        HTML;
        $block_counter++;
    }
    $stimulus = $experimental_data[$i];
    $id = hide_id($stimulus[0], $stimulus[1]);
    $sentence = $stimulus[2];
    // Print stimuli and likert scale
    print <<<HTML
        <div>
            <p class="id">$id</p>
            <p class="stimuli">$sentence</p>
            <p class="likert">&amp;#9744;&amp;#9744;&amp;#9744;&amp;#9744;&amp;#9744;</p>
        </div>
    HTML;
}

// Print HTML footer
print <<<HTML
</body>
</html>
HTML;

// Write output buffer to output file
$ob = ob_get_contents();
$fh = fopen($output_file, 'w+');
fwrite($fh, $ob);close($fh);
APPENDIX

//NB: Output buffer will be flushed to STDOUT at end of script!

// FUNCTIONS

//∗∗∗
//Randomise list of stimuli
//∗
//This function takes an array of stimuli in two conditions and merges these
//into one single list, assigning a pseudo random order to every single item.
//∗
//∗param array $stimuli A two-dimensional array of stimuli to be randomised
//∗return array Returns a flat array of stimuli in pseudo random order
//∗/

function rand_stimuli($stimuli) {
    $keys = array_keys($stimuli); // get all keys
    $keys = array_merge($keys, $keys); // double keys (+A and −A conditions)
    shuffle($keys); // pseudo-randomisation
    $result = array();
    foreach ($keys as $key) {
        if (count($stimuli[$key]) > 1) {
            $cond = rand(0, 1) ? 'ad' : 'oa'; // pseudo-random +A or −A
            if (substr($stimuli[$key][$cond], 0, 1) == '#') {
                continue; // skip references to #xx
            }
            $result[] = array($key, $cond, $stimuli[$key][$cond]);
            unset($stimuli[$key][$cond]);
        } else {
            $cond = array_keys($stimuli[$key]); // either oa or ad
            $cond = $cond[0];
            if (substr($stimuli[$key][$cond], 0, 1) == '#') {
                continue; // skip references to #xx
            }
            $result[] = array($key, $cond, $stimuli[$key][$cond]);
            unset($stimuli[$key][$cond]);
        }
    }
    return $result;
}

//∗∗∗
//Hide Stimuli ID and condition
//∗
//This function takes the id and condition description (oa or ad) from a
//stimuli and returns a string masking these in some predetermined pseudo
//random numbers, which can be converted back into the original stimuli
//∗
//∗ID and condition. Specifically a pseudo random number between 0 and 4 is
//∗assigned to the condition 'ad' and one between 5 and 9 to 'oa'. This is
//∗followed by the stimuli ID, with a leading zero where applicable. Another
//∗pseudo random number between 0 and 9 is appended at the end.
//∗
//∗param string $cond A string indicating the experimental condition, either
//∗'oa' (+A) or 'ad' (−A)
//∗param int $id The stimuli ID
//∗return string Returns a string of numbers encoding condition and ID
//∗/

function hide_id($id, $cond) {

97
if($cond == 'ad') {
    $p = rand(0, 4); //0–4 to mark -A
} else {
    $p = rand(5, 9); //5–9 to mark +A
}
if(strlen($id) == 1) {
    $id = '0'. $id; //enforce preceeding 0
}
$t = rand(0, 9); //random trailing number
return $p.$id.$t;
?>

Listing 6: Sample format of stimuli.php file

```php
<?php
/
∗∗∗
∗ Sample of Format for stimuli.php
∗
∗ This file contains a sample of the formatting in which the file
∗ stimuli.php, used by the make_questionnaire.php script, should be.
∗ This should contain two arrays, $TRAIN for the training stimuli and
∗ $FINAL for the test stimuli, both following the format indicated
∗ below. The text may include references to other stimuli in the form
∗ #NN giving their index number in the array, these will be skipped in
∗ the final output.
∗
∗ PHP Version 5.3
∗
∗ LICENSE: This piece of software was developed as part of a BA (Hons)
∗ dissertation at Bangor University. It may be freely distributed and used by
∗ anybody whomever, so long as the author is acknowledged and no changes
∗ are made to the source code without prior agreement with the author.
∗
∗ @author Florian Breit <f.breit@univ.bangor.ac.uk>
∗ @copyright 2012 Florian Breit
∗ @version 1.0.0
∗ /

$TRAIN = array(1 => array('oa' => 'Dwyt Wyn ddim isio yfed bara brith o gwbl!',
    'ad' => 'Wyn ddim isio yfed bara brith o gwbl!'),
    2 => array('oa' => 'Next sentence in overt auxiliary condition',
    'ad' => 'Next sentence in auxiliary deletion condition'),
    3 => array('oa' => '...',
    'ad' => '...',
    //etc.
    );

?>

APPENDIX

98
Listing 7: Script for entering background data

