Swansea University
Prifysgol Abertawe

The Production, Comprehension and Processing of Welsh Grammatical Gender in Welsh-English Adult Bilinguals

Tesni Rhianedd Galvin

Submitted to Swansea University in fulfilment of the requirements for the Degree of Doctor of Philosophy

Swansea University

2024

Abstract

Grammatical gender in Welsh is a persistent area of difficulty in child bilingual acquisition, yet it is well-established in adult control data (Binks & Thomas, 2019; Sharp, 2012). Typically, gender has been investigated through its use in combination with the mutation system. This thesis extends these findings to the use of gender in Welsh-English bilingual adults and establishes whether gender is as robust when in conjunction with mutations or independent of mutations. The first experiment explored the comprehension and production of gender in 40 Welsh adult bilinguals and considered language dominance and Welsh linguistic proficiency. The results from a comprehension task showed fairly accurate use of gender in comprehension, with similar performance when gender is independent of mutations and in conjunction with mutations. The results from a production task revealed that the bilinguals have a good productive command of gender when it is marked via numerals. There is also evidence to suggest stronger performance for the production of nouns in numeral-noun contexts, when gender is not involved with mutations, than when gender is in conjunction with mutations. Welsh linguistic proficiency predicted better performance in both comprehension and production tasks, while language dominance did not. The follow-up experiment investigated the processing of gender in 21/40 Welsh bilinguals and considered cognitive and environmental individual differences. The results showed evidence of sensitivity to determiner-noun gender agreement errors when gender was encoded through mutations via the determiner ‘y’, but not when gender was in conjunction with mutations nor independent of mutations. There was no evidence to suggest that any of the individual difference variables predicted increased sensitivity to gender agreement errors in Welsh, despite some variables predicting overall faster reading times in the bilinguals. These findings demonstrate that Welsh adult bilinguals from a more diverse population than those tested previously made use of gender in comprehension and production, and processed gender agreement errors during real time.

Declarations and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed [redacted] (candidate)

Date: 20/08/2024

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed [redacted] (candidate)

Date: 20/08/2024

STATEMENT 2

I hereby give my consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available to outside organisations.

Signed [redacted] (candidate)

Date: 20/08/2024

Table of Contents

Acknowledgements

First and foremost, I would like to thank the *Economic and Social Research Council Wales Doctoral Training Partnership* for funding this project (ES/P00069X/1).

My greatest thanks go to my primary supervisor, Dr Vivienne Rogers, for her continual support and guidance throughout this project. I am extremely grateful not only for her invaluable academic insight, but also for her never-ending patience, positivity, and encouragement. I would like to thank her for all of the time and attention she has given me over the years in helping me write this thesis. I would like to express my gratitude to my previous second supervisors, Dr Rocio Perez-Tattam, Dr Alex Lovell and Mr Steve Morris, for their time, guidance, feedback and suggestions.

I must also acknowledge the help I received from Dr Leigh Fernandez, Dr Pablo Bernabeu, Dr Fernando Martín-Villena and Dr Emily Lowthian for their assistance with my statistical queries. I am also grateful to Dr Alexia Bowler, Dr Cornelia Tschichold, Dr Xuehong He and Professor Tess Fitzpatrick in the Department of Applied Linguistics at Swansea University, who provided me with advice and support throughout my time at Swansea. I would like to thank the examiners, Professor Vicky Chondrogianni and Dr Kyle Jones, for the constructive feedback they have given me to improve this thesis. My thanks also go to the Welsh speakers who took part in my two experiments. Without their willingness to participate, none of this would have been possible.

I offer special thanks to my family. My mother, Siân, and my father, Tony, who have always believed in my potential. They have given me endless love and support during my PhD and throughout my life. I have been able to succeed because I have had such parents as you beside me; thank you. I sincerely thank my siblings too for your help in times of need. I would also like to express my gratitude to my friends Rachael, Amelia, and Harriet, with whom I have shared this journey, and to the life-long friendships I have made over the years, including Sarah, Fernando, Lewis, Pernelle and Anna-Marie.

My deepest gratitude goes to my husband, Carwyn. It would not have been possible for me to write this thesis without your support, love, patience, and care. Thank you for being a comfort when I needed it most and for being the shoulder I can always lean on.

Diolch o galon i chi gyd.

List of Tables

List of Figures

List of Examples

Abbreviations

SM	Soft mutation
AM	Aspirate mutation
NM	Nasal mutation
DP	Determiner phrase
NP	Noun phrase
L1	First language
L2	Second language
N	Noun
Adj	Adjective
V	Verb
SLA	Second language acquisition
AoA/O	Age of acquisition / onset
IH	Interpretability Hypothesis
MSIH	Missing Surface Inflection Hypothesis
SPR	Self-paced reading
DP	Declarative / Procedural (model)
WM	Working memory
WMC	Working memory capacity
CE	Central executive
LTM	Long-term memory
STM	Short-term memory
PWM	Phonological working memory
RST	Reading Span (task)
OSPAN	Operation Span (task)
CVMT	Continuous Visual Memory Test
TOH	Tower of Hanoi (task)
TOL	Tower of London (task)
RTs	Reaction time(s)
(G)LMM	(Generalised) Linear mixed effects model

Chapter 1
Introduction

Wales is one of the four countries in the United Kingdom and Welsh is one of two official languages in Wales, the other being English. The Welsh language (Cymraeg), along with Breton and Cornish, is a member of the Brythonic subgroup of the Celtic languages. It is thought to go back at least 2,500 years and is acknowledged as ‘the senior language of the men of Britain’ by J. R. R. Tolkien (Davies, 2014). Welsh has evolved from Early Welsh (from its beginnings to *c*.850), to Old Welsh (*c*.850 to *c*.1100), to Middle Welsh (*c*.1100 to *c*.1550), to Early Modern Welsh (*c*.1550 to *c*.1600) and now to Late Modern Welsh (*c*.1600 to present day) (Borsley, Tallerman, & Willis, 2007; Davies, 2014). During the last century, the Welsh language declined from around one in every two people in Wales speaking Welsh, to one in five (Davies, 2014)[1]. According to the censuses, the proportion of Welsh speakers had fallen from 80% in 1800 to 54.4% in 1891 and to 18.91% in 1981 (Borsley et al., 2007). Today, the official figure of Welsh speakers is documented by the census that is carried out every 10 years in Wales. The latest figures from the 2021 census shows that there are approximately 473,060 Welsh speakers born in Wales, aged 3 years and over. This is roughly 22.3% of the population c.3.1 million and such figures mark Welsh as a minority language in Wales. On the census, individuals are required to answer whether they can understand, speak, read and/or write in Welsh, by ticking a box, or by responding with *none of the above*. It is also worth noting that the last recorded monolingual speakers of Welsh were in 1981 and since this census, no adult monolinguals have been officially documented.

Of the proportion who are able to speak Welsh, this varies significantly by local authority. For instance, 50.4% of Carmarthenshire’s population recorded that they could speak Welsh (west Wales), 41.0% of Conwy’s population recorded that

[1] The decline can be attributed to various factors: (1) the industrial revolution, i.e., the immigration of English and Irish workers to industrial south Wales, forming a language shift to English (Borsley et al., 2007), (2) the Welsh Not, where children were discouraged from speaking Welsh at school, by marking out those who were heard speaking the language, (3) the World Wars perhaps accelerated the decline of the language also.

they could speak Welsh (north Wales) and 33.8% of Denbighshire's population recorded that they could speak Welsh (north-west Wales). By contrast, only 12.2% of Cardiff's population recorded that they could speak Welsh (Wales' capital city) and 11.2% of Swansea's population recorded that they could speak Welsh (Wales' second largest city in south Wales). Welsh is also spoken outside of Wales. It is estimated that there are 65,240 Welsh speakers (12.1%) aged 3 years and over, born outside of Wales, which is a slight increase from 2011 (11.8%). Welsh is also spoken in the province of Chubut, Patagonia. This Welsh speaking community was created by Welsh settlers more than 150 years ago in 1865. The latest recorded figure of Welsh speakers in Chubut is approximately 6,000 (British Council, 2023).

The Welsh language has been under a period of revitalisation since the second half of the 20th century. Considerable progress has been achieved through Government support for the Welsh language over the last 30 years. Such increasing efforts to revitalise the language are often the result of governmental initiatives and legislation that build upon one another to facilitate and support the use and availability of the language, increasing its accessibility and speaker rights. Since 1993, Welsh has been given equal status to English within the public sector, although services within the sector were slow to honour this standing for the language. Since then, the Welsh language measure was created in 2011, providing the Welsh language with official status and to be treated no less favourably than English in Wales. More recently, the Welsh Government unveiled new Welsh language policies in 2017 (Welsh Government, 2017) which sought to:

- Offer a service of equal standing to Welsh and English speakers.
- Understand and acknowledge the bilingual nature of Wales.
- Provide a wider access to activities.
- Raise the Welsh profile and recognize ownership by Welsh speakers and non-Welsh speakers alike.
- Ensure conformity to the linguistic and cultural specifications in legislation and best practice guidance as approved by the Welsh Language Commissioner.

The Welsh Government is currently consulting on proposals that will form the basis of a Welsh Language Education Bill, for pupils in Wales to become 'confident' Welsh speakers. These have all been introduced in the efforts to aid the

Welsh Government's strategy *Cymraeg 2050* in reaching a million Welsh-language speakers by 2050, which marked the fiftieth anniversary of the first Welsh Language Act (1967). The strategy hopes for Welsh to be used in every aspect of life, creating favourable circumstances throughout the country that support the acquisition and use of the language. Efforts have been made to enhance the usability and visibility across the workplaces, education, and social use of Welsh.

One of the most known organised events to celebrate the culture and language of Wales is *Yr Eisteddfod.* The first documentation of such a festival took place in 1176, however, with the anglicization of Welsh in Wales, it no longer took place until its revival in 1865. The event is held in a different part of Wales every year and involves hundreds of stalls, exhibitions, literature events, performances, and competitions. It is one of Europe's largest music and poetry festivals, with approximately 150,000 visitors each year. There are local newspapers, including *Clebran* and *Cardi Bach*, magazines such as *Golwg* and *Barn*, as well as a weekly national paper online, *Y Cymro*, all in the Welsh language. There is also a radio channel in Welsh, *BBC Radio Cymru*, as well as the television channel *S4C*. There exists a popular Welsh soap opera 'Pobl y Cwm' in addition to four television programmes which have been produced and filmed bilingually in Welsh and English, including '*Craith* / Hidden', '*Y Gwyll* / Hinterland', '*Un Bore Mercher* / Keeping Faith' and '*Dal y Mellt* / Catch the lightning'. There exists a Welsh version of the popular video game 'Minecraft' as well as a Welsh language video game inspired by Welsh folklore called the 'Maid of Sker'. There are popular music bands which sing in Welsh, such as *Alffa* and *Bwncath*, who have performed as supporting acts at the Principality Stadium and at international rugby games, as well as the up-and-coming bilingual drill rapper, *Sage Todz*. The popular novel 'The Hobbit' has been translated into Welsh and there are several Welsh fiction and non-fiction novels, including 'Coed y Brenin' and 'Llyfr Glas Nebo', with the author of the latter receiving the *Yoto Carnegie Medal for Writing*. Welsh poetry has also played a key role in Welsh literature (Baker, 2003), with well-known poets such as *Hedd Wyn*, who had written his chair-winning poem 'Yr Arwr' (The Hero) before he died in the First World War.

Promoting Welsh in Wales is seen in bilingual road and traffic signs, which was permitted by special authorisation in 1965. More recently, regulations came into force in 2016 mandating that all signs in Wales be in Welsh first. Welsh is also seen in place names, where most derive from the Welsh language. There are some Welsh place names which only have one Welsh name and do not have bilingual English names, e.g., *Llanelli*, while some Welsh place names have translated English names, some of which are almost identical both in spelling and pronunciation from the Welsh, e.g., *Caerffili* / Caerphilly, and other place names do not have similar forms, e.g., *Abertawe* / Swansea. There are other place names in Wales where the English name was adopted by the Welsh, e.g., *Wrecsam* / Wrexham, and these are typically close to the English border. There are also Welsh place names in England, such as *Pencraig* (near the border) and *Penrith* (in Cumbria).

To promote Welsh in communities, *Cymdeithas Yr Iaith* ('Welsh Language Society') was formed in 1962 which campaigns for Welsh to be used in every aspect of life. The first *Menter Iaith* ('Language Initiative') was founded in 1991 in Carmarthenshire, which hosts a wide range of activities to raise awareness and the profile of Welsh. Today, there are 22 *Menter Iaith* across Wales. The *Urdd Gobaith Cymru* ('The League of Welsh Hope') is a national voluntary youth organisation providing opportunities to children and young people in Wales to take part in experiences and activities in Welsh. It was created in 1992 and has over 56,000 members today. In the majority of universities in Wales, there are Welsh language societies known as *Y Gymdeithas Gymraeg*, which facilitate socialising in Welsh.

Welsh is also taught in the education system. There are over 500 *Ti a Fi* ('You and Me') parent and baby groups across Wales run by *Mudiad Meithrin* ('Nursery Movement'), which are available for babies and toddlers up to two years of age to play, listen to stories and sing Welsh songs together. In 1988, the Education Act made Welsh a compulsory subject in secondary schools up to age 16, after which it becomes an elective subject. For primary education (aged 4-11 years), there are four different categories of schools according to the amount of Welsh used in teaching, learning and the day-to-day conduct of the school (Welsh Government, 2013). These are:

1) Welsh-medium - main teaching medium is at least 70% in Welsh, and English is introduced formally as a subject at Key Stage 2 (7-11 years). Of the 1,219 primary schools in Wales, 450 are Welsh-medium and are spread across every county.
2) Dual stream - two types of provision exist side by side, where children can either opt for mainly Welsh-medium education or mainly English-medium education.

3) English medium with significant use of Welsh, where learning takes place in both languages with greater emphasis on English. Welsh is used as the medium of learning for 20% to 50% of the curriculum.
4) English medium with Welsh taught as a second language, where less than 20% of the teaching is through Welsh.

Secondary education in Wales is categorised in a similar way (aged 11-18 years). Welsh medium schools teach all subjects apart from English in Welsh, however, some of these schools offer mathematics and the sciences through English as well as Welsh (i.e., separate classes available). English medium schools teach subjects predominantly in English and Welsh is taught as an L2, however, between 20- 49% of subjects can be taught through Welsh, with both languages used for communication and administration. Bilingual schools range in their provision and are categorised in the following way (Welsh Government, 2016):

- Category A: at least 80% of subjects, apart from Welsh and English, are taught through Welsh. One or two subjects are taught to some pupils through English or in both languages.
- Category B: at least 80% of subjects, excluding Welsh and English, are taught through Welsh, but are also taught through English.
- Category C: 50-79% of subjects, excluding Welsh and English, are taught through Welsh, but are also taught through English.
- Category Ch: all subjects, except Welsh and English, are taught to all pupils using both languages.

As part of *Cymraeg 2050*, the Welsh Government seeks to increase the proportion of each school year group receiving Welsh-medium education from 22% in 2021 to 30% by 2031, and then 40% by 2050. Taken together, each of these initiatives have played a key role in the efforts to revitalise Welsh over the years and to reaching the target of a million Welsh speakers by 2050. In a recent interview (S4C, 2023), David Crystal[2] revealed that Welsh is in the top 7% of languages spoken globally, as far as the number of speakers is concerned and stated that Welsh,

[2] David Crystal, OBE, FBA, FLSW, FCIL, is a British linguist and an Honorary Professor of Linguistics at Bangor University.

as a minority language, has done “wonderfully well” and is “one of the success stories of the 20th stories in terms of the revitalisation process”. There is, however, still much to do to reach the 1 million speak target.

Welsh is a minority language in Wales, with English as the co-existing majority language. In a minority of areas in Wales, Welsh is the dominant language in the community where Welsh speakers are in the majority (over 50% of their respective populations speak Welsh) and where Welsh will be a child’s stronger, first language. However, English is the dominant language in the majority of areas and communities across Wales. The co-existence of Welsh and English has led to a fear that Welsh is undergoing a linguistic convergence as a result of close contact with English (Jones, 1998; Phillips, 2008). There is evidence to suggest that in colloquial speech and sometimes in the written language, there is considerable variation in adherence to the prescriptive rules in the standard Welsh language in adult speakers, particularly in regard to the use of mutations[3] (Thomas, 2001). This variation exists between and within speakers (Thomas, 2001). It has been suggested that there is a simplification of the mutation system underway, with it gradually being reduced from the original four-way alternation of the standard language (soft mutation, nasal mutation, aspirate mutation and basic form, i.e., no mutation), to a binary alternation between the basic form and soft mutation. Mutations, however, play a key part in how grammatical gender is marked in Welsh[4]. There is also evidence indicating that gender marking in Welsh is showing signs of simplification as well (Jones, 1998; Thomas, 2001). However, previous studies have typically investigated its use in combination with the mutation system (e.g., Gathercole, Thomas, & Laporte, 2001; Thomas, 2001).

For this reason, this thesis extends these findings to disambiguate the production, comprehension, and processing of Welsh gender, by establishing whether gender is as robust when in conjunction with or independent of mutations. The speakers of interest are Welsh-English bilingual adults who are from different areas across Wales, but predominantly South Wales, providing a more diverse

[3] The term ‘mutation’ refers to a set of morphophonological alternations that affect certain initial consonants of words. These are discussed in more detail in Chapter 2.

[4] Grammatical gender is a class system whereby all nouns are arbitrarily assigned into two or more different classes. Welsh has grammatical gender, like French, but English does not. Gender in Welsh is binary (masculine / feminine). The gender and mutation systems in Welsh are described in Chapter 2.

population of the Welsh adult speakers here in Wales than those previously tested (Welsh dominant areas in North Wales). Specifically, the thesis will investigate the production, comprehension and processing of grammatical gender in Welsh-English bilingual adult speakers and will use various experimental tasks to measure this domain. This thesis will also consider the possible role of individual difference variables in the use of Welsh grammatical gender.

This thesis is organised as follows. Chapter 2 introduces the linguistic domain of interest, grammatical gender, and its involvement with the mutation system. It also describes some of the linguistic properties of grammatical gender that make gender easier to acquire in some languages than others, as well as some of the language-external factors that play a role in the acquisition of gender. This is followed with a discussion of the acquisition of Welsh gender. Chapter 3 reviews the prior research on production and comprehension of grammatical gender in Welsh adults, before reviewing two theoretical approaches of ultimate attainment in L2 acquisition. After reviewing each theory, I will present the predictions for the Welsh adult bilinguals in this first experiment. Chapter 4 presents the first experiment of two which form the heart of this thesis, investigating the comprehension and production of grammatical gender in Welsh-English adult bilinguals, as well as the role of individual difference variables. Chapter 5 extends these findings from production and comprehension to processing, and reviews various theoretical approaches to the processing of grammatical gender in L2/bilingual speakers. Chapter 6 presents experiment two, exploring the processing of gender in a sub-group of the participants from experiment one and also considers the role of individual difference variables. Chapter 7 brings together the findings from the two experiments, chapters 4 and 6, and presents a general discussion. It also summarises the main conclusions that can be drawn from this thesis, together with the discussion of some of its limitations and some possible considerations for future research.

Chapter 2

The Linguistics of Grammatical Gender

The previous chapter introduced the overarching theme of this thesis, namely, how Welsh-English adult bilinguals comprehend, produce and process grammatical gender in Welsh. This second chapter introduces the linguistics of gender. First, a definition of grammatical gender is provided, describing its agreement and assignment systems. Second, the gender system in English is briefly outlined, then, an introduction to the syntax of Welsh is given, outlining in detail the Welsh gender system and the role of mutations in the gender system. The final sections briefly review the different properties of grammatical gender that make gender easier to acquire in some languages than others, followed by a discussion of the acquisition of Welsh gender.

2.1 What is gender?

Gender is the categorisation of nouns and it is a central feature in some (e.g., French, Spanish, Welsh) but not all languages (e.g., Persian) (Corbett, 1991). There are two types of gender – grammatical gender and semantic gender. Gender is an inherent, arbitrary lexical property of the noun and speakers need to classify a noun as belonging to a particular gender class (Chondrogianni, 2024; Corbett, 1991). Depending on the language, there are two or more such classes or genders (Corbett, 1991). Gender is realised differently across languages. For instance, some languages, such as French, Spanish and Welsh, have a two-way gender system (masculine and feminine), though in other languages, such as German and Russian, these have a three-way gender system (masculine, feminine and neuter) (Corbett, 2014). Not all binary gender systems classify by a masculine/feminine distinction, some systems classify nouns by a common/neuter distinction (e.g., Danish, Norwegian) or animate/inanimate (e.g., Basque and many Native American languages such as Dakotan or Hopi) (Corbett, 1991).

For some languages, such as those mentioned above, the main principles for assignment are grammatical, while for others, they are semantic in the sense of involving features of a noun's referents and in many cases, semantic and grammatical features play a role in assignment (Corbett, 2014). In languages that show semantic gender, such as English, gender distinctions are reflected in the semantics of the nouns, the things in the world that it is used to represent. Distinctions can largely be made for those referents that can be distinguished by sex. For example, in English, ‘father’ is masculine because it refers to a male human and ‘mother’ is feminine because it refers to a female human (Corbett, 1991).

Semantic gender exists in all languages, whereas grammatical gender only exists in some languages (Corbett, 2014). Notably, grammatical gender and semantic gender are not always the same[5]. For many languages, both semantic and formal features play a role in assignment (Corbett, 2014). Grammatical gender can be a complex system to acquire - in languages that possess grammatical gender, every noun has its own gender. Gender can be marked or unmarked on the noun itself. For instance, gender is straightforwardly marked on the noun in Russian, where noun endings indicate noun gender, whereas, in languages such as French, gender is marked morphologically only in the form of the words associated with the noun, such as the determiner or an adjective. In French, there also are certain stems (suffixes and/or affixes) that indicate the noun’s gender, but often, the determiner is an obligatory gender marked prefix on the noun. In contrast, some languages lack any kind of gender distinction altogether, such as Finnish, where the same grammatical forms are used when talking about male and female beings (Corbett, 1991).

It has been proposed that a fundamental feature of grammatical gender systems is the presence of agreement (Corbett, 2014). Therefore, a language that has a binary gender system has two classes of nouns that are distinguished syntactically by the agreement relationships they enter into (Corbett, 1991). Agreement is responsible for the syntactic cohesion between words within a phrase and features such as grammatical features, such as number, case, or gender are mandatory to establish grammatical agreement in the noun phrase (NP) (Corbett, 2014). Gender agreement systems follow rules which are language specific and often, these can be easily learned (Corbett, 1991).