```html
<html>
<head>
<title>Participant Background Data</title>
<style type="text/css">
input {
  margin:1px;
  padding:2px;
  border:1px solid #999;
}
input:active, input:focus {
  border:1px solid black;
  background:#FFA;
}
label[for] {
  font-weight: bold;
  min-width:40px;
  display:inline-block;
}
</style>
</head>
<body>
<?php
error_reporting(E_ALL);
if (!empty($_POST)) {
    $code = str_replace(array('\', '/'), '', $_POST['code']);
    $fh = fopen('./results/background-' . $code . '.csv', 'w+');
    fwrite($fh, 'age,gender,education,wherefrom,southnorth\r\n');
    $age = $_POST['age'];
    $gender = $_POST['gender'];
    $education = $_POST['education'];
    $wherefrom = $_POST['wherefrom'];
    $southnorth = $_POST['southnorth'];
    fwrite($fh, '$age\r\n');
    fwrite($fh, '$gender\r\n');
    fwrite($fh, '$education\r\n');
    fwrite($fh, '$wherefrom\r\n');
    fwrite($fh, '$southnorth\r\n');
    fclose($fh);
    print('Data written to file: background-' . $code . '.csv<br />');
    print('<iframe src=./results/background-' . $code . '.csv'></iframe><br />');
}
</html>
</body>
```

<h1>Participant Background Data</h1>
<form method="post">
  <label for="age">Code</label>
  <input type="text" name="code" size="4" />
  <br />
  <label for="age">Age</label>
  <input type="text" name="age" size="2" />
  <br />
  <label for="gender">Gender</label>
  <input type="radio" name="gender" value="1" />
  <br />
  <label>Male</label>
  <input type="radio" name="gender" value="2" />
  <br />
  <label for="education">Education</label>
  <input type="radio" name="education" value="1" />
  <br />
  <label>GCESEs</label>
  <input type="radio" name="education" value="2" />
</form>

APPENDIX

Listing 8: Script for entering questionnaire results

```php
<?php
function resolve_id($hidden_id) {
    // resolve condition (ad < 5 => od)
    $cond = substr($hidden_id, 0, 1);
    if ($cond < 5) {
        $cond = 'ad';
    } else {
        $cond = 'oa';
    }
    // extract id
    $id = (int) substr($hidden_id, 1, 2);
    return array('id' => $id,
                  'cond' => $cond,
                  0 => $id,
                  1 => $cond);
}
?>
</html>
```

<input type="radio" name="education" value="3" />
<label>(Some) HE</label><br />
<input type="radio" name="education" value="4" />
<label>(Some) PG Ed</label><br />
<label for="wherefrom">Where are you from?</label>
<input type="text" name="wherefrom" size="20" /><br />
<label for="southnorth">South/Mid/North-Walian?</label>
<input type="radio" name="southnorth" value="1" />
<label>South</label><br />
<input type="radio" name="southnorth" value="2" />
<label>Mid</label><br />
<input type="radio" name="southnorth" value="3" />
<label>North</label><br />
<button type="submit">Submit</button>