[5] For instance, in German, ‘das Mädchen’ is *neuter* but means *girl.*

2.1.1 Gender agreement

The gender agreement system operates in languages that have formal gender systems (e.g., Russian, French, German, Spanish, Welsh, etc) (Corbett, 1991). The term is used to describe the relationship between two elements, whereby the form of one element determines the form of another element. The noun in the phrase forms an agreement between words in the same construction, such as adjectives, articles, and numbers. These typically agree in gender, number, and person, but not all gendered languages mark all of these things. In languages such as French and Spanish, nouns assign gender to the words that modify them (e.g., determiners & adjectives) and the pronouns which co-refer with them. In most gendered languages, the noun is almost always associated with a determiner, and often adjectives when present. Below are some examples in French and Spanish.

Example 2.1	*La*	*grande*	*maison*	*(French)*
	The $_{\text{DET (f)}}$	big $_{\text{ADJ (f)}}$	house $_{\text{NOUN (f)}}$	
	'The big house'			
Example 2.2	*Le*	*grand*	*bâtiment*	*(French)*
	The $_{\text{DET (m)}}$	big $_{\text{ADJ (m)}}$	building $_{\text{NOUN (m)}}$	
	'The big building'			
Example 2.3	*La*	*casa*	*grande*	*(Spanish)*
	The $_{\text{DET (f)}}$	house $_{\text{NOUN (f)}}$	big $_{\text{ADJ (f)}}$	
	'The big house'			
Example 2.4	*El*	*gran*	*edificio*	*(Spanish)*
	The $_{\text{DET (f)}}$	big $_{\text{ADJ (m)}}$	building $_{\text{NOUN (m)}}$	
	'The big building'			

Gender is an inherent feature of nouns and from a generative grammar perspective, there is a general assumption that gender is a *phi*-feature specification, and these come in two varieties: interpretable and uninterpretable features (Carstens, 2010). Interpretable features are features which are found on the noun, thus effecting semantic interpretation (i.e., contribute to meaning), whereas, uninterpretable features do not contribute to meaning and are only relevant for syntactic operations (e.g., determiner) (Carstens, 2010; Chondrogianni, 2024). Interpretable features on nouns are checked against the uninterpretable features on other syntactically agreeing elements within the noun phrase, through the syntactic operation of Agree (Carstens, 2005). Agree involves two elements: a probe (uninterpretable features) and a goal (interpretable features), where the probe seeks matching features (goal). It copies the gender *phi*-feature of one constituent onto a dependent constituent. For instance, the Spanish NP “La casa grande” (example 2.3), the definite article “la” is the probe and precedes the singular feminine noun “casa” which is the goal. The features of the definite article match the features of the noun to agree in gender, showing feminine gender agreement. Similarly, the adjective “grande” agrees in gender with the noun “casa”. Gender agreement entails a syntactic component (gender agreement), as well as a lexical component (gender assignment) (Chondrogianni, 2024). Gender assignment is introduced below.

2.1.2 Gender assignment

Gender assignment is the term used to refer to an inherent lexical feature on nouns and is considered a complex feature that varies across languages (Corbett, 1991, 2014). There are two major principles involved in the assignment of gender. The first is semantic principles, which means that nouns are assigned to gender according to their meaning (semantic gender). While the second, grammatical gender, assigns gender according to their form and this overwrites semantic gender. Corbett (1991) argued that although gender systems of certain languages, such as French and Spanish, are often outlined as if they were arbitrary, the form of the noun plays a major role in the assignment of gender. For instance, many noun endings in Spanish are highly transparent, 96.3% of nouns ending with ‘a’ are feminine and 99.9% of nouns ending with ‘o’ are masculine (Teschner & Russell, 1984), thus providing reliable cues to recognise the gender of a noun. Although, there are some exceptions to these generalisations, e.g., masculine *cometa* "comet" and feminine *foto* "photo". While in French, noun-endings do co-occur with gender and these are important indicators of grammatical gender. Tucker, Lambert and Rigault (1997) claimed that 60% of French nouns have high predictive endings, with 96% of nouns ending with ‘ette’ are feminine, and 99% of nouns ending with ‘ain’ are masculine. Gender assignment rules are not absolute, they are probabilistic, in the sense that they are reliant to different degrees and for all rules, there are exceptions (Ayoun, 2022). This makes learning gender assignment in some languages more difficult than others.

Studies have found that noun endings are used to assign gender to nouns (e.g., Karmiloff-Smith, 1979; Macwhinney, 1978; Pérez-Pereira, 1991). Corbett (1991) claimed that grammatical gender assignment can be learned from both the form of the noun and the semantics of the noun. The regular noun endings can be used to predict noun gender, which then accelerates lexical access and processing. The use of morphophonological features has been one of the ways in which researchers have explored how L1 speakers acquire grammatical gender, and these results have been compared to second language acquisition of gender (e.g., Karmiloff-Smith, 1979; Macwhinney, 1978; Pérez-Pereira, 1991). There are different linguistic properties which make some gendered languages easier to acquire than others, such as transparency / opacity of features, the presence / absence of morphophonological cues and the locality of gender effects / structural distance. These properties are discussed in more depth in section 2.5 of this chapter, but first, the two following sections outline gender in English and Welsh because the current dissertation investigates how grammatical gender is produced, comprehended, and processed in Welsh-English adult bilinguals.

2.2 Gender in English

This section briefly outlines the syntax of English and its gender system. English does not have a formal grammatical gender system, i.e., a system with an arbitrary relationship between the referent and the label. English has a semantic gender system which is visible on some nouns (based on semantic information) and pronouns. It has three genders; masculine, feminine and neuter, in the sense that the noun ‘boy’ (he) is masculine because it refers to a male, while ‘girl’ (she) is feminine because it refers to a female. Animate non-humans, inanimate nouns and groups of humans are usually treated as neuter (i.e., it) and the pronoun ‘they’ is also used for people identifying as non-binary. There are very few exceptions to this. Sometimes gendered pronouns are applied to inanimate nouns in English, for example, ‘ship’ can be referred to as ‘she’. Typically, the personification of sexless objects is done for poetic effect and is referred to as metaphorical gender (Thomas, 2001).

English is a pronominal gendered language because gender is marked almost exclusively on pronouns. Gender is marked on reflexive (himself/herself) and possessive (his/her) pronouns for animates, as well as on 3rd person singular pronouns (he/she/it). English has three grammatical cases where the three-gender system is seen (masculine, feminine, neuter). The case system for third person singular pronouns is presented in the table below.

Table 2.1

Case System for 3rd Person Singular Pronouns

Case	Masculine	Feminine	Neuter
Nominative	He	She	It
Objective	Him	Her	It
Genitive	His	Her	It

In English, the pronoun must agree with the antecedent and it must agree in number, i.e., a singular antecedent requires a singular pronoun, where a plural antecedent requires a plural pronoun. Two examples of this are given below (Thomas, 2001).

Example 2.5 *'The boy washed his car'*

Example 2.6 *'The boys washed their cars'*

While in French, the possessive adjective agrees with the noun, rather than the antecedent, as it would in English.

Example 2.7	*Il*	*aime*	*sa*	*mère*
	He $_{\text{PRONOUN (m)}}$	loves $_{\text{VERB}}$	Her $_{\text{PRONOUN (f)}}$	Mother $_{\text{NOUN (f)}}$

'He loves his mother'

Rather than: 'Il aime son mère'

Pronouns are not the only markers of gender in English. There are some feminine noun suffixes for humans that can be added, such as '*ess*' to mark feminine gender. This is typically seen in job titles, for instance:

Example 2.8 *Steward (masc) / Stewardess (fem)*

Example 2.9 *Actor (masc) / Actress (fem)*

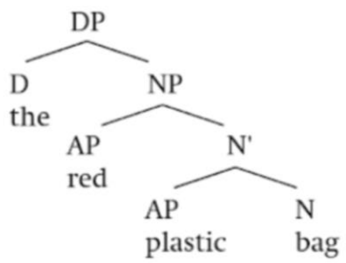
Example 2.10 *Waiter (masc) / Waitress (fem)*

While there is an agreement between the pronoun and co-referential noun in the phrase, there is no determiner and adjective agreement system in English. English follows subject-verb-object (SVO) word order and adjectives appear pre-nominally. The typical determiner phrase (DP) order is *Determiner – Numeral – Adjective – Noun – Complement.* Example 2.11 shows the DP structure (Koeneman & Zeijlstra, 2017, p. 104, ex. 65a) and Figure 2.1 displays the full tree for the English DP (Koeneman & Zeijlstra, 2017, p. 104, ex. 65b).

Example 2.11	*The*	*red*	*plastic*	*bag*
	The DET	red ADJ	plastic ADJ	bag NOUN
	'The red plastic bag'			

Figure 2.1

English Determiner Phrase (DP)



This section briefly outlined the gender system in English, noting how English has a semantic gender system and gender is seen via pronouns, but no formal grammatical gender system. The following section outlines the Welsh grammatical gender system, highlighting the differences between the two languages.

2.3 Gender in Welsh

This section briefly introduces the syntax of Welsh, followed by a detailed outline of the Welsh grammatical gender system. In contrast to English, Welsh is a verb-initial language that follows a verb-subject-object (VSO) word order. Example 2.12 shows the word order in Welsh and the word order in English (SVO).

Example 2.12	*Mae*	*Ffion*	*yn*	*hoffi*	*teisen*
	Be 3SG	Ffion PROPER NOUN	PARTICLE	like INF	cake NOUN (f)
	(V)	(S)			(O)
	'Ffion likes cake'				
	(S) (V) (O)				

The DP structure in Welsh differs to the DP structure in English. While adjectives appear pre-nominally in English, adjectives usually appear post-nominally in Welsh. Example 2.13 and Example 2.14 show the noun-adjective order in Welsh.

Example 2.13	*Y*	*gath*	*ddu*
	The $_{DET}$	cat+SM $_{NOUN (f)}$	black+SM $_{ADJ}$
	'The black cat'		
Example 2.14	*Y*	*mynydd*	*uchel*
	The $_{DET}$	mountatin $_{NOUN (m)}$	high $_{ADJ}$
	'The high mountain'		

However, there are a limited number of adjectives that can appear pre-nominally, including the adjectives '*hen'* (old), '*prif'* (main/head/chief), '*hoff'* (favourite) and '*annwyl'* (dear). Example 2.15, Example 2.16 and Example 2.17 show pre-nominal adjectives in the DP in Welsh.

Example 2.15	*Yr*	*hen*	*ddyn*
	The $_{DET}$	old $_{ADJ}$	man+SM $_{NOUN (m)}$
	'The old man'		
Example 2.16	*Y*	*prif*	*dŷ*
	The $_{DET}$	main $_{ADJ}$	house+SM $_{NOUN (m)}$
	'The main house'		
Example 2.17	*Yr*	*hoff*	*blentyn*
	The $_{DET}$	favourite $_{ADJ}$	child+SM $_{NOUN (m)}$
	'The favourite child'		

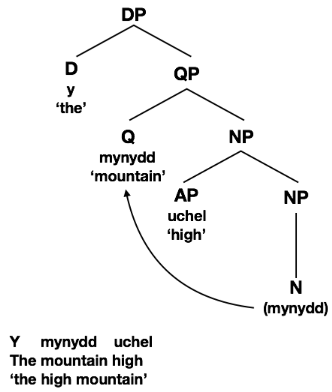
There are also certain adjectives that can appear both before and after the noun, but the meaning changes depending on where it appears. For instance, the adjective 'unig' can mean *only* when it appears pre-nominally, and *lonely* when it appears post-nominally. This is shown in Example 2.18 and Example 2.19.

Example 2.18	*Yr*	*unig*	*blentyn*
	The $_{DET}$	only $_{ADJ}$	child+SM $_{NOUN (m)}$
	'The only child'		
Example 2.19	*Y*	*plentyn*	*unig*
	The $_{DET}$	child $_{NOUN (m)}$	lonely $_{ADJ}$
	'The lonely child'		

Borsley et al (2007) argued that the DP structure in Welsh is head initial and that the noun-adjective word order in Welsh is a result of noun raising, where there is a leftward movement of the noun. The N (noun) raises to Q (quantifier phrase), which is located in a functional position between D (determiner) and N. The N bypasses the adjectives adjoined to it and raises overtly higher than the adjective. Yet, the noun does not raise as far as the determiner. It is claimed that the noun phrase is contained within the functional phrase, and this is within a DP (Borsley et al., 2007). The structure of a simple noun phrase with noun-adjective order is shown below (Borsley et al., 2007, p. 187, example 132).

Figure 2.2

Noun Phrase with Noun-Adjective Order



The noun does not raise as far as the determiner, because there are a limited number of elements that can precede the noun and intervene between D and N in the phrase. There are a limited number of adjectives (noted previously), various quantifiers and low numerals, which appear pre-nominally between the D and N. All other elements follow the head noun. Examples of DPs where quantifiers and numerals appear pre-nominally are shown in the examples below.

Example 2.20	*Yr*	*holl*	*ystafell*
	The $_{\text{DET}}$	whole $_{\text{ADJ}}$	room $_{\text{NOUN (f)}}$
	'The whole room'		
Example 2.21	*Yr*	*holl*	*bobl*
	The $_{\text{DET}}$	whole $_{\text{ADJ}}$	people+SM $_{\text{NOUN (f)}}$
	'All of the people'		

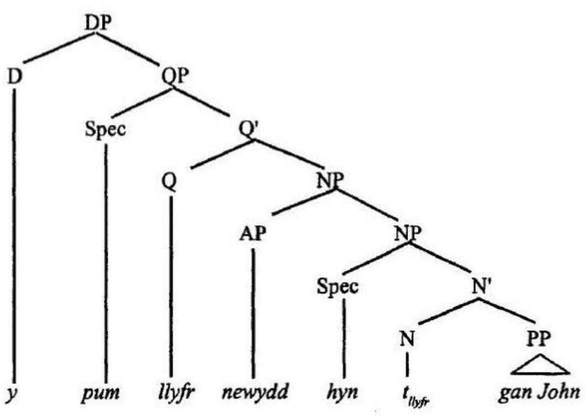
Example 2.22	*Y*	*tair*	*cadair*
	The DET	three NUMERAL (m)	chair NOUN (f)
	‘The three chairs’		
Example 2.23	*Pa*	*ddwy*	*eglwys?*
	Which QUESTION	two+SM NUMERAL (f)	church NOUN (f)
	‘Which two churches?’		

As shown above, certain numerals precede the head noun. The usual order of a noun phrase is *Determiner – Numeral – Noun – Adjective – Demonstrative / Possessor – Complement – Relative* (Borsley et al., 2007). The full DP structure is shown in the example below (Roberts, 2005, p. 92, example 13) and shows the full tree for the Welsh DP (Roberts, 2005, p. 92, example 14).

Example 2.24	*Y*	*pum*	*llyfr*	*newydd*	*hyn*	*gan*	*John*
	The DET	five NUMERAL	book NOUN (M)	new ADJ	this PRON-PL	by PREP	John
	‘These five new books by John’						

Figure 2.3

Welsh Determiner Phrase (DP)



The difference between the DP structures in Welsh and English is due to a strong uninterpretable D feature in Welsh (Borsley et al., 2007). There is also no noun raising in English. Not only do Welsh and English differ in terms of word order and DP structure, the two languages also differ with regards to gender.

In terms of semantic gender in Welsh, animate nouns are usually (but not always) marked by their biological gender (Binks, 2017). For example, the nouns ‘buwch’ (cow [f]) and ‘tarw’ (bull [m]), as well as ‘mab’ (son [m]) and ‘merch’ (daughter [f]), all bear the expected gender. However, this is not the case for all animate nouns, as the Welsh nouns ‘arth’ (bear) and ‘cath’ (cat) are feminine.

As noted previously, English does not have a formal grammatical gender system, but Welsh does. Welsh operates a binary gender system in which nouns, animate and inanimate, are marked for either masculine or feminine gender, while gender oppositions are normally neutralized in the plural. Approximately 69% of nouns in Welsh are masculine, 27% are feminine and roughly 4% are indeterminate (Hammond, 2016). In Welsh, the noun form provides little information about noun gender. However, in standard Welsh, there are small sets of adjectives that have marked feminine forms that do agree with the gender of the co-occurring noun. The examples below illustrate this (taken from Mittendorf & Sadler, 2006, p.5; and Thomas, 2001, p.74).

Example 2.25	*trwm (masculine)*	*heavy*	*trom (feminine)*
Example 2.26	*brith (masculine)*	*speckled*	*braith (feminine)*
Example 2.27	*melyn (masculine)*	*yellow*	*melen (feminine)*
Example 2.28	*gwyn (masculine)*	*white*	*gwen (feminine)*
Example 2.29	*byr (masculine)*	*short*	*ber (feminine)*
Example 2.30	*cryf (masculine)*	*strong*	*cref (feminine)*

In colloquial spoken Welsh, the adjectives in the predicative position rarely agree with gender or in number with their noun. However, there are some exceptions. Some of the exceptions are illustrated in the examples below.

Example 2.31	*Plaid Werdd (Wales)*	*Green Party*
Example 2.32	*Llaw fer*	*Shorthand*
Example 2.33	*Tylluan wen*	*Barn owl*
Example 2.34	*Arth wen*	*Polar bear*

Research has found that the gender distinctions inherent in some adjectives are being lost in Welsh, where the agreement marking in adjectives is no longer productive (Watkins, 1993). The feminine adjectives forms are mostly optional, as masculine forms can be selected with feminine nouns even when those particular adjectives have feminine forms (Mittendorf & Sadler, 2006). However, Thomas

(2001) argued that the two adjective forms for *white,* ‘gwyn’ [m] and ‘gwen’ [f], have remained in both spoken and written Welsh.

In addition to a limited number of gender marked adjectives, some morphological generalisations can be made. There are trends in the language to help infer the gender of the noun, but these are not absolute (Hammond, 2016). For example, the following endings are usually feminine: *-aeth, -iaeth, -as, -fa -ach, -en, -ell* while the following endings are usually masculine*: -ad, -aint, -awd, -od, -dod, -deb, -der, -ter, -did, -tid, - dra, -dwr, -edd, -had, -i, - iad, -iant, -ineb, -ni, -ioni, -id, -rwydd, -wch, -yd, -yn, -cyn, -we* (Thomas, 2001). There are exceptions, however, for example, *triwanti**aeth*** “truancy” is grammatically masculine but ends in *-aeth* a suffix usually associated with feminine nouns, and *gwa**wch*** “squawk” is grammatically feminine, but ends in *–wch* a masculine related suffix. However, most words ending with these suffixes are considered ‘literary forms’ and many of these forms do not feature regularly in normal, day-to-day speech (Hammond, 2016). There are other features and patterns beyond noun form that can be used to identify the gender of a noun. These are listed below (King, 2016):

- Verbal nouns are always masculine
- Semantically related nouns usually share the same gender, although exceptions exist, for example:
  - Days, months, and seasons are masculine.
  - *Gwyl* “festivals” are feminine, main holidays are masculine.
  - Rivers and streams are feminine, while placenames that begin with Nant- (“stream”) tend to be masculine.
  - Countries, cities, towns, and villages tend to be feminine; however, names for land and area are masculine.
  - Fruits and vegetables are usually feminine, however masculine exceptions exist, such as *lemon* “lemon”, and *afal* “apple” and *oren* “orange” can be either gender.

Whilst such regularities may help a learner and/or speaker of Welsh build knowledge of the system, it remains the case that the irregularities and exceptions are widespread that one cannot rely solely on such patterns whilst building the system. There are variations from regions to region, where a number of nouns that differ in gender depending on geographic region / location and dialect, including (examples extracted from *The Welsh Academy English-Welsh dictionary*, Griffiths & Jones, 2014):

Masculine in standard Welsh, but feminine in North Wales.

- Castell *Castle*
- Cwpan *Cup*
- Tywydd *Weather*

Masculine in standard Welsh, but feminine in South Wales.

- Cinio *Lunch*
- Cyflog *Wage*
- Cyngerdd *Concert*

Feminine in standard Welsh, but masculine in North Wales.

- Breuddwyd *Dream*
- Crib *Summit*
- Eiliad *Second (in time)*

While the nouns listed above differ according to geographic location and dialect, there are some nouns that do not seem to be fully integrated into the gender of choice, e.g., ‘munud’ (minute) is masculine in North Wales and feminine in South Wales. There appears to be no reason for these differences. There are also some nouns which are marked for both genders [mf], indicating that they are both masculine and feminine, including:

- Troed *Foot*
- Sbanner *Spanner*

Additionally, there are certain nouns which have more than one Welsh translated form, according to North and South Wales. The two nouns are typically marked for the same gender; however, they do not share lexical similarities. For instance:

Table 2.2

Forms According to North and South Wales

English form	N. Wales	S. Wales
Boy	Hogyn [m]	Bachgen [m]
Girl	Hogan [f]	Merch [f]
Woman	Dynes [f]	Menyw [f]
Grandad	Taid [m]	Tad-cu [m]

There are certain nouns which have more than one form and differ in gender (unlike the examples above). An example of this is given below, where both translated forms are accepted in written and spoken Welsh and are not a consequence of geographic location.

Example 2.35 Table Bord [f](NW) Bwrdd [m] (SW)

There are also nouns which differ in gender marking according to different Welsh dictionaries. For instance, *The Welsh Academy English-Welsh dictionary* (Griffiths & Jones, 2014) marks 'sbanner' (spanner) for both genders [mf], while *Y Geiriadur Mawr: The Complete Welsh-English, English-Welsh Dictionary* (Meurig Evans & Thomas, 2021, first published in 1958) only marks 'sbanner' for masculine gender [m]. There are also examples of nouns which appear to have changed gender across time. For instance, in *Y Geiriadur Mawr: The Complete Welsh-English, English-Welsh Dictionary* (Meurig Evans & Thomas, 2021), the noun 'pensel' (pencil) is marked for feminine gender [f], however, in the 1995 edition of the same dictionary, 'pensel' (pencil) is marked for masculine gender [m].

These irregularities and inconsistencies within the system makes grammatical gender in Welsh highly opaque and likely to be more difficult for learners to acquire. This is particularly the case as the grammatical gender system in Welsh is also intertwined with the mutation system. This discussion point is revisited in the following section.