```

<!-- Output goes here -->
```
<?php
error_reporting(E_ALL);
$train_results = array();
$final_results = array();
if (!empty($$_POST)) {
    $code = str_replace(array(‘\’, ‘/’), ‘,’ . $_POST[‘code’]);
    $fh = fopen('.results/offline-train-$code.csv', ‘w+’);
    fwrite($fh, ‘id,cond,rating\n’);
    foreach($_POST[‘train_id’] as $index => $id) {
        list($id, $cond) = resolve_id($id);
        $value = $_POST[‘train_value’][$index];
        fwrite($fh, ‘"$id"","$cond","$value"\n’);
    }
    fclose($fh);
    print('Data written to file: offline-train-$code.csv<br />');
    print('<iframe src='./results/offline-train-$code.csv'></iframe><br />');
    $fh = fopen('.results/offline-$code.csv', ‘w+’);
    fwrite($fh, ‘id,cond,rating\n’);
    foreach($_POST[‘id’] as $index => $id) {
        list($id, $cond) = resolve_id($id);
        $value = $_POST[‘value’][$index];
        fwrite($fh, ‘"$id"","$cond","$value"\n’);
    }
    fclose($fh);
    print('Data written to file: offline-$code.csv<br />');
    print('<iframe src='./results/offline-$code.csv'></iframe><br />');
}
<form method=‘post’>
    <b>Code:</b> <input type=‘text’ name=‘code’ />
</form>
<h2>Block 1</h2>
<ol>
    <?php
    for ($i=0; $i<10; $i++) { ?>
        <li>
            <input type=‘text’ class=‘id’ name=‘train_id[]’ />
            <input type=‘text’ class=‘value’ name=‘train_value[]’ />
        </li>
    <?php } ?>
</ol>
<h2>Block 2</h2>
<ol>
    <?php for ($block=2; $block<=6; $block++) { ?>
        <li>
            <input type=‘text’ class=‘id’ name=‘id[]’ />
            <input type=‘text’ class=‘value’ name=‘value[]’ />
        </li>
    <?php } ?>
</ol>
<button type=‘submit’>Submit</button>
Listing 9: Script for merging results from online and offline tasks

```php
<?php

// SETUP
// Some PHP stuff
error_reporting(E_ALL);
ini_set('display_errors', 1);

// Paths
$results_path = './';
$db_path = './judgement_data.sqlite';