Unlike some other grammatical gendered languages, such as French, Spanish, and German, gender is not marked on the determiner in Welsh. The Welsh definite article is 'y', which appears before a consonant, between two consonants or between a consonant and a consonantal 'w'. The definite article also changes to 'yr' in front of a vowel, diphthong, or ‘h’. It also appears between a word ending with a consonant and starting with a vowel and it can appear as ‘’r’ after a vowel or diphthong. In Welsh, there is no indefinite article (‘a’). The examples below show the different forms.

Example 2.36	*Y*	*seren*	
	The DET	star NOUN (f)	
	‘The star’		
Example 2.37	*Yn*	*y*	*siop*
	In PREP	the DET	shop NOUN (f)
	‘In the shop		
Example 2.38	*Yr*	*eglwys*	
	The DET	church NOUN (f)	
	‘The church’		
Example 2.39	*Yr*	*haul*	
	The DET	sun NOUN (m)	
	‘The sun’		
Example 2.40	*O'r*	*tŷ*	
	From PREP the DET	house NOUN (m)	
	‘From the house’		

Although gender is not marked on the definite article, gender is marked locally in some numerals. As seen in Example 2.22 and Example 2.23, numerals appear pre-nominally and the head noun in the numeral phrase remains singular, as there is no number agreement[6]. The numerals in Welsh with both masculine and feminine forms are numbers two, three and four. They mark gender by means of internal vowel shift. In younger speakers’ varieties, a tendency to avoid the gender-marked forms of these numerals has been observed (Frenda, 2011; Jones, 1998). The different forms are given in Table 2.3, followed by examples of these different forms in the DP. The examples below include mutated nouns and numerals. Gender and mutation are discussed in section 2.4.

[6] There is the option of number + preposition ‘o’ (of) + plural form, e.g., twenty men ‘ugain dyn’ or ‘ugain o ddynion’.

Table 2.3

Gendered Marked Numerals in Welsh

Masculine	Numeral	Feminine
dau	*two*	dwy
tri	*three*	tair
pedwar	*four*	pedair

Example 2.41	*Y*	*ddau*	*gastell*
	The DET	two+SM NUMERAL (m)	castle+SM NOUN (m)
	'The two castles'		
Example 2.42	*Y*	*ddwy*	*gath*
	The DET	two+SM NUMERAL (f)	cat+SM NOUN (f)
	'The two cats'		
Example 2.43	*Y*	*tri*	*chastell*
	The DET	three NUMERAL (M)	castle+AM NOUN (m)
	'The three castles'		
Example 2.44	*Y*	*tair*	*cath*
	The DET	three NUMERAL (f)	cat NOUN (f)
	'The three cats'		
Example 2.45	*Y*	*pedwar*	*castell*
	The DET	four NUMERAL (M)	castle NOUN (m)
	'The four castles'		
Example 2.46	*Y*	*pedair*	*cath*
	The DET	four NUMERAL (f)	cat NOUN (f)
	'The four cats'		

In written standard Welsh, gender is also marked in larger numbers, using the masculine and feminine forms for two, three and four. For example:

Example 2.47	*Dau ar bymtheg [m]*	*Seventeen*	*Dwy ar bymtheg [f]*
Example 2.48	*Tri ar ddeg [m]*	*Twenty-three*	*Tair ar ddeg [f]*
Example 2.49	*Pedwar ar hugain [m]*	*Twenty-four*	*Pedair ar hugain [f]*

Gender in Welsh is also marked in demonstratives. These typically follow the noun they modify, and the definite article precedes the noun. They also show agreement for gender and number with the head noun. For example, 'y tŷ hwn' (this house). However, the masculine (dem/pron) *hwn* and *hwnnw*, as well as the neuter (dem/pron) *hyn* and *hynny* can appear before the head noun. In some varieties of colloquial Welsh, it has been noted that masculine and feminine demonstratives in attributive use are replaced by the neuter demonstrative pronoun ‘hyn’. For instance, ‘y dyn hyn’ *this man* (instead of ‘y dyn hwn’), ‘y wraig hyn’ *this woman* (instead of ‘y wraig hon’) (Frenda, 2011). The demonstrative forms are given in the table below.

Table 2.4

Demonstrative Forms in Welsh (see Borsley et al, 2007, p. 176, table 5.5)

	proximal	*distal*	*‘new’ distal*
Masc. dem. / pron.	hwn	hwnnw	hwnna
Fem. dem. / pron.	hon	honno	honna
Neut. Pron. (dem.)	hyn	hynny	hynna
Plural demonstrative	hyn	hynny	hynna
Plural pronoun	rhain	rheini	rheina

Gender in Welsh is not marked straightforwardly, nor is it probabilistically marked (Hammond, 2016). The Welsh grammatical gender system is complex and opaque, with a limited number of gender marked cues. Some of these gender cues are marked via the numerals (as discussed above), however, gender is also marked via mutations. This is because gender is Welsh is in some instances, encoded through mutations or in conjunction mutations, while it is independent of mutations in other instances (e.g., some gendered numerals) (this is discussed in more detail below). The term ‘mutation’ refers to a set of morphophonological alternations that affect certain initial consonants of words, in turn affecting a sub-set of nouns. Although this results in phonological changes, mutations are not triggered by phonological factors, rather, mutations are conditioned by lexical and morphosyntactic factors (Borsley et al., 2007). The grammatical gender and the mutation systems operate in conjunction with one another but the two systems are not interchangeable (Thomas & Gathercole, 2005), as they can also operate independent of one another. The role of mutations in the grammatical gender system is outlined in the following section.

2.4 Mutation and Gender in Welsh

There are three sets of mutations in Welsh. These are the soft mutation (SM) (*treiglad meddal*), the nasal mutation (NM) (*treiglad trwynol*) and the aspirate mutation (AM) (*treiglad llaes*). Only 9 out of the 22 consonants in Welsh undergo mutation. The complete table of mutations is given in Table 2.5.

Table 2.5

Consonant Mutation in Welsh (see Borsley et al, 2007, p. 20, table 1.5)

Radical (original consonant)	**Soft**	**Nasal**	**Aspirate**
p [p]	b [b]	mh	ph [f]
t [t]	d [d]	nh	th [θ]
c [k]	g [g]	ngh	ch [x]
b [b]	f [v]	m [m]	
d [d]	dd [ð]	n [n]	
g [g]	-	ng [ŋ]	
m [m]	f [v]		
ll [ɬ]	l [l]		
rh [r̥]	r [r]		

Only two of the mutations are relevant to grammatical gender – the soft mutation (SM) and aspirate mutation (AM), therefore, the nasal mutation is not discussed any further. Of the two relevant mutations, the AM affects the smallest number of consonants. It changes the voiceless stops /p/, /t/, and /k/ to fricatives /f/, /θ/, and /x/, and orthographically, /p/, /t/, k/ correspond to p, t, c and /f/, /θ/, /x/ correspond to ph, th and ch. However, SM affects the highest number of consonants, and it is a general process of lenition. It changes the voiceless liquids (/ɬ, r̥/) and stops (/p, t, k/) to become voiced (/b, d, g, l, r/), and the voiced stops (/b, d/) and the labial nasal (/m/) to become fricatives (/v, ð, v/), /g/ is deleted. If the target word does not have a mutable consonant, then the target will not change. The SM and AM play a crucial role in the way gender is marked in Welsh and the SM is extremely pervasive, in other contexts as well as gender.

Mutations can be triggered by both syntactic (grammatical relation) and lexical (adjacent items) triggers (Borsley et al., 2007). A large number of mutation triggers are lexical and typically, a mutation is triggered by an adjacent or a preceding lexical item (Thomas, 2001). For instance, many of these lexical triggers are high frequency items, including prepositions, as well as some adverbs and possessives. To illustrate this, the preposition ‘i’ (English: to) is a trigger of soft mutation, giving ‘i **G**ymru’ (to Wales) from ‘i Cymru’ (Cymru+SM, c > g). In these instances, SM does not function as a marker of gender.

There are two types of lexical triggers which are related to grammatical gender. The first are *pure lexical triggers* (Borsley et al., 2007). Here, the triggers always "cause" a particular mutation of any word that follows them. The second are *categorical lexical triggers* (Thomas, 2001). For instance, feminine nouns trigger SM on following adjectives. Although the mutation is not triggered by a specific lexical item, it is triggered by a morpho-syntactic feature [+feminine]. Borsley et al (2007) have proposed three approaches to consider the mutations triggered by feminine singular nouns on following adjectives, questioning whether mutations are phonology or morphosyntax.

The first is a *phonological* approach, which suggests that the mutation has a lexical trigger, and this is phonological. It is argued that there is floating auto-segment at the end of every feminine noun which re-attaches itself to the following word, and this is interpreted by the phonological component as an instruction to change the initial segment of that word. The second is a *morphosyntactic* approach, which suggests that the mutation has a morphosyntactic trigger. It is proposed that the feminine noun triggers a mutation on the first phrase that it immediately precedes and c-commands (i.e., a following sister or its descendants). The third is an *agreement* approach, which maintains that the mutation is an agreement operation, where feminine adjectives agree with their head nouns and the mutation is a morphological manifestation of this agreement. These various explanations suggest that it is unknown as to what mutations are, questioning whether mutations are phonology or morphosyntax, particularly in this context where mutations are triggered by feminine singular nouns on following adjectives.

As noted previously, the definite article in Welsh ‘y’ does not inflect gender. However, it triggers SM on following feminine singular nouns (head noun) and the postnominal adjective but the definite article does not trigger SM on masculine nouns (Thomas, 2001). For instance:

Example 2.50	*Y*	*gath*	*dew*
	The $_{\text{DET}}$	cat+SM $_{\text{NOUN (f)}}$	fat+SM $_{\text{ADJ}}$
	‘The fat cat’		

Example 2.51	*Y*	*goeden*	*werdd*
	The DET	tree+SM NOUN (f)	green+SM ADJ
	'The green tree'		

Here, the SM is a "local" marker in the sense that the marker occurs on the noun itself or "its immediate modifiers" (Gathercole et al., 2001). Adjectives following feminine nouns undergo SM (while adjectives following masculine nouns do not mutate). The SM is not associated with singular masculine nouns, nor is it associated with plural nouns of either gender in "local" gender marked contexts (i.e., following the definite article). There is also no soft mutation on the modifying adjectives.

Example 2.52	*Y*	*cathod*	*tew*
	The DET	cat-PL NOUN (f)	fat ADJ
	'The fat cats'		
Example 2.53	*Y*	*ci*	*tew*
	The DET	dog NOUN (m)	fat ADJ
	'The fat dog'		
Example 2.54	*Y*	*cwn*	*tew*
	The DET	dogs-PL NOUN (m)	fat ADJ
	'The fat dogs'		

The SM only applies to the sub-set of letters (outlined in Table 2.5). The SM with feminine nouns only applies to these sub-set of letters, however, 'll' and 'rh' are exceptions and do not mutate with feminine nouns, meaning that feminine and masculine nouns are often indistinguishable from one and other. Feminine nouns borrowed from English also do not undergo SM, e.g., *gêm* (game) and *gôl* (goal).

The definite article also triggers SM on the cardinal number 'two', for both the feminine and masculine forms (dau & dwy) and the SM on *dau (m)* and *dwy (f)* are said to be stable in spoken Welsh (Borsley et al., 2007). Both gendered numeral forms also trigger SM on following nouns. However, only the adjective following the feminine noun undergoes SM, while the adjective following the masculine noun does not. The noun following the numeral 'two' is always singular in form and does not 'agree' in number (i.e., it is not plural). For example:

Example 2.55	*Y*	*ddwy*	*gath*	*ddu*
	The DET	two+SM NUMERAL (f)	cat+SM NOUN (f)	black+SM ADJ
	'The two black cats'			
Example 2.56	*Y*	*ddau*	*gi*	*du*
	The DET	two+SM NUMERAL (m)	dog+SM NOUN (m)	black ADJ
	'The two black dogs'			

The numbers 'three' and 'four' also agree in gender with the head noun, but the definite article does not trigger SM on these cardinal numerals, irrespective of the head noun. The nouns following the feminine and masculine forms of the numeral four, *pedwar* and *pedair*, are produced in their bare form (i.e., no mutation). This is also the case for the noun following the feminine numeral three, *tair*. However, the noun following the masculine form of the numeral three *tri* undergoes aspirate mutation (AM), however, this is arguably used erratically in the spoken language (Binks, 2017). Adjectives following masculine nouns are not mutated, while adjectives following feminine nouns are SM for numerals three and four. For example:

Example 2.57	*Y*	*tair*	*cath*	*ddu*
	The DET	three NUMERAL (f)	cat NOUN (f)	black+SM ADJ
	'The three black cats'			
Example 2.58	*Y*	*tri*	*chi*	*du*
	The DET	three NUMERAL (m)	dog+AM NOUN (m)	black ADJ
	'The three black dogs'			
Example 2.59	*Y*	*pedair*	*cath*	*ddu*
	The DET	four NUMERAL (f)	cat NOUN (f)	black+SM ADJ
	'The four black cats'			
Example 2.60	*Y*	*pedwar*	*ci*	*du*
	The DET	four NUMERAL (m)	dog NOUN (m)	black ADJ
	'The four black dogs'			

The number one '*un*' does not have separate masculine and feminine forms, but feminine nouns following '*un*' undergo a soft mutation, but masculine nouns following '*un*' do not mutate. This is the same as when singular nouns follow the definite article. For example:

Example 2.61	*Un*	*ferch*	/	*Y*	*ferch*
	One	girl+SM $_{\text{NOUN (f)}}$		The $_{\text{DET}}$	girl+SM $_{\text{NOUN (f)}}$
	'One girl'			'The girl'	
Example 2.62	*Un*	*bachgen*	/	*Y*	*bachgen*
	One	boy $_{\text{NOUN (m)}}$		The $_{\text{DET}}$	boy $_{\text{NOUN (m)}}$
	'One boy'			'The boy'	

The acquisition of gender agreement between numeral and noun is typically unproblematic (Thomas, 2001) and gender marking within the noun phrase in the speech community, e.g., 'y gath dew' and 'un ferch', continues to remain in use and appears to be stable in local gender marked constructs (Frenda, 2011; Thomas & Gathercole, 2005).

It is worth highlighting and discussing in more depth, that gender operates in three different ways - independently of mutations, in conjunction with mutations and encoded through mutations. Some examples of these different instances have already been given above, however, the table below shows clear examples of these instances, before discussing them in more detail.

Table 2.6

Examples of How Gender Operates (with and without mutations)

Instance	Linguistic property	Examples	Locality
Gender encoded through mutations	Definite article	'y' + SM [f] 'y' + no mutation [m]	Local
	3rd person sing proclitic	'ei' + AM [f] 'ei' + SM [m]	Non-local
Gender in conjunction with mutations	Numeral two	'dwy' + SM [f] 'dau' + SM [m]	Local
	Numeral three	'tair' + no mutation [f] 'tri' + AM [m]	Local
Gender independent of mutations	Numeral 4	'pedair' + no mutation [f] 'pedwar' + no mutation [m]	Local
	3rd person sing pronouns	'hi' 'fe'	Local and non-local

Table 2.6 shows examples of how gender operates with and without mutations. When gender is encoded through mutations, there are two examples. The first is via the definite article ‘y’ (the), because ‘y’ triggers SM on following feminine nouns but not masculine nouns. The second are the mutations following the third-person singular proclitic *ei* (his / her). Although in Welsh, the SM functions as a marker for feminine gender in “local” contexts, the SM also functions as a marker for masculine gender in “non-local” contexts, i.e., when there is distance between agreeing elements such as the agreement trigger and agreement target (Chondrogianni, 2024). The third person possessive adjectives in non-local marking must agree with the gender of the antecedent noun within the sentence, and this agreement is marked by different types of mutation that appear on the following word. After *ei*, there is a SM on the noun to mark masculine gender while AM on the noun marks feminine gender. Table 2.7 shows the agreement of possessive forms with antecedents, followed by two examples.

Table 2.7

3rd Person Possessive Adjectives in Welsh

Type	Form	Mutation / No mutation
3rd person singular possessive: masculine	His *‘ei’*	SM marks agreement with a masculine antecedent
3rd person singular possessive: feminine	Her *‘ei’*	AM marks agreement with a feminine antecedent
3rd person plural possessive: neuter	Their *‘eu’*	Triggers no mutation(s)

Example 2.63	*Mae'r*	*bachgen*	*wedi*	*colli*	*ei*	*gi*
	Be 3SG + the DET	boy NOUN (m)	has VERB	lost VERB	his PRON	dog+SM NOUN (m)
	‘The boy has lost his dog’					
Example 2.64	*Mae'r*	*ferch*	*wedi*	*colli*	*ei*	*chath*
	Be 3SG + the DET	girl NOUN (f)	has VERB	lost VERB	her PRON	cat+AM NOUN (f)
	‘The girl has lost her cat’					

This is one of the very few instances where AM plays a role in the gender system, and it is less pervasive than the SM. It is possible that this may cause difficulty in acquiring the gender system and using SM correctly when gender is encoded through mutations, as SM functions as a local marker for feminine gender but a non-local marker for masculine gender. Previous research has noted that the use of AM is often avoided in natural speech by many speakers, with use varying from region to region and that certain environments which are formally contexts for AM tend to be supplanted by SM in the modern spoken language (King, 2016). However, the AM following *ei* to mark feminine gender and agree with the feminine antecedent noun is arguably one of the most robust environments for AM in the spoken language, where it is frequent after the feminine possessive adjective *ei* (Borsley et al., 2007; Thomas, 2001).

Gender also operates in conjunction with mutations, as shown in Table 2.6. An example of this is the numeral two. As shown in Example 2.55 and Example 2.56, the number two in Welsh has masculine and feminine forms (‘dau’ [m] & ‘dwy’ [f]). Both gendered forms trigger SM on the following noun, e.g., ‘Dau gi (c>g)’ (‘Two dogs’ [m]) and ‘Dwy gath (c>g)’ (‘Two cats’ [f]). Here, the speaker SM the target noun if it starts with a mutatable letter, following the two gendered forms. However, if the noun starts with a letter that does not participate in the mutation system (see Table 2.5), then there is no mutation. Another example is the numeral three. The numeral *three* also has masculine and feminine forms (‘tri’ [m] & ‘tair’ [f]). There is an AM on masculine nouns following *tri* [m] but there is no mutation on feminine nouns following *tair* [f], e.g., ‘Tri chi (c>ch)’ (‘Three dogs’ [m]) and ‘Tair cath’ (‘Three cats’ [f]). The speaker must differentiate between masculine and feminine nouns, understanding that there is AM on masculine nouns but no mutation on feminine nouns.

In contrast, the gender system also operates independently of mutations and an example of this is the numeral four, as shown in Table 2.6. As shown in Example 2.59 and Example 2.60, the number four in Welsh has masculine and feminine forms (‘pedwar’ [m] & ‘pedair’ [f]). Neither gendered forms trigger any mutation on the following nouns, e.g., ‘Pedwar ci’ (‘Four dogs’ [m]) and ‘Pedair cath’ (‘Four cats’ [f]). Here, the speaker must understand that nouns following these two gendered forms are produced in their bare form, with no mutation. Another example of when gender operates independently of mutations, is via the third person singular pronouns ‘fe’ / ‘hi’ (he / she). The two pronouns mark masculine and feminine gender and these pronouns are instances of gender when it does not operate in conjunction with mutations, rather, it is independent of mutations. Notably, there is no form for the

English word ‘it’ in Welsh, therefore, nouns are referred to as ‘fe’ / ‘hi’ when in the singular form.

Given the complexity of the gender system and its intricate relationship with the mutation system, it is important to understand the interpretability of these different instances of gender when it is encoded through mutations, in conjunction with mutations and independent of mutations. When gender operates independently of mutations via the forms ‘pedwar’ [m] / ‘pedair’ [f] + no mutation, this may be considered a relatively straightforward acquisition process for the Welsh learner. This is because ‘pedwar’ / ‘pedair’ are interpretable due to their specific numeral values. Additionally, the absence of a mutation on the following target noun (i.e., the noun is produced in bare) is also interpretable, as this contributes directly to the interpretation of syntactic structure, and it is a consistent feature of the language. With regards to the 3rd person singular pronouns ‘fe’ / ‘hi’ (he / she), these are also interpretable, given that they mark for gender, and this is a well-established feature in Welsh.

When gender operates in conjunction with mutations via the forms ‘dau’ [f] / ‘dwy’ [m] + SM, this may also be considered a relatively straightforward acquisition process for the learner. This is due to the interpretability of the gendered forms ‘dwy’ / ‘dau’ given their semantic content, indicating a specific numerical value. The SM on the following nouns may also be considered interpretable because this is a consistent and predictable grammatical rule in Welsh. For both the absence of mutation following ‘pedwar’ / ‘pedair’, and the presence of mutation following ‘dau’ / ‘dwy’, Welsh speakers likely intuitively understand these rules as part of the language's morphophonological system. The second example of gender operating in conjunction with mutations, via the gendered numeral forms ‘tri’ [m] and ‘tair’ [f], are also interpretable due to their numerical value. However, they trigger different instances of mutation (‘tri’ + AM) / no mutation (‘tair’ + no mutation) on post-numeral nouns. It could be argued that the presence or absence of the mutation reflects an uninterpretable feature.

However, the picture is more complex with regards to instances of gender when it is encoded through mutations. For example, the definite article ‘y’ (the) is interpretable, because it carries meaning related to definiteness. Yet, it must carry an uninterpretable feature as it triggers something different to happen on the following masculine and feminine nouns with regards to mutation (SM [f] / no mutation [m]).

This is also the case with the third-person singular proclitic '*ei*' (his / her), where the pronoun itself is interpretable, but the mutations on the following nouns are reflecting an uninterpretable feature as they must agree with the gender of the antecedent noun, and this agreement is marked by SM for masculine gender and AM for feminine gender. These are two instances of gender when it is encoded through mutation(s), with its complexity lending itself to the fact that one must understand the mutation in order to understand the gender. It may be possible that when gender is encoded through mutations, it becomes problematic due to the opaqueness of the features.

The description of the Welsh gender and mutation systems has detailed some of the properties which make acquiring Welsh gender a complex task. These properties include the fact that the definite article 'y' does not inflect gender, but rather triggers SM on following feminine nouns and no mutation on following masculine nouns. Therefore, 'y' may not serve as a reliable cue to anticipate upcoming information regarding the nouns gender. Additionally, the SM is multifunctional, in that SM occurs in contexts unrelated to gender, such as prepositions, while for gender, it is a marker of feminine gender in local contexts but is a marker of masculine gender in non-local contexts. This lack of one-to-one correspondence between noun form, the gender it encodes and mutations, results in a complex form-function mapping between gender and mutations. Further to this, the role of distance (i.e., local / non-local) and the consequence of this in relation to gender and mutations, also contributes the opaqueness of the gender system.