// MAIN SCRIPT

// Set up and flush database
$fh = @fopen($db_path, 'w'); // Flushes DB
if ( $fh === false ) {
    die('nError: Could not open file n'.$db_path' for writing.');
}
fclose($fh);
$db = new SQLite3($db_path, SQLITE3_OPEN_READWRITE);
$result = $db->exec('CREATE TABLE participants
(
    p_id INTEGER PRIMARY KEY,
    p_code TEXT,
    p_age INTEGER,
    p_gender INTEGER
);');
```

CREATE TABLE participants

(bcode INTEGER,
age INTEGER,
gender TEXT,
education INTEGER,
wherefrom TEXT,
southnorth INTEGER
);

CREATE TABLE results

(p_id INTEGER,
s_id INTEGER,
scond INTEGER,
s_duration INTEGER,
r_on_response INTEGER,
r_on_rtime INTEGER,
r_off_rating INTEGER
);
$stmt->bindValue(':p_code', $code);
$stmt->bindValue(':p_age', $age);
$stmt->bindValue(':p_gender', $gender);
$stmt->bindValue(':p_education', $education);
$stmt->bindValue(':p_wherefrom', $wherefrom);
$stmt->bindValue(':p_southnorth', $southnorth);
$stmt->execute();
$p_id = $db->lastInsertRowID();

// Add online results
$stmt = $db->prepare('INSERT INTO results 
(p_id , s_id , s_cond , s_duration , r_on_response , r_on_rtime
) VALUES 
(:p_id , :s_id , :s_cond , :s_duration , :r_on_response , :r_on_rtime
);');
for ($i=1; $i<count($online_data); $i++) {
// ignore training data
if (substr($online_data[$i][2], 0, 5) == 'TRAIN') {
    continue;
} elseif ($online_data[$i][0] == 'oa') {
    $cond = 1;
} else {
    $cond = 2;
}
$dur = $online_data[$i][1];
$cid = $online_data[$i][3];
if ($online_data[$i][4] == 'z') {
    $resp = 1;
} elseif ($online_data[$i][4] == 'm') {
    $resp = 2;
} else {
    $resp = null;
}
$rt = $online_data[$i][5];
if ($rt == 'timeout') {
    $rt = null;
}
$stmt->reset();
$stmt->bindValue(':p_id', $p_id);
$stmt->bindValue(':s_id', $id);
$stmt->bindValue(':s_cond', $cond);
$stmt->bindValue(':s_duration', $dur);
$stmt->bindValue(':r_on_response', $resp);
$stmt->bindValue(':r_on_rtime', $rt);
171 $stmt->execute();
172 
173 // Add offline results
174 $stmt = $db->prepare('UPDATE results
175 SET
176   r_off_rating = :r_off_rating
177 WHERE
178   p_id = :p_id
179   AND
180   s_id = :s_id
181   AND
182   s_cond = :s_cond
183 ');
184 for ($i=1;$i<count($offline_data);$i++) {
185   $s_id = $offline_data[$i][0];
186   if ($s_id == '0') {
187     $s_id = '9'; // correct for programming error
188   }
189   $s_cond = $offline_data[$i][1];
190   if ($s_cond == 'oa') {
191     $s_cond = 1;
192   } else {
193     $s_cond = 2;
194   }
195   $r_off_rating = $offline_data[$i][2];
196   if (!is_numeric($r_off_rating)) {
197     $r_off_rating = null;
198   }
199   $stmt->reset();
200   $stmt->bindValue(':p_id', $p_id);
201   $stmt->bindValue(':r_off_rating', $r_off_rating);
202   $stmt->bindValue(':s_id', $s_id);
203   $stmt->bindValue(':s_cond', $s_cond);
204   $stmt->execute();
205 }
206 
207 } // Create data views in the database
208 $result = $db->exec('CREATE VIEW
209 combined
210 AS
211 SELECT
212   participants.p_id,
213   p_age,
214   p_gender,
215   p_education,
216   p_southnorth,
217   s_id,
218   s_cond,
219   s_duration,
220   r_on_response,
221   r_on_rtime,
222   r_off_rating
223 FROM
224   participants,
225   results
226 WHERE
participants.p_id = results.p_id

$result = $db->exec('CREATE VIEW combined_per_sentence AS
    SELECT s_id, s_cond, round(avg(r_on_response), 2) AS avg_on_response,
    round(avg(r_on_rtime), 2) AS avg_on_rtime,
    round(avg(r_on_rtime-s_duration), 2) AS avg_on_score,
    round(avg(r_off_rating), 2) AS avg_off_rating
    FROM results
    GROUP BY s_id, s_cond
;
');

// FUNCTIONS

// Read CSV file into array
/* This function reads the specified CSV file, using the optionally defined
separator (default ',') and using quotations to assign fields (default '"').
The function returns a two-dimensional array containing the rows and fields
present in the CSV file. An empty array is returned if the CSV file is empty.
* @param string $file Path to the CSV file to be read
* @param string $sep Separator for fields, default ','
* @param string $trim Characters to be trimmed from either side of fields
* @return array Returns a two-dimensional array representing rows and columns
*/
function read_csv($file, $sep=' , ', $trim=' " ') {
    $lines = file($file);
    foreach($lines as $key => $line) {
        $line = trim($line);
        $quot = false;
        $line = csv_explode($sep, $line);
        foreach($line as $index => $value) {
            $line[$index] = trim($value, $trim);
        }
        $lines[$key] = $line;
    }
    return $lines;
}

// Explode CSV line into Array
/* This function takes a line from a typical CSV file and separates it into an
array using the given separator, much like explode(). However it ignores any

APPENDIX

289  * occurrences of the separator inside double quotation marks (""").
290  *
291  * @param string $sep The separator to be used
292  * @param string $line The CSV line to be parsed
293  * @return array Returns an array with the individual fields in the CSV line
294  */
295 function csv_explode($sep, $line) {
296     $return = array();
297     $cell_count = 0;
298     $return[0] = "";
299     $quot = false;
300     for($i=0;$i<strlen($line);$i++) {
301         if($quot) {
302             if($line[$i] == ""') {
303                 $quot = false;
304             } else {
305                 // ignore sep until unquoting
306                 $return[$cell_count] .= $line[$i];
307             }
308         } else {
309             if($line[$i] == ""') {
310                 $quot = true;
311             } else {
312                 if($line[$i] == $sep) {
313                     $cell_count++;
314                     $return[$cell_count] = "";
315                 } else {
316                     $return[$cell_count] .= $line[$i];
317                 }
318             }
319         }
320     } return $return;
321 }  
322 ?>
323

D Instructions for Judgement Experiment

Instructions for Online Task

Welcome to the auditory judgement task!

During this experiment, you will hear a short beep followed by a sentence in Welsh. Some of these sentences are perfectly fine colloquial Welsh sentences, as you could possibly hear them somewhere in the street. However, some of the sentences were changed and probably don’t sound right to you.

Your task is to listen carefully to all the sentence and decide as quickly as you can whether you think that what you’ve just heard is an acceptable example of a colloquial Welsh sentence or not. If you think it is okay, you should press the right (M) key - but if you think it doesn’t really feel right to you, press the left (Z) key!

(Page Break)