There are several other factors which contribute to the complexity of the Welsh gender system. For example, the gaps in the system, in the sense that SM is used in local gender marking and is constrained to singular feminine nouns, and that there are no gender distinctions in the plural forms in either gender. Also, gender distinctions can only be expressed on the surface of nouns and adjectives, in that feminine nouns that begin with consonants that do not undergo SM after the definite article are indistinguishable from masculine nouns. Lastly, noun forms can provide information about noun gender, however, whilst such forms exist, the numbers of exceptions outweigh their applicability.

The Welsh gender and mutation systems have been discussed in terms of their 'prescriptive' nature. However, in spoken Welsh (and even in written Welsh), gender and mutations are used in a variable manner (Thomas, 2001). This variation exists between and within speakers, which is a likely consequence of their level of competence in the language and/or their frequency of exposure to gender-marked constructs. A given speaker may not mutate at all, they may use different mutations to the above ‘rules’ (i.e., overextending SM to AM), or that they may not mutate consistently, in that they may sometimes mark a noun as one gender in one sentence but go on to mark the same the noun as a different gender in subsequent sentences. For example, a child might hear any number of the following variations in speech: y *menyw dal, y *menyw tal, y fenyw dal, y fenyw *tal “the tall woman”, making it more difficult to abstract out clearly defined regularities from within the system. Sometimes, the level of ‘accuracy’ is far below the perceived ‘target level’ according to the prescriptive nature of the two systems (Binks, 2017; Thomas, 2001). In addition to the fact that adults’ use of the system can vary greatly within and across mutation types, there are also some irregularities and inconsistencies in the gender marking of certain nouns. There are nouns which differ in gender according to region (i.e., NW > SW) and according to different dictionaries, while some nouns are marked for both genders and other nouns have changed gender over time.

All of these factors contribute to the opacity and complexity of the gender system in Welsh, which can make it a difficult task to master. While some aspects of gender are encoded through mutations and the two systems work in conjunction with one another, gender can also operate independently of mutations. Thomas (2001) postulated that even though the Welsh grammatical gender system is not robust, it is argued that the gender distinctions are robust, suggesting that it is the use of the mutation that is variable.

The Welsh gender system is claimed to be more complex than that of many other gender-marked languages in part due to its interaction with the mutation system (see Borsley et al., 2007; Thomas & Mayr, 2010; Thomas & Gathercole, 2007, for a fuller discussion of mutation and gender). Therefore, when acquiring gender in Welsh, Thomas and Gathercole have argued that it is not possible to acquire the gender system without mastering the mutation system (Thomas & Gathercole, 2007). Moreover, the acquisition of Welsh is in a minority language context, with English often as the dominant language in the community (some communities in NW remain Welsh dominant). Therefore, the frequency of occurrence of the language in its prescriptive state may be more limited, if not inconsistent and highly varied in some cases. It is suggested that the irregularities and exceptions are so widespread that children cannot rely solely on such patterns whilst building the system (Binks, 2017). Consequently, is it not just the *quantity* of input that is important but also the *quality* of that input (Unsworth, 2008). The literature investigating the acquisition of Welsh in children is briefly reviewed in section 2.7, but first, the section below discusses how the different linguistic properties of gender, such as transparency / opacity of features, the presence / absence of morphophonological cues and the locality of gender effects / structural distance, make some gendered languages easier to acquire than others.

2.5 Linguistic properties and the acquisition of grammatical gender

Children acquire grammatical gender in their first and second language at different rates, and this is largely down to the several linguistic and individual factors which play a role in the acquisition process (Chondrogianni, 2024). There are various linguistic properties of gender that make some gendered languages easier and faster to acquire than others (Chondrogianni, 2024). For instance, different languages are discussed in terms of how *transparent* or *opaque* a gender system is, and how this affects the acquisition of gender. One linguistic property which is discussed in terms of its transparency or opacity, are linguistic *cues* to gender. Cues to gender are marked in various forms across languages, such that it can sometimes be seen in determiners, noun endings and the modifying words in the construct, such as adjectives. In some languages, determiners, noun endings and adjectives can all be used as cues. However, in other gendered languages, there are fewer cues, and it may be that only one (or two) of the three aforementioned properties can be used to indicate gender. As a result, some learners are able to rely on several cues, while others can only rely on one cue (Grüter, Lew-Williams, & Fernald, 2012).

It is claimed that grammatical gender in Spanish is transparent. Spanish has a binary gender system (masculine and feminine) and gender is marked in the determiners ‘el’ [m] and ‘la’ [f] (Teschner & Russell, 1984). Research has shown that gender agreement between the definite article and noun in monolingual Spanish, is acquired as early as 3 years of age (Arias-Trejo et al., 2013; Pérez-Pereira, 1991). Research has also found that simultaneous and early sequential bilingual children (Spanish & English) lag slightly behind monolingual children of the same age, yet, this difference is largely attributed to lower levels of input and exposure to Spanish (e.g., Gathercole, 2002; Montrul & Potowski, 2007).

Similar to Spanish, French also marks gender via the determiners ‘le’ [m] and ‘la’ [f]. Despite the fact that ‘le’ [m] and ‘la’ [f] serve as markers of gender in French, research has shown that only 41.22% of noun tokens had a gender-marked determiner, while 49.76% of noun tokens are not gender-marked and the remaining 9.01% of the tokens were modified by gender-marked adjectives (Ayoun, 2010). Research exploring the acquisition of determiner-noun agreement patterns has found that it is acquired with relative ease in French monolingual children (e.g., Dewaele & Véronique, 2000; Pizzuto & Caselli, 1992). Similarly, for L1 bilingual children aged three, research has found that they use phonological cues, such as determiners, to determine the gender of the noun (Carroll, 1989). However, when French is acquired slightly later between the ages of three-six and is the less dominant of the two languages, error rates are typically higher (Granfeldt, Schlyter, & Kihlstedt, 2007; Granfeldt, 2018).

Arguably, determiners in both Spanish and French serve as clear and consistent phonological cues to grammatical gender. However, in contrast, the definite article in Welsh ‘y’ does not inflect gender - there are no masculine and feminine forms. As outlined in the previous section(s), ‘y’ triggers SM on following feminine singular nouns, but no mutation on masculine nouns. Therefore, the determiner ‘y’ itself does not serve as a cue for gender, rather, the potential cue is the presence of the SM on feminine nouns, or absence of no mutation on masculine nouns, which is seen on the noun itself and not prior to it. However, because mutations only involve a subset of letters (9/22 consonants), the presence or the absence of the mutation does not arguably serve as a ‘reliable’ cue, given that not all nouns are involved in this cue signal.

Languages also have other cues to gender, such as adjectives. Gender is visible on adjectives in Italian. For example, the adjective ‘nuovo’ [m] (new) agrees in gender with the masculine noun ‘gioco’ (toy) with the ending ‘o’, e.g., ‘Il gioco nuovo’ (the new toy), while to agree in gender with a feminine noun such as ‘casa’ (house), the suffix changes to ‘a’, e.g., ‘La casa nuova’ (the new house). Similarly, gender is also visible on adjectives in French. The adjective *new* also has masculine and feminine forms to agree in gender. For instance, ‘nouveau’ [m] agrees in gender with the masculine noun ‘jouet’ (toy) with the suffix ‘eau’, e.g., ‘Le nouveau jouet’ (the new toy). Whereas, to agree in gender with a feminine noun such as ‘maison’ (house), ‘nouvelle’ is produced with the feminine noun, e.g., ‘La nouvelle maison’ (the new house). Russian, which has a tripartite gender system - masculine, feminine and neuter - also distinguishes gender on singular adjectives. Adjectives which appear pre-nominally, have different endings to show agreement for gender (and case). For example, in the nominative case, ‘новый’ [m] agrees in gender with the masculine noun ‘город’ (city) with the suffix ‘ый’, e.g., ‘новый город’ (the new city). In contrast, to agree in gender with a feminine noun such as ‘машина’ (car), ‘новая’ with the suffix ‘ая’ is needed, e.g., ‘новая машина’ (the new car).

In (standard) Welsh, there are a small set of adjectives that have marked feminine forms that agree in gender with the noun. For example, *trwm* [m] and *trom* [f] (‘heavy’), *gwyn* [m] and *gwen* [f] (‘white’), and *byr* [m] and *ber* [f] (‘short’). In context, it is possible to say ‘*Dyma'r ddawnswraig, mae hi'n **fer**’* (Here is the (female) dancer, she is short). However, the feminine adjectives forms are mostly optional, as masculine forms can be selected with feminine nouns even when those particular adjectives have feminine forms (Mittendorf & Sadler, 2006). For example, the sentence ‘*Dyma'r ddawnswraig, mae hi'n **fyr**’* is also acceptable (‘fyr’ in place of ‘fer’). The majority of adjectives in Welsh appear post-nominally, with only a subset appearing pre-nominally, such as ‘hoff’ (favourite), ‘hen’ (old) and ‘prif’ (main). These do not have masculine and feminine forms, and they trigger SM on the following noun, irrespective of its gender (but only when it starts with a mutatable letter). Additionally, an adjective following a singular feminine noun undergoes a SM, if it begins with a letter that participates in the mutation system. For example, ‘Merch fach (b>f)’ (‘little girl’ [f]). Therefore, it is not possible to say that adjectives in Welsh act as a ‘reliable’ cue to gender, as is seen in some other gendered languages.

In addition to determiners and adjectives, the morphophonological information of the noun ending can also be used to classify nouns into gender classes (Chondrogianni, 2024). For example, in Spanish, 96.3% of nouns ending with -*a* are feminine (e.g., *la rosa* [f]– ‘the rose’), and 99.9% of nouns ending with -*o* are masculine (e.g., *el teatro* [m]– ‘the theatre’), but not all nouns end in -*a* and -*o* (Teschner & Russell, 1984). Gender is said to be predictable for 95% of the nouns in Spanish with these two endings, which contributes to the overall transparency of the gender system (Teschner & Russell, 1984). Grammatical gender in Italian is also claimed to be transparent because about 70% of inanimate nouns follow the predictable pattern of -*o/-i* for single/plural masculine and -*a*/-*e* for singular/plural feminine nouns (Ayoun & Maranzana, 2022). There are nouns which fall outside this pattern. These are said to be phonologically opaque and are considered as exceptions. For instance, the vocalic ending does not clearly indicate the noun's gender, and these represent approximately 20% of nouns, such as *mano [f] [sg]* ('hand') and p*roblema [m] [sg]* ('problem'). Moreover, nominal endings in -*e* can be either feminine, (e.g., *fede* 'faith') or masculine (e.g., *mare* 'sea'), while those ending in -*ù* or -*i* in the singular are usually classified as irregular and feminine (e.g., *crisi* 'crisis') (Ayoun & Maranzana, 2022).

While in French, although certain stems are associated with a gender of a noun, which suggests that gender is marked probabilistically (Foucart, 2008), Ayoun (2010) found that 49.76% of noun tokens are not marked for gender. It is also thought that 96% of nouns ending with 'ette' are feminine, 99% of nouns ending with 'ain' are masculine and that overall, 60% of French nouns have high predictive endings (Tucker et al., 1997). However, these rules are probabilistic, in the sense that they are reliant to different degrees, and for all rules there are exceptions, causing a lack of consistency in the system (Ayoun, 2010). Interestingly, children as young as three are sensitive to gender noun-endings in French, particularly to feminine noun-endings more so than masculine noun-endings (Seigneuric, Zagar, Meunier, & Spinelli, 2007). In languages like Dutch, however, nouns carry no morphophonological marking, unless they are in the diminutive form, which means that Dutch provides few morphophonological cues for the learner to assign gender to a specific noun (Blom & Vasić, 2011).

In Welsh, some morphological generalisations can be made with regards to noun endings. There are trends in the language to help infer the gender of the noun (Hammond, 2016). Using the CEG electronic corpus of Welsh (Ellis et al., 2001), Hammond (2016) found that using the following masculine suffixes, *-deb, -iant, -yn, -iad, -wr, -ydd, -wch, -ter/der, -rwydd* and the following feminine suffixes, *-aeth, -en, -wraig, -es, -fa,* speakers can correctly assign gender to 37,165 word tokens (20%). However, there are exceptions to some of these suffix generalisations (Hammond, 2016). For example, the words '*Llafariad'* ('vowel') ending with the suffix -*iad* and

‘Morhwch’[7] (‘dolphin’) ending with the suffix -*wch* are both feminine, and the word ‘*Hiraeth*’ (longing / nostalgia) is masculine, despite ending with the typically feminine suffix -*aeth*.

Previous research has suggested that learners rely on co-occurrence relations between nouns and gender marked modifiers to detect a noun’s gender (Grüter et al., 2012). The computation of co-occurrence relations has been shown to play a key role in children’s early acquisition of grammatical gender (Saffran, Newport, & Aslin, 1996). Therefore, for a gendered language such as Spanish, which has a transparent gender system (i.e., marking gender via the determiner and in noun endings), learners are able to rely on frequent co-occurrence relations between the gender marked modifiers and the noun. These computations are said to strengthen across time, and this allows frequent co-occurrence of sequences received in the input to become powerful cues during language learning (Grüter et al., 2012; Unsworth, 2014). However, in contrast, for a gendered language such as Welsh, if the gender marker is not a frequently co-occurring trigger of gender in the language (e.g., determiner ‘y’, adjectives & noun-endings), it will not act as a reliable cue to indicate gender.

In addition to the afore-mentioned linguistic properties that make gender easier to acquire in some languages than others, such as the determiner, noun endings and gender marked modifiers (e.g., adjectives), another such feature is the role of distance. Distance between agreeing elements affects children’s and adults’ ability to produce and comprehend related structures in real time (Chondrogianni, 2024). Notably, studies have reported that target-like use on gender marking diminishes as the distance between the noun and the modifier increases (e.g., Keating, 2010). For instance, Keating (2010) examined the role of distance on the processing of gender agreement in advanced L2 learners of Spanish (L1 English). Gender agreement between the noun and adjective was violated, and the distance between these were manipulated - placing either one, four or seven words between the agreeing elements. For example, the distance between the noun and adjective in ‘La casa es viejo’ (The house is old) is one word (the verb *es* ‘is’), whereas the distance between the noun and adjective in ‘La casa de la esquina es viejo’ (The house on the corner is old) is four words (*de la esquina es* ‘on the corner is’). Keating investigated whether the

[7] According to ‘Y Geiriadur Mawr’ (Evans & Thomas, 2021), ‘Morhwch’ is feminine. However, according to ‘The Welsh Academy English-Welsh Dictionary’ (Griffiths & Jones, 2014), ‘Morhwch’ is both masculine and feminine (marked *mf*).

increase in distance from one to four (and to seven) words plays a role in the sensitivity to gender agreement violations. The results showed that the L2 learners did not show sensitivity to agreement mismatches on predicative adjectives that are more distant from the corresponding noun, but only for local agreement. Keating interpreted these results as indicating that the learners had acquired grammatical gender agreement, but that distance affected the detection of violations outside of the DP.

In Welsh, as so few adjectives are marked for gender (see above and section 2.3 for examples), distance between agreeing elements manifests differently compared to the Spanish examples provided above. One of the ways in which gender has been examined in 'distant' or 'non-local' contexts in Welsh (i.e., distance between agreeing elements such as the agreement trigger and agreement target), is via the use of the 3^{rd} person singular pronouns *hi* / *fe* "her/him" (he/she) or by the mutation triggered after the third-person possessive adjective *ei* "his/her". In these instances, the non-local marking within a sentence must be in agreement with the gender of the antecedent noun (Binks & Thomas, 2019; Gathercole, Thomas, & Laporte, 2001). Example 2.65 shows the use of the 3^{rd} person singular pronoun and example 2.66 shows the use of the 3^{rd} person possessive adjective 'ei'.

Example 2.65 *Mae cath yn y ffenestr. Mae hi'n ddu.*
Be $_{3SG}$ cat $_{NOUN\ (f)}$ in $_{PREP}$ the $_{DET}$ window $_{NOUN\ (f).}$ Be $_{3SG}$ she $_{PRO\ 3SG\ (f)}$ black $_{ADJ}$
'There is a cat in the window. She is black.'

Example 2.66 *Mae'r plismon yn gwisgo esgid las ar*
Be $_{3SG\ +}$ the $_{DET}$ policeman $_{NOUN\ (m)\ PART}$ wearing $_{VERB}$ shoe $_{NOUN\ (m)}$ blue $_{ADJ}$ on $_{PREP}$
ei droed.
his $_{POSS\text{-}ADJ}$ foot $_{NOUN\ (f)+SM}$
'The policeman is wearing a blue shoe on his foot.'

These two different instances are examples of gender operating without mutations (first example: i.e., *hi* / *fe* "her/him") and gender operating with mutations (second example i.e., *ei* "his/her"), where gender is encoded through mutations, in non-local contexts. When gender is encoded through mutations via *ei*, gender marking is intertwined with the mutation system, therefore, one must have knowledge of mutations in order to produce / comprehend gender (Thomas, 2001).

However, this is only for the items that can show mutations, as only 9 / 22 consonants in Welsh undergo mutation, therefore, ei +SM (marking masculine gender) and ei +AM (marking feminine gender) only affect a subset of nouns in Welsh. Therefore, it is not possible to do one without the other, in this instance of non-local gender marking. This agreement pattern is complicated further by the fact that SM functions as a marker of masculine gender in non-local contexts, while SM functions as a marker of feminine gender in local contexts.

This section introduced some of the linguistic properties that makes gender easier to acquire in some languages (e.g., Spanish) than others (e.g., Welsh). In addition to linguistic properties which affect the acquisition of grammatical gender, other child-external factors also play a role, including the length of exposure to the target system, the quantity of the input as well as the quality of the input. Some of these factors are discussed in the following section, before reviewing the literature exploring the acquisition of Welsh grammatical gender in child learners.

2.6 Child-external factors and the acquisition of grammatical gender

Research has shown that language-external factors play an important role in child bilingual acquisition of grammatical gender (Rodina et al., 2020; Rodina & Westergaard, 2017). Input factors are heavily discussed within the context of bilingual gender acquisition, often referring to the quantity and quality of the input to the target language (Rodina et al., 2020). Input factors such as the role of the home language and the language at school (among others), as well as the length of exposure to the target language, are often considered (Granfeldt, 2018; Rodina & Westergaard, 2017; Unsworth et al., 2014). For instance, Unsworth et al (2014) explored the role of input quantity on the acquisition of grammatical gender in bilingual Dutch-English and Greek-English children. Data were collected from 4-17 year olds who were divided into three groups: 2L1 bilinguals who were exposed to both languages from birth, L2 speakers who were exposed to Dutch or Greek between the ages of 4 and 10, and early successive bilinguals (ESB) who were exposed to Dutch or Greek between the ages of 1 and 4. The participants completed two elicited production tasks (both in the target language) and parents completed an extensive parental questionnaire on the children’s language use and background. The elicited production tasks examined the children's knowledge of grammatical gender marking on definite determiners, by eliciting either determiner-noun (simple) or determiner-adjective-noun (complex) phrases in both languages.

The Dutch results revealed an effect of gender, where participants were better on common nouns than neuter nouns ($p < .001$) and were at ceiling for common nouns, as well as an effect of complexity, with better performance on simple compared to complex DPs ($p < .001$). For neuter nouns, accuracy was significantly higher for the L1 children compared to the ESB children ($p < .01$), and for L1 compared to L2 children ($p < .01$). The percentage of exposure to Dutch, as calculated by the parental questionnaire, was found to be a significant predictor variable for neuter nouns but not common nouns. The Greek results revealed an effect of gender, with better performance on neuter nouns in comparison to masculine and feminine nouns ($p < .01$), as well as an effect of complexity, with better performance on simple compared to complex DPs ($p < .001$). The percentage of exposure to Greek was found to be a significant predictor variable for masculine and feminine nouns but not neuter nouns. Further analysis showed that the percentage of exposure to the target language did not play a greater role in the acquisition of grammatical gender for one language over the other.

This result was somewhat surprising to the authors, given that Greek marks gender more transparently than Dutch. Greek has cues to aid the learner, as gender is consistently marked on the noun and agreeing DP constituents, whereas Dutch is more opaque, as gender marking on the noun is largely absent, and on other DP elements, it is often inconsistent or ambiguous. Therefore, the authors anticipated that the Dutch children would be affected by the amount of input to which they are exposed to, and this would hold for Greek children, but to a lesser extent. The authors concluded that input is clearly necessary to establish the target system and amount of input to which children are exposed is the best predictor for bilingual children's acquisition of grammatical gender in both languages.

Unsworth et al (2014) operationalised the amount of exposure to the target language by considering various sources of input, such as the time spent at school / day care and out of school /day care, the number of hours spent on activities (e.g., sports, clubs, computers, etc) and how much each parent spends time with their child in a week, while Rodina and Westergaard, (2017), in contrast, focused on parental input. They investigated the acquisition of grammatical gender in bilingual

Norwegian-Russian children, focusing on input factors (e.g., parental input). Data were collected from 60 children aged 4-7 living in Norway (with Norwegian as the majority language), of which 20 were Norwegian monolinguals, 20 were Russian monolinguals and 20 were Norwegian-Russian bilinguals. The bilingual children were divided into two groups, 10 of which were exposed to both Norwegian and Russian at home and the other 10 were only exposed to Russian at home. They completed an elicited production task targeting adjective-noun agreement forms in Russian and indefinite determiner-phrases in Norwegian, consisting of 30 items in each language.

The results for the Norwegian indefinite determiner-phrases revealed that the monolingual Norwegian children performed best (*M*=70%), followed by the children who had only Russian at home (*M*=57%) then the children who had both languages at home (*M*=54%). Statistical analyses showed that the difference between the monolingual and bilingual children was significant (*p*=0.00001), but the difference between the two bilingual groups did not differ (*p*=0.84). The results for the Russian adjective-noun phrases revealed that the monolingual Russian children performed best (*M*=96%), followed by the children who had only Russian at home (*M*=91%) then the children who had both languages at home (*M*=59%). Statistical analyses showed that the difference between the monolingual and bilingual children was significant (*p*=0.0004), and the difference between the two bilingual groups was significant (*p*=0.00001).

The authors argued that the amount of parental input plays a crucial role in the minority language (Russian) but not the majority language (Norwegian). Children with both parents speaking Russian at home have a clear advantage, as their accuracy rates are similar to those of Russian monolingual children (91% vs. 96%, respectively). However, in contrast, being exposed to Norwegian at home does not give the children with both languages at home an advantage, as they do not perform better than the children with only Russian at home (57% vs. 54%, respectively). Therefore, the authors concluded that the amount of exposure, especially in the minority language, is crucial.

The role of parental input was also examined by Granfeldt (2018). They investigated input effects on the longitudinal development of French gender acquisition in simultaneous and successive French-Swedish bilingual children. Data were collected from three groups of children who attended the same French school in

Sweden, including monolingual French speaking children (L1) (n=4), bilingual French-Swedish speaking children (simultaneous: 2L1) (one French speaking parent) (n=4) and Swedish-speaking children (successive: cL2) (with two Swedish speaking parents) (n=4). The children's use of indefinite (*un* vs. *une*) and definite articles (*le* vs. *la*) was assessed via free conversations about the past, present, and future, and recordings were taken approximately every 3 to 5 months for up to 3 years, covering an age span from 3;5 to 10 years. Granfeldt distinguished between variables tapping into input quantity and quality, such as the mode of acquisition of French of each parent, proportion of French spoken at home, the number of weeks spent in French speaking countries per year, contact with French speaking people, access to French television and reading habits in French.

The overall results revealed that the L1 children performed at ceiling (*M*=99.6%) as did the 2L1 children (*M*=91%), while the cL2 children did not reach the same levels of accuracy (*M*=75.3%). The difference between the L1 and cL2 children was statistically significant (*p*=.0001) as was the difference between the 2L1 and cL2 children (*p*=.0001). The results revealed that the 2L1 and cL2 children overgeneralized masculine gender, where error rates are higher for feminine nouns than masculine nouns. For both groups, the children used *le* in place of *la*, however, this was not observed for the indefinite article, *un* in place of *une*. The results also showed a positive correlation between input profile scores and gender assignment accuracy in the 2L1 children (medium to strong effect sizes ranging from .376 and .500). However, no clear correlation was observed in the cL2 children, as a child with one of the lowest input scores outperformed a child with the highest input profile score of all cL2 learners. Granfeldt suggested that this difference between the 2L1 and cL2 children is likely because the 2L1 children have a longer exposure time to the target system, where the individual input differences in the cL2 group may become apparent at later stages. Granfeldt concluded that the development of French gender assignment is to a high degree dependent on input conditions.

Sections 2.5 and 2.6 have discussed various linguistic and child-external variables that have been shown to play a role in the acquisition of grammatical gender cross-linguistically. The following section (2.7) turns its focus to the language of interest in this thesis, Welsh. Research has investigated the acquisition of Welsh grammatical gender in child learners, exploring the acquisition of local (e.g., det-noun) and non-local (e.g., 3rd person possessive adjectives) agreement in the production and comprehension of gender. The studies reviewed in the below section also pay attention to the various linguistic and language-external variables that play a role in the acquisition of gender.

2.7 Grammatical gender acquisition in Welsh

This section reviews the research examining the acquisition of grammatical gender in Welsh, paying attention to the linguistic and language-external variables which contribute to its acquisition process.

There is a growing body of literature investigating the acquisition of Welsh from both psycholinguistic and sociolinguistic perspectives, predominantly investigating child speakers from North Wales. Studies have examined how particular aspects of Welsh are acquired, including rhotic consonants and lateral fricatives (Ball, Müller, & Munro, 2001), plural morphology (Binks & Thomas, 2019; Binks, 2017; Thomas, Williams, Jones, Davies, & Binks, 2014), vocabulary (Gathercole, Thomas, & Hughes, 2008) and grammatical gender (e.g., Binks & Thomas, 2019; Thomas, 2001; Thomas & Gathercole, 2005; Thomas & Gathercole, 2007). This section briefly reviews the literature exploring the acquisition of gender (and mutations) in Welsh speaking children.

Studies have focused on the acquisition of mutations in non-gendered contexts. Research has shown that the acquisition of mutations is not complete until after the age of five (Bellin, 1984, 1988), while others have suggested that it is not until the age of nine (Gathercole et al., 2001; Thomas, 2001; Thomas & Gathercole, 2007) or even eleven (Gathercole & Thomas, 2005; Thomas & Gathercole, 2007). For instance, Thomas and Gathercole (2007) found that nine-year-old children were still acquiring SM after prepositions, even though the children were L1 speakers of Welsh (Welsh at home and Welsh at school). The authors did find, however, that despite the slow learning process, the children demonstrated development toward what they considered to be the ‘adult norm’.

The acquisition of Welsh grammatical gender has predominantly been examined through mutations, where evidence of an abstract syntactic representation (grammatical gender) has been tested through the presence of morphology (mutations). For instance, Thomas (2001) tested the use of grammatical gender in 48 Welsh speaking children aged 4.5 to 9 years from North Wales, who received 80% or more of their input at home in Welsh and has Welsh education. Thomas examined the production of gender via Det-Noun sequences. The results revealed that the children were better at producing masculine nouns in their bare form (i.e., no mutation) than SM feminine forms after the definite article (‘y’). Thomas also examined gender marking in non-local contexts in the same Welsh speaking children. The results revealed that they consistently produced SM nouns [m] more frequently than AM nouns [f], often inaccurately producing SM feminine nouns in place of AM feminine nouns. Thomas suggested that these results show that Welsh speaking children begin to produce and mark masculine forms more accurately than feminine forms. They also claimed that these findings are evidence for a piecemeal item-by-item acquisition process, as some forms are acquired sooner than others.

This finding, that performance is better on masculine nouns compared to feminine nouns, is well attested in the literature of gender acquisition cross linguistically. For instance, Pérez-Pereira (1991) observed that monolingual Spanish-speaking children were more likely to assign masculine gender than feminine gender when assigning determiners. Masculine is considered the default gender in Spanish. Granfeldt (2018) also found an overgeneralisation of the masculine definite article *le* in place of *la*, as masculine is the default form in French. Therefore, the results found by Thomas (2001) are in line with other gendered languages (e.g., French & Spanish), particularly given the fact that approximately 69% of Welsh nouns are masculine (Hammond, 2016).

Gathercole et al (2001) also examined non-local gender marking in Welsh speaking children, aged 5 to 9 years, who came from Welsh speaking homes (80%+), English speaking homes (80%+) and bilingual (W: 40%, E:60%) speaking homes in North Wales. Children selected one of two images, one of which was referred to in a non-local gender marked context via masculine or feminine pronoun or possessive forms. The results revealed that the children were more accurate in selecting the feminine nouns than the masculine nouns, which is in contrast with Thomas' (2001) findings. The results also showed that children with only Welsh at home (OWH) outperformed children who had both languages at home (WEH), and only English at home (OEH). The authors claimed that these results are evidence of a variable and protracted course of development when children acquire Welsh gender. They suggested that this difference is because of child external variables, such as the frequency and varying levels of input in Welsh. They postulated that OWH children receive more Welsh and are therefore influenced less by English [-gender], with greater dominance in English, thus leading toward a more semantic gender approach. They also suggested that the course of development is variable because of linguistic variables, such as the complexity and opaqueness of the Welsh gender system.

The suggestion that the opacity of a gender system can affect the acquisition process is in line with previous studies examining the acquisition of the Dutch gender system. The acquisition of Dutch gender in monolingual children is claimed to be a gradual process. Studies have shown that the acquisition of the neuter definite determiner '*het*' causes problems until at least the age of 6 and that children overgeneralise the common '*de*' with neuter nouns, but there is no over-generalisation in the other direction (Van der Velde, 2003, 2004). This is also the case for Dutch-English simultaneous bilingual children, where the acquisition of 'de' with common nouns is typically unproblematic but the neuter gender is problematic, gradually improving by the ages of 9/10 (Cornips & Hulk, 2008; Unsworth, 2008). Bilingual children have also been found to overgeneralise the use of '*de*' with the neuter nouns, instead of '*het*' (Unsworth, 2008). It is suggested that this may be due to the similarity between 'de' and the English determiner '*the*', inferring that at the initial stages of acquisition, there is transfer from English to Dutch (Unsworth, 2008). Similar to Welsh, research on the bilingual acquisition of Dutch gender has found it to be a long-lasting process, whereby some aspects are acquired earlier and with more ease than others.

Evidence suggesting that the acquisition of Welsh gender is a protracted route also comes from Gathercole and Thomas (2005). They investigated the productive command of grammatical gender in local and non-local gender marked contexts, in children from OWH, WEH and OEH language backgrounds, across ages 2.5 to 11, from North Wales. The children described cartoon videos, hoping to elicit D+N+A (local) and *ei*+N (non-local) phrases. For local contexts, the results revealed that the children were more accurate on masculine than feminine nouns. For non-local contexts, the results revealed performance was better with masculine *ei*+(SM) than with feminine *ei*+ (AM) nouns. Both of these findings are in line with Thomas' (2001) study. The results also showed that the OWH group outperformed WEH and OEH children, across local and non-local contexts, which was also observed in Gathercole and colleagues' (2001) study (for non-local marking). Similar to Gathercole et al (2001), Gathercole and Thomas (2005) also claimed that linguistic and child-external variables influence the rate at which different aspects of the gender system are acquired. They argued that some aspects are acquired with more ease than others, serving as evidence of a piecemeal approach. They stated that acquiring the gender and mutation systems simultaneously is challenging, given the complexity of the two systems, but this acquisition process is made easier and is typically less problematic when a child receives optimal exposure to Welsh than when they receive less exposure. They claimed that often, lower amounts of input can lead to incomplete acquisition of certain gender constructs.

Thomas and Gathercole (2007) also examined non-local gender marking via the possessive adjective '*ei'* (his/her) in 35 L1 Welsh speaking children aged 4.5 to 9 years from North Wales, who received education in Welsh and received 80% or more of their input at home in Welsh. Following previous studies (Gathercole & Thomas, 2005; Thomas, 2001), the results revealed that they selected masculine forms more accurately than feminine forms. Thomas and Gathercole stated that this is likely because SM is the most pervasive mutation in Welsh and that AM is very rarely used in relation to gender, accept after *ei* (+F noun). They also noted that SM is used at the expense of AM, where SM is overextended into this particular gender marked context, as children may associate SM with 'feminineness' because of its frequent use in local contexts (following the determiner 'y'). They suggested that such overextensions of SM on feminine forms are indicative of an overlap between the mutation and gender systems. The authors argued that the quicker progression with masculine gender over feminine gender in non-local contexts shows evidence of a lengthy and protracted route of acquisition, with children building their knowledge around individual items in a context-by-context fashion. They suggested that this is likely consequential of the linguistic properties of Welsh gender, such as the opaque mapping between gender and mutations, in addition to the fact that one requires command of mutations to be able to mark gender categories.

In two studies, Sharp (2012) investigated the use of gender in local contexts. One study assessed gender marking via N-Adj sequences in 165 children aged 4 to 9 years, from OWH, WEH and OEH language backgrounds from North Wales. The results revealed that the children were more accurate in producing masculine N+A sequences in their bare form than the feminine N+A sequences in their SM forms. The OEH children did not perform as well as the OWH and WEH children, but this difference was only present for the feminine data. Sharp argued that this difference between the groups of children is due to variables such as the amount of exposure to Welsh at home and in their earliest years, suggesting that reduced levels of exposure can influence gender accuracy.

A second study examined gender marking via gendered numerals with the same children. The children were required to produce the correct gender marked forms of the numerals two, three and four [m/f], when shown either two, three and four of items (see Table 2.3 for gendered numerals). The results revealed that children aged 4 and 5 exclusively produced masculine numerals for all masculine and feminine nouns, with no difference in performance between the home language groups. Seven-year-old children largely overused masculine numerals, however, some produced feminine numerals accurately with feminine nouns. The nine-year-olds showed some increasing awareness of feminine numerals, producing feminine numerals accurately during the task, but this use was still not considered ‘target-like’. Sharp stated that the children have not yet acquired this gender feature in Welsh. They suggested that the overuse of the masculine forms stems from the fact that these are taught as part of the Welsh counting system (un, dau, tri, pedwar, pump, etc) and not the feminine forms. Sharp also argued that children acquire Welsh in a piecemeal fashion, as some forms are learned more quickly than others (i.e., masculine numerals are acquired and accurately produced before feminine forms), which is in line with findings from previous studies (Gathercole et al., 2001; Gathercole & Thomas, 2005; Thomas, 2001; Thomas & Gathercole, 2007).

The most recent data investigating the knowledge of Welsh grammatical gender in Welsh speaking children comes from Binks and Thomas (2019), who tested Welsh-English bilingual children and teenagers, aged 12-13 and 16-17. They were categorised into L1 Welsh speakers (i.e., raised in homes where both parents spoke Welsh), 2L1 simultaneous bilinguals (i.e., raised in homes where both languages were spoken) and L2 Welsh speakers (i.e., raised in homes where both parents spoke English and Welsh was learnt at school at the age of 4), all of which attended Welsh medium schools, located either in North West Wales where 65.4% of the population speaks Welsh, or in South West Wales, where 43.9% of the population speaks Welsh (Census, 2011). Gender in non-local contexts was examined, replicating the task from Gathercole and colleagues’ (2001) study. The results revealed that the L1 group performed best across both age groups, suggesting that having Welsh from birth, at home and in school, is optimal for the development of gender constructs. There were no significant differences in the performance between the 12-13- and 16-17-year-olds, indicating that they were performing at the same level and that they might have reached a plateau on opaque structures.

Moreover, the participants correctly chose *ei* AM with feminine nouns more often than they chose *ei* SM with masculine nouns. This is in line with Gathercole et al (2001), but in contrast to Thomas (2001) and Gathercole and Thomas' (2005). Binks and Thomas noted how this was interesting given that children's knowledge of masculine gender precedes their knowledge of feminine gender in Welsh, thus, suggesting that this may be because there are fewer feminine nouns in Welsh and children therefore pay more attention to these agreement patterns. However, the L1 Welsh speakers' results only reflected 'complete' acquisition for animate referents (not inanimate) at the ages of 16-17. Binks and Thomas claimed that the results do not indicate mastery of the gender system across any bilingual categories for all noun types. They suggested that the acquisition process for all aspects of gender persist beyond the ages of 17 and continue into adulthood. They suggested that factors such as the level of exposure a speaker receives in Welsh at home, in school and the proportion of speakers who use the language in the wider society, all plays a role in the acquisition of Welsh gender in children.

These studies shed light on the acquisition of Welsh gender, in local and non-local contexts. The findings for local contexts show that children are more accurate in producing masculine nouns in their bare form, than SM feminine nouns (Gathercole & Thomas, 2005; Sharp, 2012; Thomas, 2001). However, the results for non-local contexts are mixed, where Thomas (2001) and Thomas and Gathercole (2007) found performance to be better on masculine nouns, while Gathercole et al (2001) and Binks and Thomas (2019) found performance to be better on feminine nouns. Research has also found that various child-external and linguistic variables contribute to the Welsh gender acquisition process. Specifically, the frequency of exposure to Welsh, highlighting the importance of the home language and language of education (Gathercole et al., 2001; Gathercole & Thomas, 2005; Sharp, 2012). They have also highlighted that the acquisition of complex structures is not straightforward particularly given the complexity of the linguistic properties of the system. For instance, the Welsh gender system is considered opaque. It is intertwined with mutations and gender marking in the different contexts (local vs non-local) means that there is no clear form-function mapping between gender and mutations.

Together, the studies have claimed that the acquisition of Welsh gender can take a long time to master, resulting in a protracted, error-marked and item-based process of acquisition, due to its complex system(s) and the natural conditions of reduced input of Welsh due to nature of Welsh bilingualism in Wales (Binks & Thomas, 2019).

In the afore-mentioned studies, the acquisition of Welsh grammatical gender has predominantly been examined through mutations, where evidence of grammatical gender has been tested through mutations. This may be considered somewhat problematic, because even though gender operates in conjunction with mutations, the two are not interchangeable systems, and conflating the two systems may account for the claimed protracted route of acquisition. In order to understand how grammatical gender in Welsh is acquired, produced, and comprehended in children, gender should be tested independently of mutations and in conjunction with mutations, and mutations should be examined independent of gender.

An additional consideration is the type of bilingual children tested in these studies, who come from North Wales, which is a heavily Welsh dominant area. Despite the consideration of child-external variables, such as the frequency of exposure, looking at home and school language, studies have all collected data from children who attend Welsh medium school. For instance, Binks and Thomas (2019) tested a group of participants who they categorised as L2 Welsh teenagers, however, they have and continue to receive their education through the medium of Welsh (from the age of 4). As noted in the introduction of this thesis, within Wales, there are different categories of schools according to Welsh provision. It may have been useful to collect data from children from the different education provisions to paint a clearer and more cohesive picture of how Welsh gender is acquired in Welsh speaking children.

The studies thus far, however, have not collected data from children who identify as Welsh-English bilinguals, such as early sequential bilinguals or late bilinguals. The children in the afore-mentioned studies receive a high degree of input in Welsh, which is in contrast to some of the more typical scenarios in Wales, particularly in South Wales, where a child may attend a Welsh speaking school and have Welsh at home, but the language of the immediate regional society is English. Children in Wales may also attend a Welsh speaking school but only have English at home. There are also children who would be considered as L2 learners of Welsh, with English at home and attend an English medium school, only experiencing

Welsh as a subject (i.e., hearing between 2-3 hours of Welsh a week). These bilingual profiles differ between across Wales. Future work could consider a broader take on bilingual groups that is more representative of the Welsh child population. A final consideration is the fact that adult data has only been obtained to act as a gauge of the maximum target acquisition level for the children. The adult data from these studies are discussed in the following chapter, when considering the production and comprehension of grammatical gender in Welsh bilingual adults.

In summary, these studies have explored the construction of gender in a Welsh-English bilingual child's mind. Results have shown that when a gender system offers no clear indication of noun gender or a clear cue to noun gender, (i.e., determiner is not marked for gender), as well as an intricate relationship between the gender and mutation systems, any late development may be attributed to this complexity (Thomas & Gathercole, 2005). Research has also highlighted the importance of linguistic input levels, such as the language spoken at home and language of education. These variables are all important in developing an understanding of how Welsh gender is acquired and how gender is used by Welsh speaking children.

2.8 Acquisition of gender summary

Bilingual children's language acquisition is affected by certain characteristics of both the linguistic properties of the language and by their language learning experience. Research has shown that different linguistic properties make gender easier to acquire in some languages than others, including the transparency or opaqueness of a gender system, the presence or absence of morphophonological cues, and the locality of gender effects. Research has also highlighted the importance of child-external variables, such that continual exposure to high levels of quality and quantity of input are important for bilingual language acquisition (Daskalaki, Blom, Chondrogianni, & Paradis, 2020; Grüter et al., 2012). Bilingual children naturally receive less input in each of the two languages, experiencing more variability in their input than their monolingual peers (Gathercole, 2002). The rate of gender acquisition can be predicted by the amount of input they are exposed to in their two languages (Unsworth, Chondrogianni, & Skarabela, 2018). However, the rate of grammatical gender acquisition differs across gendered languages for both L1 and bilingual children, showing a non-linear relationship between input and linguistic development (Unsworth et al., 2014). Several factors play a role in how gender is acquired by children in bilingual contexts, such as the quality and quantity of input, as well as the linguistic properties and characteristics of the different gendered systems.

2.9 Chapter Summary

To summarise, this chapter introduced the principal interest of the present thesis – grammatical gender. First, gender was defined and its operations in terms of agreement and assignment were described. Then, the English gender system was briefly outlined, followed by a detailed description of the gender system in Welsh and its operation with the complex morphophonological mutation system. An overview of some of the different linguistic properties that make gendered languages easier to acquire than others was also described, as well as some of the language-external factors that play a role in the acquisition of gender. This was followed with an overview of the current Welsh gender acquisition literature. The following chapter discusses the production and comprehension of grammatical gender in adult L1 and L2 bilingual speakers, as well as two theoretical approaches to L2 ultimate attainment.

Chapter 3

The Production and Comprehension of Grammatical Gender

The previous chapter outlined the linguistics of gender, paying close attention to the complexity and intricacy of the gender and mutation systems in Welsh. Chapter 2 also outlined the various linguistic properties that make gender easier to acquire in some languages than others, as well as some of the language-external factors that play a role in the acquisition of gender. This third chapter turns to adult data, as adults are the focus of this thesis, and the first of two experiments investigates the production and comprehension of grammatical gender in Welsh-English bilingual adults.

Previous research has investigated the nature of second language (L2) acquisition by adult learners and have observed divergent behaviour in several domains of language in comparison to L1-speakers. Theoretical approaches attempt to explain the initial state of L2 acquisition and ultimate attainment in L2 acquisition. However, as the Welsh adult speakers in the present thesis are not L2 adult learners of Welsh, rather, the speakers acquired both languages simultaneously from birth, Welsh [+gender] as their L1 and English second (before or at the age of 11), or English [-gender] as their L1 and Welsh second (before or at the age of 11), initial state theories of L2 acquisition are not discussed.

Theories of ultimate attainment in L2 acquisition are discussed here. The Welsh bilinguals are not L2-like in the sense that they have learned a second language during adulthood, nor are they like L1-monolingual speakers, as they are all bilingual speakers of both Welsh and English. However, the Welsh bilingual speakers have two languages in their minds. It is therefore appropriate to discuss theoretical approaches of ultimate attainment in L2 acquisition, as it may be that their proposals are extendable and relevant to the current Welsh-English bilingual adult data. Various approaches have suggested that errors are attributable to different reasons, and two of these approaches, namely the *Interpretability Hypothesis* (Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007) and the *Missing Surface Inflection Hypothesis* (Prévost & White, 2000), are discussed below. These two approaches were designed to account for variability in L2 adult learners and not adult bilingual speakers who have acquired two languages simultaneously from birth, or even those who acquired the second language before adolescence. However, there exists no theoretical approach or hypothesis that directly accounts for this bilingual population. Therefore, these two accounts are discussed in the following sections, to provide theoretical grounding and discussion for the current population of bilingual speakers investigated in this thesis.

3.1 Overview

The following sections will review the studies which have examined Welsh adult speakers' production and comprehension of grammatical gender in Welsh, who have typically served as a gauge of the maximum target acquisition level for children acquiring Welsh. I will then outline the *Interpretability Hypothesis* (Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007) and the *Missing Surface Inflection Hypothesis* (Prévost & White, 2000), and review studies supporting each approach. Predictions for the Welsh-English bilingual adults are put forward for the theoretical approaches. Finally, I will describe some of the factors that have been found to predict different outcomes in bilingual performance and were selected for investigation in this thesis.