Don’t worry whether you think the sentences are ‘proper Welsh’ - most of them aren’t, and we don’t really care. What we want to know about is your personal intuition, what you would think if you heard this in real life. So remember that you are the real expert in this experiment!

We will now first give you 10 sentences to practice, as this task takes a little getting used to at first. After this you will have the chance to take a little break (as at several points during the experiment!) before the real thing starts. Should you have any problems you can ask the researcher for help during the break.

To start the practice session press any key...
Instructions for Offline Task

In this task you are presented with a number of colloquial Welsh sentences similar to those you’ve heard in the computer-aided task you have just completed. Again some of these sentences will be just fine and some of them will probably seem rather odd to you.

As opposed to the previous task however, this time we want you to rate how acceptable the given sentences seem to you. For this you will see a five point scale next to every sentence. You should use the number 1 to indicate that you feel the sentence is completely unacceptable and the number 5 to indicate that it feels completely acceptable to you. Use any of the numbers in-between to indicate that you have a tendency to say it is acceptable or unacceptable, or the box in the middle if you cannot decide at all.

Again this is about what you feel is appropriate in the colloquial, spoken language and that this is not about what you may have been taught about Welsh in school. As you will surely know sometimes what people do can be very different from what they teach! So remember that this is about your personal opinion about the language you speak and so you are the real expert!
### E Sample of Offline Judgement Questionnaire

**Block 2**

<table>
<thead>
<tr>
<th>Question</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y ti’n casglu’r plant o’r ysgol.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Pwy wyt ti’n ei warhodd?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti ddim yn cael mynd adra eto.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wyt ti’n dilyn Pobol y Cwm ar S4C?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Dillad ti’n prynu.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti neud dy waith cartref di yn dda ddoe.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti’n cael cawod beno?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ymarfer Karate wyt ti.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti’i di golchi’r llestri.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Siarad efo ffrind wyt ti.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Yr heddlu wyt ti’n osgoi.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wnest ti fwydo’r planhigion mwy nag wythnos yn ôl.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>F’n licio hufen iâ.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti sguobo ’fory.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti’n chwarae tennis yn dda.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti’n ffonio fi neithiwr.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ni’n neidio o gwmpas ar y gwely.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wyt ti’n dilyn Pobol y Cwm ar S4C?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti’n tecstio at dy ffrindiau yn aml.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I’r canolfan hamdden ti’n mynd.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I’r canolfan hamdden wyt ti’n mynd.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Dwyt ti ddim yn cael mynd adra eto.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wyt ti’n bwytia cinio rwan?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Ti’n siarad efo fy athro i wythnos nesaf.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Rwyt ti’n eistedd yn y ’staffell fwy.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Efo pwy ti’n dawnsio?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Mae Rhian yn siarad Almaeneg hefyd.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Pwy ti’n siarad efo?</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Rwyt ti’n tecstio at dy ffrindiau yn aml.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Wnest ti neud dy waith cartref di yn dda ddoe.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
Are you a native Welsh speaker? (30mins, £5!)

I am currently conducting some research on colloquial Welsh. For this I need native Welsh speakers to take part in an experiment in which you will be played some recorded sentences in Welsh and also given a list of sentences which you will then be asked some questions about.

You don't need to be good with grammar or spelling or even think that your Welsh is “good”, all you need is to be over 18 years old and a fluent, native Welsh speaker!

The experiment should take about half an hour or less (depending on how fast you are) and on completion you will also be reimbursed £5 for your time.

If you are interested please get in touch with Florian at elub45@bangor.ac.uk or at 07932 902 250.
G Consent Form for Judgement Experiment

Bangor University’s ‘Code of Practice for the Assurance of Academic Quality and Standards of Research Programmes’ (Code 03)
https://www.bangor.ac.uk/ar/main/regulations/home.html

COLLEGE OF ARTS & HUMANITIES

Participant Consent Form

Researcher’s name: Florian Breit
E-Mail: elub45@bangor.ac.uk
Phone: 07932 902 250

The researcher named above has briefed me to my satisfaction on the research for which I have volunteered. I have been informed that the researcher intends to use the data collected for a dissertation submitted to the School of Linguistics & English Language at Bangor University and in the potential publication of an article in an academic journal. I have also been informed that the researcher intends to make the entire set of data collected during this research, in anonymised and non-identifying form, publicly available. I understand that I have the right to withdraw from the research at any point without any explanation by alerting the researcher of this and that any data collected from me will subsequently be destroyed in this case. I also understand that my rights to anonymity and confidentiality will be respected.

Signature of participant .................................................................

Date ............................................................................................

This form will be produced in duplicate. One copy should be retained by the participant and the other by the researcher.