3.2 The use of grammatical gender in Welsh adults

This section reviews the studies which have explored the production and comprehension of gender in Welsh speaking adults. Typically, the adult data have not been the focus of the studies, but rather, the data have been collected as a measure of the ultimate attainment one might expect of children receiving different levels of exposure to Welsh. The studies are presented below according to the production (3.2.1) and the comprehension (3.2.2) of Welsh gender.

3.2.1 The production of grammatical gender in Welsh

One of the earliest studies to collect data from children and adults on a large scale in exploring the production of grammatical gender, comes from Thomas (2001). In one study, Thomas collected data from 30 adult L1-Welsh speakers from Anglesey, Bangor and Ogwen (North Wales). The participants were divided into three age groups, 16-30 (n=8), 31-50 (n=10) and 51+ (n=12). They completed a story telling task which elicited a semi-naturalistic production of grammatical gender, to elicit an equal amount of feminine and masculine nouns in a Det-N sequence. The participants were scored for noun gender mutation accuracy. Participants received 1 point for SM feminine nouns following the determiner and 0 for producing the feminine nouns in AM or bare form, or 1 point for producing masculine nouns in bare form following and 0 for producing masculine nouns in AM or SM forms. The results revealed an effect of gender, but no effect of animacy or age group. The gender effect showed that the adults were errorless in producing masculine forms in their bare forms, while accuracy for feminine nouns was 88%. The feminine noun errors were due to instances of producing the noun in bare form (i.e., no mutation) and two instances of double mutation (assuming the base form was actually a mutated form and mutating again). Thomas suggested that the gender system is well established in adult speech, particularly in local contexts, but the adults were more accurate in masculine (no mutation) than feminine (with SM) noun production.

In a second study, Thomas investigated the production of grammatical gender in non-local gender marked contexts via the possessive adjective ‘ei’ (see Table 2.7). Fifteen adults from first experiment participated (5 from each age group). Participants were shown 88 different inanimate noun prompts, balanced for gender and animacy, and were instructed to produce what they saw in the picture. The pictures showed the different possessor nouns with something ‘on’ it, targeting the elicited noun following ‘ei’. Participants received 1 point for correct use of mutation/no mutation in appropriate contexts. Incorrect or absent mutations received 0 points.

The results showed that accuracy was higher when the possessor was masculine (87.6%) than feminine (29.3%). Analysis showed that accuracy was best for masculine nouns beginning with vowels (no mutation) than SM nouns beginning with *p, t, k, b, d, g, r̥, ɫ* and *m*. Some masculine nouns were produced with no mutation or AM instead of SM. For feminine nouns, accuracy was high for AM and no mutation noun production. Some feminine nouns were SM instead of AM, with some instances of no mutation noun production. Thomas concluded that overall, the adults were more accurate and consistent on masculine forms than feminine forms and suggested that this is likely due to three reasons. First, all masculine nouns with mutatable letters undergo SM, whereas there are fewer letters that AM (9 letters for SM vs 3 letters for AM). Second, the SM is the most pervasive in the Welsh language and is realised most often as it has the most triggering environments, while AM is reported to be use rarely except after 'ei'. Third, SM may be used at the expense of AM and is being overextended into non-local contexts for feminine forms. Thomas questioned whether the results are due to a lack of gender knowledge or rather, they are an indication of adults' lack of ability and consistency with producing mutations in general. This final point serves as one of the fundamental ideas for the present thesis, which investigates whether the production of gender is more robust when gender is independent of mutations, rather than when it is in conjunction with mutations. This consideration is revisited in the summary section of this chapter.

Sharp (2012) investigated adults' knowledge and productive use of Welsh gender in three different experiments. Fifteen L1 Welsh speaking adults participated in all three experiments, which had been raised in homes where both parents spoke Welsh to them from birth onwards (age range 18-50, *M*=25,5). The first experiment assessed local gender marking via SM in Det-N-Adj contexts. The task included 48 trials, which were balanced for gender and animacy. Examples of the elicited spoken sentences by the participants are below:

Masculine:	*Dyma'r*	*ci*	*glas*	
	This PRON is VERB the DET	dog NOUN (m)	blue ADJ	
	'This is the blue dog'			
Feminine:	*Dyma'r*	*gath*	*las*	
	This PRON is VERB the DET	cat+SM NOUN (f)		blue +SM ADJ
	'This is the blue cat			

The results revealed that the adults performed close to ceiling levels (80%+) for both feminine nouns (SM) and masculine nouns (bare form) following the determiner and that there was no effect of noun gender. There was no overall effect of noun animacy, however, further analysis showed that accuracy was higher for feminine animate nouns when compared to feminine inanimate nouns. The absence of an effect of animacy is in line with Thomas' (2001) findings, however, the absence of an effect of gender is in contrast with Thomas (2001), who found that the adults were errorless in producing masculine forms in their bare forms following the determiner, while accuracy for feminine nouns was 88%. Sharp suggested that the

Welsh speaking adults have a productive command of the SM feature in relation to grammatical gender in local gender marked contexts. Sharp also suggested that the Welsh that the children are exposed to by the adults largely includes correct forms of nouns, providing children with the data they need to master the gender system. Sharp noted that the adults' mutation patterns in an experimental setting may not reflect their usage in informal speech, and even in an experimental setting, variation in adults' application of the mutation rule was still evident.

In a second experiment, Sharp explored the production of grammatical gender in non-local contexts, via the pronouns 'hi' (she/it) and 'fo' (he/it). Participants completed an elicitation task (with 48 items balanced for gender and animacy), where they were prompted with a sentence (along with an image) and were required to produce the appropriate pronoun referring to the noun in the picture presented to them. The results revealed an effect of gender, as accuracy was higher for masculine pronouns (79%) than feminine pronouns (55%) in accordance with the antecedent noun. Although the elicited anaphoric reference is different to what was assessed in Thomas' (2001) study ('ei' vs 'hi/fo'), Sharp's results are in line with those found by Thomas, in that performance was better on masculine forms than feminine forms in non-local constructs. There was also an effect of animacy in both genders, with higher accuracy for animate nouns compared to inanimate nouns. Sharp suggested that adults were only able to confidently produce pronouns when semantic gender information was available, asserting that pronouns are not established as a grammatical feature in L1 Welsh speaking adults. Sharp also claimed that the pronouns are treated as a marker of semantic gender, likely because they carry semantic load, and so the distinctions they make at a semantic level are particularly accessible.

The third experiment assessed local gender marking via numerals. The adults were presented with two prompt sentences which referred to two images and they were asked to produce one target sentence referring to the third image presented to them. The task included 48 trials (balanced for numerals and gender). The nouns were presented with a different number of morphosyntactic cues (e.g., no mutation present, one SM or two SM present) as well as the different gendered numeral forms. The results revealed that the morphosyntactic cues were utilised as a cue to elicit the correct sentence (i.e., the more cues present, the more accurate the responses). Sharp suggested that the adults showed an ability to link aspects of the mutation and the grammatical gender systems together. The results revealed that the masculine gendered numeral forms were produced more accurately than feminine gendered numeral forms, where the participants produced masculine gendered numeral forms 63% of the time and feminine forms 37% of the time. Sharp argues that this result reflects the higher percentage of masculine nouns in comparison to feminine nouns in Welsh. Additionally, Sharp attributes this higher percentage to the fact that the masculine forms of the numbers 2, 3 and 4 are used in the Welsh counting system, therefore, they are naturally more frequent in the input. Results also revealed that they performed significantly better on trials including human nouns, in comparison to animal and inanimate nouns. However, Sharp did not report any results concerning the production of the noun accuracy according to SM, AM or no mutation, as the different numeral forms trigger different mutations on the following noun (i.e., no mutation following 3 [f], 4 [f] and 4 [m], SM following 2 [m] + [f], AM following 3 [m]).

Sharp concluded that Welsh speaking adults can clearly distinguish between masculine and feminine forms in local contexts, particularly for Det-N pairs and numeral-noun pairs, suggesting that gender features in local contexts are well established in production. While for non-local contexts, the adults were only able to accurately produce pronouns when semantic gender information was available. This suggests that adults are able to link formal gender features with real world sex distinctions to inform their gender marking decisions; however, it does not suggest that pronouns are well-established as a grammatical feature in adult Welsh speakers.

3.2.2 The comprehension of grammatical gender in Welsh

This section now turns to review the studies exploring the comprehension of gender in Welsh adults. There are three studies which have tested the receptive knowledge of gender in Welsh adults through the means of a comprehension task – Binks (2017), Binks and Thomas (2019) and Thomas and Gathercole (2005). A subset of the data is reported in Binks and Thomas' (2019) study, therefore, Binks' (2017) PhD thesis will be the focus.

Thomas and Gathercole (2005) explored Welsh adult speakers’ receptive abilities with non-local gender marking. Forty-three adults from Gwynedd and Anglesey were divided into two age groups (18-50 years and 51-82 years) and came from two types of homes, where Welsh was spoken over 80% of the time (n=25) or where Welsh and English were spoken between 40% and 60% of the (n=18).

The experiment assessed non-local gender marking via the possessive adjective form '*ei*' and the anaphoric pronouns '(f)*o*' (3rd person singular masculine) or '*hi*' (3rd person singular feminine). Participants were presented with a sentence including two nouns (assisted with a depiction of these two nouns), followed by a second sentence referring to one of the nouns, with altered pictures to convey the meaning of the second sentence. An example is shown below (taken from Binks and Thomas' study, 2019, p. 1028-1029; images taken from Binks, 2017, Appendix B).

First sentence

Dyma'r fwyell $_{\text{FEM}}$ [< bwyell] frown [<brown] a dyma'r gwely $_{\text{MASC}}$ coch.

Here is- the – axe- brown- and –here is- the – bed - red.

"Here's the brown axe and here's the red bed."

Second sentence

Ond mae ei goes/ei choes wedi plygu

But is $_{\text{POSS}}$ leg SM / AM has bent.

"But his/her leg's bent."

There were 36 pairs of sentences, balanced for gender and animacy. The results from an ANOVA did not reveal an effect of age, indicating that the two age groups performed similarly, but revealed effects of gender and animacy. The gender effect showed that accuracy was higher for interpreting the "feminine" possessor in sentences involving feminine referents (85%) than interpreting the "masculine" possessor in sentences involving masculine referents (79.4%). They suggested that this is due to the saliency of AM after 'ei' as signalling feminine possession. The animacy effect showed that accuracy was higher on human nouns (98.8%), than animal (76.6%) and inanimate nouns (71.2%). Further analyses showed that performance was best on human feminine nouns, than animal and inanimate feminine nouns, and performance was better for animal than inanimate nouns, while performance was best for human masculine nouns than animal and inanimate masculine nouns, but there was no difference between animal and inanimate masculine nouns. These results show that the adult speakers' comprehension of 'ei' was high when the antecedent noun was human, however, accuracy decreases when the antecedent noun was an animal or inanimate object. No further results were reported in relation to home languages (WEH and OWH).

In a more recent study, Binks and Thomas (2019) explored the receptive knowledge of grammatical gender in 12-13- and 16-17-year-old Welsh speakers (reported in section 2.7). To provide a basis for comparison for the child and teenager data and to provide a measure of the ultimate attainment, data were collected from 36 Welsh-English bilingual adults and were divided into three groups. Forty-five adults were categorised as L1 Welsh bilinguals, who had been raised in homes where both parents spoke Welsh to them from birth onwards, nine were categorised as 2L1 Welsh bilinguals, who had been raised in homes where both Welsh and English were spoken to them, and 12 were categorised as L2 Welsh bilinguals, who had been raised in English only homes. Participants came from different areas of Wales and all participants had Welsh education.

The materials used were adapted from Gathercole et al (2001), Gathercole and Thomas (2009), Gathercole and Thomas (2005) and Thomas and Gathercole's (2005) studies. In Binks and Thomas' (2019) study, only 30 pairs of sentences were included in the task - 18 pairs included animate nouns (6 human, 6 feminine animals and 6 masculine animals) and 12 included inanimate nouns (6 feminine and 6 masculine). In Binks' (2017) PhD thesis, the results revealed that the adult L1 Welsh bilinguals achieved the highest average score (*M*=85.23, *SD*=2.08), narrowly followed by the L2 Welsh bilinguals (*M*=85.00, *SD*=5.77) and the 2L1 bilinguals (*M*=77.41, *SD*=5.21). Results from a one-way ANOVA revealed no effect of bilingual group. Binks (2017) interpreted this finding to suggest that by adulthood, any differences between bilingual groups disappear.

Binks and Thomas (2019) found the adults to perform significantly better than the 12-13- and 16-17-year-old Welsh speakers. They suggested that the Welsh–English bilingual teenagers may well continue to learn through adulthood or may never acquire the language to the same levels as the adults, perhaps due to their greater exposure to and dominance in English. However, no further results concerning the adult groups were provided, such as gender or animacy. Furthermore, Binks and Thomas noted the imbalance in sample sizes across groups and the fact that the adults who chose to take part in the study may have had a special interest in the Welsh language, as well as other socioeconomic factors that were not controlled for that may have influenced the results. In turn, they suggested that future research should investigate how Welsh–English bilingual adults' lives influence their language proficiency in more detail, in order to establish which factors predict and influence higher performance on measures of morphology.

3.2.3 Welsh data summary

The studies reviewed in section 3.2.1 have shown that adults can distinguish between masculine and feminine forms in local contexts in production, suggesting that the grammatical gender system in Welsh is well-established in adult speakers when gender is marked in local contexts. However, accuracy for gender in non-local contexts is lower, which may be because of the role of SM and AM in the gender system.

The studies reviewed in section 3.2.2 investigated Welsh adult speakers' receptive abilities with non-local gendered contexts via the pronouns and the possessive adjectives 'ei'. The results were mixed. In order to understand how Welsh adults comprehend gender in non-local contexts, future work should investigate what factors predict higher performance on receptive measures of grammatical gender (Binks & Thomas, 2019).

What is clear from the studies reviewed above is that the adult speakers' use of gender and mutations in Welsh is variable. Performance is more consistent when gender is marked in local contexts, however, when gender is marked via mutations in non-local contexts, there is variation in adults' adherence to mutations and thus, the production and comprehension of gender is more inconsistent. Therefore, it then follows that if gender and mutation are not always marked consistently in the input and is transmitted to children under minority language conditions (i.e., where reduced input in highly possible), it is no surprise that children appear to struggle to acquire the system.

There is one idea that I raised earlier that is worth considering at this point in more depth. Thomas (2001) questioned whether the variable use of gender and mutations in Welsh adult speakers is due to a lack of gender knowledge or rather, that is it an indication of adults' lack of ability and consistency with producing mutations in general. The studies thus far have examined the production and comprehension of grammatical gender through mutations, where evidence of grammatical gender has been tested through the presence of mutations. However, this may be considered somewhat problematic, because even though gender operates in conjunction with mutations on the relevant subset of nouns, conflating the two systems may account for the variable use in adults and the claimed protracted route of acquisition in children. In order to better understand how grammatical gender in Welsh is produced and comprehended in adults, gender should be tested independently of mutations, as well as in conjunction with mutations. It is possible that when the two systems are teased apart, gender is more robust when it is separate from mutations. Even if this does not appear to be the case, the findings will nonetheless provide a greater insight into the production and comprehension of gender in Welsh adult speakers.

In addition to this, the results obtained from the Welsh adult speakers have not been discussed in light of any theoretical approaches within the realm of bilingualism, in an attempt to explain the variability at a more detailed level. I suspect that this has not been done as the adult Welsh speakers have not typically been the focus of the study, rather, the adult data helps determine the baseline which children are aiming to achieve (Binks & Thomas, 2019; Sharp, 2012).

It is also possible that this has not been done as there exists no specific theory which attempts to account for variability in bilinguals who have acquired two languages simultaneously from birth, or even those who have acquired the second language before adolescence. Attempting to explain the (possible) variability observed in the Welsh-English bilingual adults' data in the present thesis under theoretical approaches of L2 acquisition will help situate the findings in the broader field of bilingualism and can only develop our understanding of the different types of bilingual speaker profiles that exist in Wales. The following sections review two theoretical approaches which may be relevant and applicable to the Welsh-English bilingual adult data presented in this first experiment.

3.3 Interpretability Hypothesis

The Interpretability Hypothesis (IH) (Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007) is an account that supports a representational deficit view for second language acquisition, specifying that the acquisition of features is governed by their interpretability. It is based on a minimalist distinction between interpretable features, which are accessible at the LF (Logical Form) interface and are those that make an essential contribution to the meaning (e.g., tense), and uninterpretable features, which are only relevant at the PF (Phonetic Form) interface and serve a purely syntactic role without contributing to meaning (e.g., agreement). The IH compares accessibility between L1 and L2 learners, arguing that uninterpretable features will constitute a problematic area for L2 learners. The IH assumes that uninterpretable features are susceptible to critical period effects, in other words, they cannot be acquired after late childhood unless they are instantiated in the L2 learner's L1, however, in contrast, interpretable features remain intact and are always accessible (Chondrogianni, 2024). Therefore, it is assumed that uninterpretable features on nominal elements within the noun phrase (e.g., the determiner) are not acquirable (Chondrogianni, 2024).

Under the assumption that the unavailability of uninterpretable features in the L1 thus renders them unattainable for L2 learners, there are two options in L2 acquisition posited by the IH (Tsimpli & Mastropavlou, 2007). One suggestion is that L2 learners will exhibit transfer of their L1 parametric values even at advanced stages of acquisition. The L2 learners will fail to analyse the input and as a result, show optionality in the use of morphosyntactic elements. The second suggestion is that the L2 properties which cannot be analysed due to their uninterpretable feature specification will be misanalysed by L2 learners, resulting in building non-target features specifically and representations, even at advanced stages of acquisition (Tsimpli, 2003).

3.3.1 Support for the IH

Within in the domain of grammatical gender, support for the IH comes from Tsimpli and Mastropavlou (2007). They investigated the acquisition of determiners and pronominal clitics in ten Greek L2 child learners (L1 Turkish, age range 8-12 years) and six Greek L2 adult learners (L1 Russian & Turkish, age range 27-46 years). Data were collected via an oral production task, examining first-/second object clitics, third person accusative clitics and third person possessive clitics in adults, but only third person accusative and possessive clitics in children, as well as definite and indefinite articles in both groups. In Greek, grammatical gender is marked on definite (singular and plural) and indefinite (singular) articles, showing a tripartite gender distinction: neuter, masculine and feminine. Gender is also marked on third person accusative and possessive clitics (singular), showing neuter, masculine and feminine distinction. Greek gender marking is considered to be systematic, largely unambiguous, and relatively transparent. The authors deduced that based on the IH, the data would show determiner omission likely due to L1 transfer, but with higher omission rates of definite than indefinite articles. Also, there will be higher rates of omission for third person accusative clitics than first-/second person object clitics, as well as higher omission rates for third person accusative clitics compared to third-person possessive clitics, likely due to the role of interpretability in developing L2 grammars.

The adult data for article production revealed that the use of definite articles (49%) is significantly lower that indefinite articles (95%) in obligatory contexts. This is also the case for clitics, with significantly lower production of third person accusative clitics (37%) than first-/second person clitics (83%) and third person possessive clitics (88%). The child data were split into two, grouping children by age (n=5: ‘younger’ 8-9 years & n=5: ‘older’ 11-12 years). The younger children produce the definite article (62%) significantly less than the indefinite article (90%), which is similar to the adult data. However, this pattern was not emulated in the older children, as they use the definite (89%) and indefinite (96%) article similarly. Regarding clitics, the younger group used third-person accusative (27%) and possessive (23%) clitics similarly. In contrast, the older group produced third-person accusative clitics (56%) significantly less than possessive clitics (91%), similar to the adults.

Taken together, the authors argued that the results show support for the interpretability distinction set out by the IH. This is claimed based on the evidence of a priority for indefinite over the definite article in both child and adult data, in addition to the fact that third-person accusative clitics appear to be problematic for both adults and children.

Beyond the domain of grammatical gender, support for the IH comes from Prentza and Tsimpli (2013), who investigated the acquisition of null and postverbal subjects in L2 English intermediate and advanced learners of L1 Greek. Seventy-two learners and 25 L1 English speakers completed a grammaticality judgement task, which assessed the acceptability of null and overt subjects in one part, and a tested the acceptability of postverbal and preverbal subjects in a second part. Greek allows null subjects (i.e., Greek is a pro-drop language), which involve uninterpretable syntactic features, whereas English requires overt subjects, which involve interpretable features. Additionally, Greek allows postverbal subjects under certain conditions (involving interpretable features), whereas English does not, which typically requires preverbal subjects (involving uninterpretable features).

The results showed that both the intermediate and advanced learners were significantly more accurate in judging grammatical than ungrammatical null subjects. The intermediate learners demonstrated partial suppression of null subjects in non-null-subject contexts but continued to produce errors. In contrast, the advanced learners produced fewer errors, showing more accurate usage of null subjects compared to the intermediate learners. When compared to the L1 English speakers, the learners accepted more null subject structures. With regards to postverbal subject structures, the results revealed that L2 learner performance improved with increasing proficiency yet deviated from L1 performance (i.e., English). The results showed a clear developmental pattern, as the advanced learners were significantly more accurate than the intermediate learners for both grammatical and ungrammatical items. However, the learners were not as accurate as the L1 English controls for either the grammatical or ungrammatical items.

The authors took these results to lend support to the IH, because there is evidence suggesting that L1 syntactic options of null and postverbal subjects are active in advanced L2 grammars, thus, validating the claim that uninterpretable features are inaccessible in adult L2 acquisition. They suggested that null and postverbal subjects cause prolonged learnability problems because of their different interpretable feature specification between the Greek / English interlanguage.

While Tsimpli and Mastropavlou (2007) and Prentza and Tsimpli (2013), show support for the IH, the IH is scrutinised by others, who argue that their findings are not compatible with the predictions made by the IH. For instance, Chondrogianni (2008) investigated the acquisition of definite articles and direct object clitics in sequential bilingual children, who were L2 learners of Greek (L1 Turkish) aged 7-12 with different levels of proficiency (low, lower intermediate, upper intermediate, high). Greek definite articles and clitics host uninterpretable features, as they inflect for case, number, and gender. To examine the acquisition of definite articles, learners completed a story-telling task as this controlled for inflectional class and grammatical gender in a naturalistic way. The results revealed that the production of articles was low for the low proficiency learners (26.95%), however, accuracy was significantly higher for the intermediate (lower: 83.79%, upper: 95.98%) and high (99.52%) proficiency learners. This was interpreted to suggest a fast development of the definite article after the initial stage. To examine the acquisition of clitics, learners also completed an elicited production task. The results from the story telling task showed that the low proficiency learners did not produce clitics, however, as proficiency increased, the production of clitics increased (lower: 55.82%, upper: 80.4%, high: 96.78%). This production pattern was also seen in the elicited production task, with increasing proficiency resulting in increased production (2%, 23.45%, 71.20% and 93.33%, respectively).

Taken together, Chondrogianni suggested that it was not possible to explain the findings under the *Interpretability hypothesis*. This is because the IH would predict that articles and clitics would be equally omitted by the learners at initial stages, given their uninterpretable features. However, this was not supported by the data, as the low proficiency learners produce articles (albeit low levels) in the story-telling task. Rather, Chondrogianni suggested that the learners are able to re-assemble the features related to definiteness from their L1 and therefore produce an overt inflectional element before the noun. It is also suggested that the acquisition of complex operations, such as articles and clitics, are modulated by age and input, particularly for clitics, as they need the integration of different domains.

A recent paper by Solaimani, Myles and Lawyer (2023) investigated the acquisition of relative clauses (RCs) in adult L2 learners of English (L1 Persian and L1 French) and found that their findings were incompatible with the *Interpretability Hypothesis*. The L1 English speakers and L2 English learners completed a grammaticality judgement task (GJT). The GJT examined the acceptability of resumptive pronouns and gaps in RCs to see whether there was evidence of learners transferring a resumptive strategy from L1 to L2, given that Persian allows resumptive RCs (object and object-of-preposition RCs, but not subject RCs), while

French and English do not. The results from the GJT revealed that the L1-Persian speakers were more likely to accept resumptive RCs than L1 English and L1 French speakers. L1-Persian speakers' resumption was judged to be more acceptable in direct object and object-of-preposition RCs, than subject RCs, while the L2-French speakers showed no preference between the three. The results revealed that all three groups preferred RCs with gaps over resumptive pronouns and for gapped RCs to be more acceptable with *who* than with *that* (relativizer form). The L1-Persian speakers judged gapped subject RCs to be more acceptable than gapped direct object and object-of-preposition RCs. This was also the case for L1 French speakers, but gapped subject RCs were only judged to be more acceptable for the direct object RCs. The results also showed that both groups of higher proficiency learners were less likely to accept resumption pronouns and that learners who had more immersion experience were also less likely to accept resumptive RCs. However, there was no interaction between working memory capacity (as measured by an RST) and resumption acceptability for either L2 group.

The authors argued the results were incompatible with the *Interpretability Hypothesis*. This is because the IH predicts that the L1-Persian speakers would have different acceptability patterns to the L1-French and L1-English speakers, specifically, that they be more tolerant of resumptive RCs and show non-target-like behaviour for the choice of relativizer. Although the results suggested that L1-Persian speakers were more likely to accept resumptive RCs than L1 English and L1 French speakers, the L1-Persian preferred RCs with gaps over resumptive pronouns. Additionally, the L1-Persian speakers behaviour regarding relativizer form did not differ from the L1 English and L1 French speakers. The authors suggested that this is evidence that the L2 speakers have access to uninterpretable features in the acquisition of English RCs and show capability in acquiring the morphosyntactic phenomena. They also highlighted that the L1-Persian speakers were lower proficiency than the L1-French and L1-English speakers, which likely influenced the higher resumption acceptability rate. This is arguably in line with previous research, suggesting a reduction of processing burden at high proficiency levels when reading complex syntactic structures (e.g., relative clauses). The concluded that given increased proficiency and linguistic exposure, L2 grammars can potentially match L1 grammars in terms of the complexity of the underlying linguistic system.

3.3.2 Predictions of IH for Welsh-English bilingual adults

The previous section outlined the Interpretability Hypothesis (Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007) and reviewed some of the studies supporting and opposing it. Even though the IH was designed for L2 adult learners, it may be possible to extend its basic assumptions to the Welsh adult bilingual data. The Welsh-English speakers are neither L1 monolingual speakers of Welsh nor are they L2 adult learners of Welsh, rather, they sit along the bilingual spectrum / continuum, possibly demonstrating behaviours observed in L1 monolingual data and L2 adult learner data. Therefore, based on the IH, the following predictions are made:

- When gender operates independently of mutations, involving *interpretable* features, this should not present itself as a problematic area for the bilingual speakers. For example:
    - In local gender marking, nouns following the gender marked forms for the numeral *four* ‘pedwar’ [m] and ‘pedair’ [f], are produced in their bare form (i.e., are not mutated). Accuracy should be high given the interpretability of the feature, where the absence of a mutation contributes directly to the interpretation of the syntactic structure.
- When gender operates in conjunction with mutations, also involving *interpretable* features, this should not be the locus of difficulty for the bilingual speakers. For example:
    - In local gender marking, nouns following the gender marked forms for the numeral *two* ‘dau’ [m] and ‘dwy’ [f] are soft-mutated. Accuracy should be high given the interpretability of the feature because it is a consistent and predictable grammatical rule in Welsh.
    - In local gender marking, nouns following the gender marked masculine form of the numeral *three* ‘tri’ [m] triggers an aspirate mutation while nouns following the feminine form ‘tair’ [f] are not mutated. Despite triggering different instances of mutation on following nouns, this should not present itself as an area of difficulty due to the specific numeral values.

- When gender is encoded through mutations, involving *uninterpretable* features, this, however, may constitute a problematic area for the Welsh bilingual speakers. For example:
    - In non-local gender marking, third-person singular proclitic '*ei*' (his / her) is interpretable, but the mutations on the following nouns reflect an uninterpretable feature as they must agree with the gender of the antecedent noun, and this agreement is marked by SM for masculine gender and AM for feminine gender.

It is possible that the uninterpretable features of gender when they are encoded through mutations are vulnerable, while those bearing interpretable features are not. Furthermore, the IH does not explicitly differentiate between local and non-local gender marking, rather, it focuses on whether the grammatical features, such as gender, are interpretable or uninterpretable. Additionally, the IH does not explicitly differentiate between comprehension and production performance, assuming symmetrical performance between production and comprehension (Chondrogianni, 2024). Therefore, evidence of asymmetrical performance between the two could not be directly explained by the IH.

3.4 Missing Surface Inflection Hypothesis

The Missing Surface Inflection Hypothesis proposed by Prévost and White (2000) (henceforth MSIH) assumes that abstract categories and features with complete functional projections are present in L2 grammar, but failures to produce consistent inflectional morphology stem from problems at the morpho-syntax interface. The MSIH claims that the difficulty lies in mapping the abstract features to their appropriate forms on surface morphology under real-time processing pressure, in particular in production (Prévost & White, 2000). In these situations, L2 learners often rely on default forms or other less specified forms because they fall short of mapping the appropriate target lexical form into the given contexts (Prévost & White, 2000). With the proposal that the errors are specifically related to production, a second prediction of the MSIH is the dissociation between production and comprehension (Prévost & White, 2000). Performance is arguably poorer on production measures than comprehension measures as a by-product of communication pressure, because production happens during real-time, requiring fast, spontaneous and automatic processing, while comprehension tasks do not typically involve such pressures (Alarcón, 2011; Grüter et al., 2012; Hopp, 2013, 2016).

Prévost and White (2000) first proposed the MSIH in a study which examined the use of inflectional morphology in the spontaneous production of two L2 learners of French (L1: Arabic) and two L2 learners of German (L1: Spanish and Portuguese), in terms of the optionality of finite and non-finite verbs in non-finite contexts, as well as verbal agreement. The study was longitudinal in that the learners were recorded approximately once a month for two to three years. Prévost and White predicted that finite verbs should not appear in non-finite contexts, and the results supported this. The results showed that the learners showed high accuracy infiniteness placement regarding the use of finite and non-finite verbs, leading Prévost and White to claim that the L2 learners appropriately distributed finite forms in non-finite contexts. The results concerning subject-verb verbal agreement showed that the learners were largely accurate in matching inflected verbs appropriately to the given subjects. The French learners were correct in over 94.5% of all contexts, while the L2 Germans still showed high accuracy of subject-verb agreement in over 87.8% of the contexts.

Based on the results, Prévost and White argued that there is no evidence to support a deficit view, as the learners did not show high variability in finite verb forms and subject-verb agreement. Instead, they claimed that that optionality is syntactically constrained and any of the observed errors during spoken production do not reflect the impairment or deficiency of the underlying syntactic representation, but rather, they are attributable to a failure in selecting the appropriate morphology at the moment of production. They also argued that optionality in L2 does not lie in the domain of syntax but in its interface with morphology, where syntactic nodes are fully specified in the learners, but the lexical entries may not be. Therefore, the problem is in accessing the relevant lexical item by which inflection is realized and this arises specifically during oral production (Prévost & White, 2000).

3.4.1 Support for the MSIH

Since its' proposal, studies examining syntactic phenomena, such as grammatical gender, have shown support for the MSIH (e.g., Alarcón, 2011; Black & Tararova, 2020; Montrul, 2011; Unsworth, 2013). As predicted by the MSIH, the functional gender feature and gender agreement is acquirable in L2, as the feature gender is represented in the speakers' grammar at an abstract syntactic level (Grüter et al., 2012). However, any gender errors that occur during production are due to a production specific performance problem. It is claimed that speakers fall short of mapping the appropriate gender forms to features in the given grammatical context and that this reflects a problem with spelling gender-marked forms out in production, and not due to a representational deficit or impairment (Black & Tararova, 2020; Hopp, 2013, 2016b).

Alarcón (2011) showed support for the MSIH. A comprehension task and a production task assessing Det-N-Adj gender agreement in Spanish were completed by Spanish heritage speakers (bilingual with English), advanced L2 Spanish learners (L1 English) and L1 Spanish monolinguals. The results from the comprehension task showed that the three groups performed similarly, with the L1 speakers scoring 99.6% accuracy, the heritage speakers scoring 98% and the L2 learners scoring 96.9%. Statistical analysis revealed that there were no differences between the heritage speakers and the L2 learners, however, no comparisons were made between the L1 and heritage/L2 speakers. The results from the production task, however, varied more in comparison to the comprehension task. The results showed that the L1 speakers scored 100% accuracy, while the heritage speakers scored 97.1% and the L2 learners scored 85.6%. Statistical analysis revealed that the difference between the heritage speakers and the L2 learners was statistically significant, however, no comparisons were made between the L1 and heritage/L2 speakers. Based on the fact that the heritage learners were equally accurate in gender comprehension and production and that the L2 learners were significantly more accurate with gender comprehension than production, Alarcón argued that these results show support for the MSIH. Alarcón claimed that the differences between L2 comprehension and production is due to a mapping problem, in that the performance errors stem from difficulties in accessing the abstract features and mapping them onto their surface forms. Alarcón also concluded that the heritage and L2 learners have gender features in their underlying grammars despite their difficulties in mapping the abstract gender feature to its appropriate form.

Montrul (2011) also examined Spanish Det-N-Adj gender agreement in a production task and two written tasks in Spanish heritage speakers (bilingual with English), L2 Spanish learners (L1 English) and L1 Spanish monolinguals. The results from the three tasks showed that the L1 monolinguals barely made any errors across the three tasks: 99.9% accuracy on the production task, 98.8% accuracy on the written picture identification task and 97.5% accuracy on the written morphology recognition task. However, the heritage speakers and L2 learners were less accurate in comparison, scoring 85.5% and 74.9% on the production task, 80.2% and 89.7% on the written picture identification task, and scoring 79.9% and 89.3% on the written morphology recognition task (respectively). Statistical analyses revealed that in the two written tasks, the L2 learners were more accurate than the heritage speakers, while in the oral task, the heritage speakers were more accurate than the L2 learners. Montrul suggested that the L2 learners have acquired the abstract gender feature but have difficulty in realizing gender morphology in production, due to a production performance problem, which is consistent with the MSIH. However, Montrul claimed that this explanation is not extendable to the heritage speakers, given that they performed best on the oral production task. The fact that the MSIH cannot be used to account for the heritage speaker results is unsurprising, as the MSIH was designed to explain variability in L2 adult learner results and the fact that heritage speakers do not always have schooling / written exposure in their heritage language, therefore, they are typically better at spoken than written tasks. Since its proposal, there have been attempts to extend it to different bilingual profiles, such as heritage speakers (e.g., Alarcón, 2011). In some instances, the MSIH can be applied to these different bilingual populations, yet this is not always the case.

Support for the MSIH also comes from Unsworth (2013) who investigated the production and judgement of Det-N phrases in Dutch-English bilingual children, who had been raised with both languages from birth. The results from two elicited production tasks showed that they were significantly more accurate in producing the determiner *de* for common nouns than *het* for neuter nouns. The results from the GJT showed that they were highly accurate in producing the appropriate determiners for both common and neuter nouns. However, in the GJT, accuracy was higher for the producing *het* with neuter nouns than *de* for common nouns, which is in contrast to the production task. Unsworth noted that these results are consistent with the MSIH, in that performance is better overall on the GJT than the production tasks. Unsworth suggested that the children's failure to consistently produce the target definite forms in production reflects a production-specific performance problem, rather than a failure to acquire those grammatical features and rules. It is possible that during production, the children resorted to the default common form *de* where the more specified form *het* is required, as predicted by the MSIH, because the bilingual children fall short of specifying certain nouns with the target gender feature.

In a more recent study, Black and Tararova (2020) examined Spanish gender agreement between Det-N and N-Adj in Spanish L2 learners (L1 English) using a production task and a written recognition task. The results showed that the learners were more accurate with masculine nouns than with feminine nouns, showing that the masculine form is being treated as a default form and is overextended to feminine nouns. The results also revealed that the L2ers performed better on the written recognition task (84%) than the oral production task (64%). The authors argue that the difference in performance between the two tasks can be accounted for under the MSIH. They claimed that the L2 learners have the feature gender represented in their L2 grammars at an abstract syntactic level, but gender errors occur due to production or assembly problem, or a computational difficulty during unplanned oral production.

Turning to a study that has not argued for its results under the MSIH, but whose findings may be explained by the MSIH, is a study by Fhlannchadha and Hickey (2021). They examined the productive use of grammatical gender in proficient adult Irish speakers, targeting specific initial mutations[8]. The results from three expressive measures showed variability in gender marking accuracy (30 – 90%). The authors accounted for this variability in terms of age and the influence of English as accuracy scores were higher in older and more dominant adult Irish speakers. However, it could be suggested that this variability in gender marking accuracy, using initial mutations, could be accounted for under the MSIH. Among factors such as age and dominance, it is also possible to interpret the morphosyntactic variability as a 'mapping problem', where the Irish adults have difficulty mapping gender forms to features in real-time processing (Hopp, 2016a). Since agreement morphology between gender and mutation requires highly complex activities (Adila & Ma'mun,

[8] Nouns in Irish are marked for gender using initial mutations depending on the morphosyntactic context and these changes are governed by definite articles and possessives (Fhlannchadha & Hickey, 2017). This is for a sub-set of letters, similar to Welsh.

2020), it could be suggested that the adults have trouble spelling them out correctly when involved with the mutation system during production. They may even have difficulty expressing appropriate agreement marking even if they know the gender of a lexical item (Spinner & Juffs, 2008). It is also possible that some of the younger adults have been exposed to variability in the input in the community, which may have given rise to variability in mapping from representations to forms in the speakers' grammars (Sarko, 2009). This evidence of variability in gender marking can likely be attributed to the complexity and opacity of the gender system with its involvement with mutations, among other factors such as the instability of the Irish community and Irish's minority language status in its country. This is similar to Welsh, in that these various factors influence the speakers' use of the complex gender system in Welsh.

3.4.2 Predictions of MSIH for the Welsh-English bilingual adults

This section outlined the *Missing Surface Inflection Hypothesis* proposed by Prévost and White (2000) and reviewed some studies supporting it. Even though the MSIH was designed for L2 adult learners, it may be possible to extend it to the Welsh adult bilinguals, given that previous research has applied the MSIH to simultaneous bilingual children (e.g., Unsworth, 2013) and adult heritage speakers (e.g., Alarcón, 2011). The Welsh-English speakers are neither L1 monolingual speakers of Welsh nor are they L2 adult learners of Welsh – they sit along the bilingual spectrum / continuum, possibly demonstrating behaviours observed in L1 monolingual data (i.e., ceiling effects) and L2 adult learner data (i.e., large amounts of variability). Therefore, based on previous findings and the MSIH, the following predictions are made:

- There will be evidence of a dissociation between syntax (grammatical gender) and morphology (mutations), where accuracy will be higher when gender is examined independent of mutations than when it is in conjunction with mutations.
- When the Welsh bilingual speakers need to produce a soft mutated form, they may produce a default form (i.e., no mutation) in place of the more specified form (i.e., soft mutation).

- When the Welsh bilingual speakers need to produce a mutated form, they may produce a less specified form (i.e., soft mutation) in place of a more specified form (i.e., aspirate mutation).
- Evidence of variable behaviour may predominantly surface in production given that the MSIH predicts problems spelling out the relevant grammatical features in production, and it does not make explicit predictions for comprehension.

The predictions made under the MSIH, initially, will not distinguish between local and non-local gender marking. However, they may be extended to discuss locality, given that this could be a new testing ground for the MSIH and the function of distance in Welsh gender agreement.

3.5 Individual differences variables

Having reviewed the Welsh adult speaker data and three theoretical approaches of L2 acquisition which may be applicable to a bilingual population, I now turn to briefly describe some of the factors that have been found to predict different outcomes in bilingual performance. The previous sections have alluded to some of these individual factors, such as age of onset to the L2, input quantity/quality, linguistic proficiency, or cognitive resources. Previous Welsh data focused on home language (during childhood) and the language of education. Notably, Binks and Thomas (2019) suggested that future research should investigate how Welsh–English bilingual adults' lives influence their language proficiency in more detail in order to establish what factors predict and influence higher performance on measures of morphology. This suggestion is taken forward in this thesis, where a selection of these factors is tested in this first experiment, namely, language dominance and linguistic proficiency.

3.5.1 Language dominance

Bilinguals are typically a heterogeneous group considering differences in variables such as proficiency, exposure and use, age of onset and attitudes towards the respective languages, among others. Different combinations of these variables generally give rise to distinct yet not differentiated bilingual profiles, which are the result of varied bilingual experiences (Luk & Bialystok, 2013). Two important components of an individual's language profile and bilingual competence are proficiency and dominance (Montrul, 2016; Vicente et al., 2019). Dominance is related to proficiency in several ways, however, they are not uniquely equitable (Birdsong, 2018). Dominance has previously been defined as the relative weight and proficiency of the bilingual person's two languages (Gathercole & Thomas, 2009; Montrul, 2016). It is understood in terms of dimensions and relativistic terms, not absolute, because it considers the two languages of the bilingual speaker. Dominance considers biographical variables, including age of acquisition, birthplace, languages of the environment and current residency, the type and amount of input in the two languages, as well as a bilinguals' estimate of their own competence (Gertken, Amengual, & Birdsong, 2014; Montrul, 2016; Oppenheim, Griffin, Peña, & Bedore, 2020). Research has found it important to consider bilinguals' current and cumulative language experience and use, given that gradient changes in the speakers linguistic environment results in shifts between the two (or more) languages (Oppenheim et al., 2020).

It is important to note that different conceptualisations of language dominance in bilinguals has given rise to different measures and its subsequent inclusion as a predictor in studies testing its potential effects (Silva-Corvalán & Treffers-Daller, 2016). Language dominance has been used as a categorical predictor that determines which of the languages of the linguistic repertoire of a bilingual is their dominant language (Treffers-Daller, 2016, 2019). However, previous research has found it to be problematic when dominance is treated as a categorical variable because dominance, arguably is inherently gradient, not categorical (Birdsong, 2016; Solís-Barroso & Stefanich, 2019). Solís-Barroso and Stefanich (2019) claimed that bilinguals would not be consistently placed into the same dominance group depending on which assessment is use and that this would vary depending on when the participant is assessed.

An alternative and more favourable approach is to treat bilingualism as a spectrum and to therefore consider dominance as a continuous or gradient multi-faceted construct that results from a composite score of different relevant measures (Birdsong, Gertken, & Amengual, 2012; Birdsong, 2016; Rothman et al., 2023;

Treffers-Daller, 2016). As a result, several questionnaires that include different dimensions within the wider construct of language dominance have been designed and broadly used: e.g., the *Bilingual Language Experience Calculator* (BiLEC) (Unsworth, 2013), the *Bilingual Dominance Scale* (BDS) (Dunn & Fox Tree, 2009), the *Language Experience and Proficiency Questionnaire* (LEAP-Q) (Marian, Blumenfeld, & Kaushanskaya, 2007), the *Language and Social Background Questionnaire* (LSBQ) (Luk & Bialystok, 2013), and the *Bilingual Language Profile* (BLP) (Birdsong et al., 2012; Gertken, Amengual, & Birdsong, 2014). Each of these are considered useful tools when collecting relevant data (Rothman et al., 2023).

In order to capture the multiple dimensions of language dominance as a construct, dominance is operationalized and analysed as a continuous variable in this experiment (Birdsong, 2018) and the BLP is considered a suitable tool to investigate the possible role of language dominance in the production and comprehension of Welsh gender. This is because the BLP allows for a more fine-grained characterisation of a bilingual's language dominance, containing detailed questions on dimensions and domains related to dominance, than other questionnaires specific to measuring dominance (e.g., BDP, Bonvin, Brugger, & Berthele, 2021) (Rothman et al., 2023). The BLP has been widely used to account for variability in different bilingual populations, e.g., simultaneous (Amengual, 2016a, 2016b; Bonvin et al., 2021; Perpiñán, 2017) and sequential bilinguals (Black, Joanisse, & Rafat, 2020; Bonvin et al., 2021; Garraffa, Obregon, & Sorace, 2017; Olson, 2017). Each of these studies crucially suggest that using a measure that conceptualizes language dominance to be a gradient variable and not a categorical one, has the potential to explain (at least) some of the variability found within bilingualism (Rothman et al., 2023). The use of the BLP as a measure of language dominance is revisited in section 4.1.4 of this thesis.

Given that language dominance, understood as a multi-faceted construct including has been found to modulate linguistic outcomes in a number of different phenomena and different bilingual populations, it is predicted that dominance will play a role in both the comprehension and the production of grammatical gender in the Welsh-English bilinguals.

3.5.2 Linguistic proficiency

Closely related to dominance is an individual's linguistic proficiency (Montrul, 2016; Vicente et al., 2019). Although related in several ways, linguistic proficiency, which is globally defined as linguistic ability and fluency in a language (Montrul, 2016), is recognised as a component of dominance, rather than an interchangeable construct (Treffers-Daller & Silva-Corvalán, 2016).

Previous research has found bilingual speakers' linguistic proficiency levels to be a predictive factor in the production, comprehension and processing of grammatical gender cross-linguistically (Foucart & Frenck-Mestre, 2011; Silvina Montrul, Foote, & Perpiñán, 2008; Sagarra & Herschensohn, 2010). For example, Montrul et al (2008)) found more target-like production of Spanish gender agreement in oral production among heritage speakers of Spanish than the L2 Spanish learners. However, the L2 learners outperformed the heritage speakers in written comprehension. The authors explained that these divergences can be accounted for in terms of language proficiency. Sagarra and Herschensohn (2010) found that L1-like processing of gender can be attained, but that it depends on proficiency, the results from a SPR task revealed that intermediate L2 learners and L1 speakers were sensitive to the gender agreement violations, but the L2 beginners were not. The authors suggested that L2 learners with a certain proficiency level can develop processing patterns that are qualitatively similar to those of L1 speakers. Similar results were obtained by Foucart and Frenck-Mestre (2011) who examined gender agreement violations in French L1 speakers and advanced L2 French learners (L1 German). The results revealed that both groups evidenced a P600 response, suggesting that L2syntactic processing is affected by proficiency levels in the target language. They suggested that high-proficiency L2 learners who receive enough exposure to their L2 can process gender in a similar way to L1 speakers.

Researchers have employed different tools to assess linguistic proficiency in their subjects, including standardized language tests (e.g., TOEFL) and cloze tests (i.e., fill in the gaps). The former provides the researcher with a standardized proficiency score. However, standardized tests are often designed for specific decision-making purposes, which means they may not be appropriate for a given study (Tremblay & Garrison, 2010). Additionally, not all languages have standardized tests available to assess the linguistic proficiency levels of subjects, including Welsh. Cloze tests have been a popular tool to measure linguistic proficiency for several years (e.g., Bachman, 1985; Jonz, 1976; Oller, 1972).

Arguably, they are a good proficiency assessment measure by testing standards (e.g., reliability measured by Cronbach's alpha) (Bachman, 1985; Brown, 1980; Tremblay & Garrison, 2010). Research has shown that different cloze tests are internally consistent, which do not vary substantially across cloze test with varying methods (e.g., Bachman, 1985; Brown, 1980; Chapelle & Abraham, 1990).

Following previous research, the cloze test method was selected as a general proficiency test to measure Welsh and English linguistic proficiency levels in the Welsh-English bilingual adults because they are an independent measure of proficiency (Ayoun, 2019), they are often sufficiently integrative (e.g., it not only assesses morphosyntactic competence, but also lexical and discourse competence) and scores can discriminate between the bilingual participants of interest (Tremblay & Garrison, 2010). As there is no reliable proficiency assessment tool available in Welsh and with the need to assess proficiency with a test, the cloze test method suited the needs of the current research outcomes (Tremblay, 2011; Tremblay & Garrison, 2010). The use of cloze tests as a measure of linguistic proficiency is revisited in section 4.1.8 of this thesis.

Given that linguistic proficiency has been found to modulate linguistic outcomes in a number of different phenomena and different bilingual populations, it is predicted that both Welsh and English proficiency will play a role in both the comprehension and the production of grammatical gender in the Welsh-English bilinguals.

3.6 Chapter summary

This chapter has reviewed studies examining the production and comprehension of gender in Welsh speaking adults, as well as two theoretical approaches which may be extendable and relevant to the Welsh-English bilingual adult data in the following chapter. Chapter 4 presents the first experiment of two, which will:

- tease apart the gender and mutation systems, addressing the point put forward by Thomas (2001) who questioned whether the variable use of gender and mutations in Welsh adult speakers is due to a lack of gender knowledge or rather, is it an indication of adults' lack of ability and consistency with producing mutations in general

- examine the production and comprehension of grammatical gender in a more diverse Welsh adult population than those previously tested
- discuss the results in light of the theoretical approaches presented in this chapter

Welsh speakers vary in regard to their use and input of the Welsh language, including the different possible combinations of home and school languages when growing up, the use of Welsh in professional environments and the language of the immediate community. These factors differ between individuals in any case, however, some of these factors may have a more prominent influence when considering the differences between Welsh adult speakers from South Wales and North Wales. For instance, Welsh is the dominant language of schools and the wider community in several regions of North-West Wales, whereas English is dominant and is the norm in South Wales (e.g., Swansea, Cardiff). In light of this, the first experiment will also:

- consider language dominance and linguistic proficiency, in order to investigate whether these individual difference variables predict / influence higher performance on measures of morphology in Welsh-English bilingual adults

Chapter 4 presents the first experiment, detailing the methodology, results, and a discussion of the findings in light of the literature reviewed in this third chapter.

Chapter 4
Experiment 1: The Comprehension and Production of Welsh Grammatical Gender

This chapter presents the first of two experiments. It is divided into three main sections: the methodology used to empirically test the comprehension and production of Welsh grammatical gender (see section 4.1), the results of these tasks, including the roles of proficiency and dominance (see section 4.2) and a discussion of the findings (see section 4.3).

4.1 Methodology: Overview

Section 4.1 outlines the methodology used to investigate the comprehension and production of grammatical gender in Welsh-English bilingual adults, in addition to the possible roles of language dominance and linguistic proficiency in the comprehension and production of Welsh gender. Details are given on the research questions and hypotheses, participants, procedure, task details and analyses.

4.1.1 Research Questions and Hypotheses

Considering the theoretical accounts and empirical studies examining the comprehension and production of grammatical gender presented in Chapter 3, the following research questions and hypotheses were formulated. Moreover, dominance and proficiency are considered to provide novel insights that may potentially explain the comprehension and production findings. Therefore, the following research questions guide this experiment:

1) To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in comprehension when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?
    a) What are the roles of dominance and proficiency in Welsh-English adult bilinguals' comprehension of Welsh grammatical gender, when gender is

independent of mutations and when gender is encoded through mutations, in non-local contexts?

2) To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in production when gender is independent of mutations and in conjunction with mutations, in local contexts?
   a) What are the roles of dominance and proficiency in Welsh-English adult bilinguals' production of Welsh grammatical gender, when gender is independent of mutations and in conjunction with mutations, in local contexts?

The two theoretical approaches outlined in Chapter 3, the *Interpretability Hypothesis* and the *Missing Surface Inflection Hypothesis*, make different predictions for Welsh-English adult bilinguals. These predictions are summarized in the table below.

Table 4.1

Summary of Predictions

Predictions		
Comprehension	**IH**	**MSIH**
Show accurate use of gender in *comprehension,* in non-local contexts, when gender operates independently of mutations via anaphoric pronouns 'fe' + 'hi' (gender - mutations)	✓	?
Show accurate use of gender in *comprehension,* in non-local contexts, when gender is encoded through mutations via the possessive pronouns 'ei' + SM [m] / AM [f] (gender + mutations)	X	?
Production (gender only)	**IH**	**MSIH**
Show accurate use of gender in production, when gender operates independently of mutations and is marked via gendered numeral forms: 'dau' [m] / 'dwy' [f] (2), 'tri' [m] / 'tair' [f] (3), 'pedwar' [m] / 'pedair' [f] (4)	✓	✓
Production (gender +/- mutation)	**IH**	**MSIH**
Show accurate use of gender in *production,* in local contexts, when gender operates independently of mutations via 'pedwar' / 'pedair' + no mutation on nouns (gender - mutations)	✓	✓
Show accurate use of gender in *production,* in local contexts, when gender operates in conjunction with mutations via 'dau' / 'dwy' + SM on nouns (gender + mutations)	✓	? / X
Show accurate use of gender in *production,* in local contexts, when gender is conjunction with mutations via 'tri' + AM / 'tair' + no mutation on nouns (gender +/- mutations)	? / ✓	X

I consider whether the predictions can be used to explain the patterns emerging from the Welsh-English bilingual adult data, when gender is independent of mutations (e.g., through 4 & anaphoric pronouns), when gender is in conjunction with mutations (e.g., through 2 & 3) and when gender is encoded through mutations (e.g., through possessive pronouns). Therefore, these predictions will be empirically tested with Welsh-English bilingual adult speakers in the current experiment. These bilinguals are described in the following section.

4.1.2 Participants

Forty self-reported Welsh-English bilingual adults participated in the experiment - ten males, 29 females and an individual who identified as non-binary (age *M*=34, range 19-64). The group are highly educated with a variety of occupations; six participants with PhDs, 14 with master's degrees, 13 with undergraduate degrees, while the remaining participants all had A-level qualifications and were studying at undergraduate degree level at the time of testing.

The Welsh-English adult bilinguals in this experiment are considered along a continuum of language dominance (Amengual, 2015), and are not categorized according to age of acquisition. However, considering bilinguals are often a heterogenous group that varies greatly, even if they acquire both languages early in life, the adult participants' bilingual type is noted here for interest.

The participants generally fit into three categories. These are simultaneous bilinguals who acquired both Welsh and English from birth (n=11), early sequential bilinguals who acquired their second language between ages three and six (n=23) and later-bilinguals who acquired their second language after the age of seven and before or at the age of 11 (n=6). Of the 23 early sequential bilinguals, 16 acquired Welsh first then English, while the other seven acquired English first then Welsh. Of the six later bilinguals, three acquired English first and began learning Welsh at age 11, while the other three later bilinguals acquired Welsh from birth and began learning English in an instructional setting at seven.

In a number of studies by Thomas, Gathercole and colleagues (outlined in Chapter 2), Welsh speaking child participants are grouped in the following way: Welsh at home and Welsh at school (L1 Welsh), English at home and Welsh at school (L2 Welsh), and both Welsh and English at home and Welsh at school (simultaneous). If the 40 Welsh-English bilingual adults in the current experiment were categorised according to these three adult groups in Binks and Thomas' study (i.e., Welsh medium education until 18 years and either Welsh at home, both Welsh and English at home, or English at home until 18 years), the participants would be categorised in the following way: 21 L1 participants, two simultaneous participants and two L2 participants (total=26). This leaves 14 participants unaccounted for according to these group categorisations. The Welsh speakers who are unaccounted for, are of the following bilingual combinations and profiles:

- Welsh at home, Welsh at primary education, English at secondary education (Welsh taught as a subject) (n=1)
- Welsh at home, Welsh at primary education, bilingual (dual track) secondary education (some subjects in English and some in Welsh) (n=2)
- English at home, English at primary education (Welsh taught as a subject) and Welsh at secondary education (n=2)
- English at home, bilingual (dual track) primary and secondary education (n=1)
- English at home, English at primary and secondary education (Welsh taught as a subject) (n=5)
- Both Welsh and English at home, bilingual (dual track) primary and secondary education (n=1)
- Both Welsh and English at home, English primary and secondary education (Welsh taught as a subject) (n=1)
- Both Welsh and English at home, English primary education (Welsh taught as a subject) and Welsh at secondary school (n=1)

As evidenced above, it is not possible to categorise the Welsh-English adult bilingual participants in the same way as Binks and Thomas (2019). The different home and school language combinations are not unusual for Welsh speakers in Wales, particularly speakers from the South Wales, with the main language of education changing between primary and secondary school, or for Welsh to be taught as a subject in school (not as the main medium of communication). Therefore, the Welsh-English bilingual adults in this study are not divided into these three groups, and are considered along a continuum of language dominance (Amengual, 2015).

The bilinguals in this experiment come from more geographically diverse areas than those in previous studies. This is because there are few areas across the south of the Wales where Welsh speakers speak Welsh at home, had a Welsh education and where Welsh is the dominant language in the immediate community. Notably, the Welsh speaking adults in Binks and Thomas' study come from different areas of Wales, however, it is unclear which areas these were. The child participants were recruited from two areas, a region of the northwest of Wales, where 65.4% of the population speaks Welsh, or in a region of southwest Wales, where 43.9% of the population speaks Welsh (Office of National Statistics, 2011). The Welsh bilingual adults in the present experiment are from the following areas. Information regarding the percentage of speakers in that local authority is provided in Table 4.2 (Office of National Statistics, 2011).

Table 4.2

Participants by Local Authority

Local authority	Percentage of Welsh speakers in that local authority	Number of participants living at the time of testing in that local authority
Swansea	11.2%	n=16
Cardiff	12.2%	n=9
Carmarthenshire	39.9%	n=5
Gwynedd	64.4%	n=3
Isle of Anglesey	55.8%	n=1
Conwy	25.9%	n=1
Pembrokeshire	17.2%	n=1
Rhondda Cynon Taf	12.4%	n=1
Caerphilly	10.5%	n=1
Newport	7.5%	n=1
Blaenau Gwent	3.8%	n=1

Table 4.2 shows that 25 out of the 40 participants lived in Swansea and Cardiff at the time of testing, where only 11.2-12.2% of the speakers in that local authority speak Welsh. This figure is below the national percentage of Welsh speakers in Wales, which sits at 17.8%. Moreover, ten participants lived in local authorities where there is a higher percentage of Welsh speakers than the national average, including Carmarthenshire, Gwynedd, Isle of Anglesey and Conwy (39.9%,

64.4%, 55.8% and 25.9% respectively) and four participants lived in local authorities where Welsh is spoken between 7.5% and 17.2% of the population (Newport, Caerphilly, Rhondda Cynon Taf and Pembrokeshire). One participant lived in Blaenau Gwent at the time of testing, which has the smallest percentage of Welsh speakers in all of Wales (3.8%).

At the time of testing, thirty-five out of the 40 participants lived in South Wales, while 5 lived in North Wales. Speakers from South Wales are typically exposed to less environmental Welsh than Welsh speakers in North Wales, as the dominant language in the majority of communities (in South Wales) is English. This consideration is taken forward into the measures empirically researching the comprehension and production (and processing) of grammatical gender in Welsh-English bilingual adults. Moreover, testing Welsh bilingual speakers who come from more geographically diverse areas than those tested in previous studies helps to develop our understanding of the different bilingual profiles of the Welsh speakers that we have here in Wales.

A short brief of the study was used to recruit participants. This was posted on my personal social media platforms, including *Facebook*, *Instagram,* and *Twitter*. The information was also posted in an online Welsh reading group and a Welsh monthly newsletter as part of *Swansea's Welsh Language Initiative* community-based organisation. Several 'Mentrau Iaith' (*Welsh Language Initiative* organisations) across Wales were also contacted. Additionally, a small number of participants were contacted directly by myself and my supervisor, including friends, family, and colleagues, as well as those who had previously completed my MA study (5/40). Participation was voluntary and unpaid. In total, 40 participants completed the battery of tasks.

4.1.3 Battery of tasks and Procedure

A battery of tasks was administered to the participants to investigate the comprehension and production of grammatical gender in Welsh adults and the possible effects of individual differences, including language dominance and linguistic proficiency. The battery included six items[9] and were completed in the following order:

- Bilingual Language Profile questionnaire (detailed background questionnaire and language dominance measure)
- comprehension of gender task (comprehension measure)
- elicited production task (production measure)
- productive vocabulary task (vocabulary measure)
- cloze test in Welsh (Welsh proficiency measure)
- cloze test in English (English proficiency measure)

For all of the battery items, besides the cloze tests (both Welsh and English), the data were collected online via the software *Gorilla.sc*. Participants completed the experiment on their own devices in their own homes while on Zoom with the researcher (myself). Data collection lasted between 45 minutes and 2 hours. Participants first read an information sheet and signed a consent form (in the language of their choice: Welsh or English) and all task instructions were presented bilingually. If the participants wished to take a quick break, it was suggested they do so after the elicited production task. The two cloze tests were completed on a separate link sent to the participants via the chat function on Zoom, on the *learningapps.org* website, after they completed the battery of tasks on *Gorilla.sc*. The participants were provided with a bilingual debrief sheet upon completion. Data were collected on Zoom for the following reasons: (1) to talk the participant through the process of the study, (2) to be of assistance at any point during the experiment and (3) to record the participants speaking during the production tasks.

Data collection took place during October, November, and December in 2020. Each participant was given the right to withdraw at any given time, and all participant data were anonymised with an ID code in line with GDPR. The study was granted ethical approval by the Faculty of Humanities and Social Science Research Ethics Committee (approval number 200720/2995) and complied with the ethical guidelines for conducting research with adults, as outlined in the university's guidelines. The following sub-sections describe each task in turn, outlining the

[9] An elicited imitation production task and a Flanker task were also administered as part of the battery. However, these items were removed due to some issues with administration and scoring. The EI task was administered after the BLP, and the Flanker task was administered after the vocabulary task.

stimuli and how the data were used to address the research questions put forward in this experiment.

4.1.4 Background Questionnaire

Section 3.5.1 discussed language dominance and reported that the BLP was selected to measure language dominance in the Welsh-English bilinguals. The Welsh and English versions of the BLP can be found in appendix A.

The BLP is a sociolinguistic questionnaire that assesses language dominance through self-reports. It systematically conceptualises and evaluates bilingual language dominance, providing a continuous dominance score and a general bilingual profile. It was developed using factor analytical techniques that support the internal dimensionality of the instrument (Bonvin et al., 2021). It also considers the multiple dimensions of dominance, such as age of acquisition of the L1 and L2, frequency and contexts of use, competence in different skills and attitudes towards each language. It is highly detailed and provides depth to its evaluation, operationalising dominance in a succinct and measurable way (Gertken et al., 2014). The BLP is freely available, quick to administer, easy to complete and can be adapted to specific linguistic situations (Bonvin et al., 2021; Gertken et al., 2014). Gertken et al (2014) also state that it is a reliable, valid, and highly practical instrument. More recently, Olson (2023) claimed that the BLP shows excellent levels of test-retest reliability in each of the four main individual components (following an introductory biographic section). Taken together, the BLP is considered to be a valid and reliable method of assessing language dominance (Gertken et al., 2014; Olson, 2023).

There are five sections in the BLP. The first section is an introductory section for biographic information, with the other four sections including questions about language history, language use, language proficiency, and language attitudes. Each of its 19 questions is asked for both languages (i.e., Welsh and English). The language history section contains six questions, asking about language learning onset, the age at which a participant felt comfortable using the language, the number of years pursuing school classes in each of the two languages, residing in a region/country where the languages are spoken, living with a family, and working in an environment where the languages were spoken. The second section contains five questions and gathers information on language use. The first three questions are about the average percentage of use of Welsh and English in an average week with friends, family and at work/university. The fourth and fifth questions ask how often the participant thinks and counts in Welsh and English. In the third section, four questions are asked on language proficiency, specifically abilities in speaking, understanding, writing, and reading. Participants self-evaluate their abilities on a 7-point scale. The last section on language attitudes contains four questions on a 7-point Likert scale. The questions asked about the degree to which speakers feel like themselves when speaking each language, whether they identify with that language speaking culture, how important using hat language like a native/L1 speaker is to them and whether they want other to think that they are a native/L1 speaker of that language.

The responses to the questionnaire generate a global score for each language (Welsh and English), a language particular score for each section which receives equal weighting, and a global score of dominance which ranges from -218 (Welsh dominant) to +218 (English dominant). A score close to 0 (+/- 20) indicates similar proficiency between the two languages, suggesting that the speaker is more of a balanced bilingual.

The BLP yields a quantitative score on a continuum, so that language dominance can be treated as a continuous variable as well as a categorical variable (Gertken et al., 2014; Olson, 2023). However, previous research has found it to be problematic when dominance is treated as a categorical variable. Solís-Barroso and Stefanich (2019) claimed that bilinguals would not be consistently placed into the same dominance group depending on which assessment is used (e.g., BDP / BLP) and that this would vary depending on when the participant is assessed. Solís-Barroso and Stefanich also argued that collapsing participants across the group level can obscure patterns that would be otherwise present if their linguistic behaviour was based upon a continuum of their dominance. Because of this, and the fact that the construct of dominance is inherently gradient and not categorical (Birdsong, 2016), language dominance is treated as a continuous variable in this experiment (and the follow-up experiment in this thesis).

A number of additional questions were added to the questionnaire, to the biographical information, language history and language use sections. This was to better understand the profile of the participants and to get a more complete picture of the bilinguals in this study. It is worth noting that these new questions were not incorporated in the calculation of the dominance score that has been described above for comparability purposes. However, these questions may prove useful in the discussion of the results, in providing invaluable supplementary information about the Welsh-English bilingual speakers. These questions can be found in appendix A.

Global dominance scores are generated to understand the possible role of language dominance in the Welsh-English bilingual adults' production and comprehension of grammatical gender. The scores will be used to address the two sub research questions (1a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' comprehension of Welsh grammatical gender, when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?' and (2a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' production of Welsh grammatical gender, when gender is independent of mutations and in conjunction with mutations, in local contexts?'.

4.1.5 Comprehension of Grammatical Gender Task

The first experimental task in the battery was the comprehension of grammatical gender task. This was a close replication of Binks and Thomas' (2019) task, which was discussed in section 3.2.2. The task was used to assess the comprehension of grammatical gender in non-local gender marked contexts. The task assessed two different instances of gender marking: non-local gender marking via the possessive adjective form '*ei*' (third person singular masculine and feminine) and via anaphoric pronouns '*fe*' (3rd person singular masculine) or '*hi*' (3rd person singular feminine) (see Section 2.4 Mutation and Gender in Welsh). When gender is marked via anaphoric pronouns, this is an instance of when gender operates independently of mutations. Whereas, when gender is marked via the possessive adjective, this is an instance of when gender is encoded through mutations (see chapter 2 for a more detailed discussion).

The participants were given pairs of sentences. The first sentence included two nouns, while the second sentence included either a possessive adjective or an anaphoric pronoun, referring to one of the nouns in the first sentence. The pairs of sentences were accompanied with a depiction of the two nouns, conveying the meaning of the second sentence. An example is shown below, taken from Binks and Thomas' study (2019, 1028-1029; images taken from Binks, 2017, appendix B).

Example 4.1 First sentence

Roedd	y	drwm	tenau	a'r	bêl	frown	mewn	bocs
Be-PST	the DET	drum NOUN (m)	thin ADJ	and the	ball+SM NOUN (f)	brown+SM ADJ	in PREP	[a] box NOUN (m)

"The thin drum and the brown ball were in a box"

The first sentence always included one masculine noun and one feminine noun, each followed by an adjective. The feminine noun and modifying post-nominal adjective are marked by the soft mutation (noun: pêl > bêl, adjective: brown > frown), while the masculine noun and adjective do not undergo mutation following the determiner 'y'. Here, the SM is a local gender marker in the sense that the marker occurs on the noun itself or its immediate modifiers (Gathercole, Thomas & Laporte, 2001). Example 4.2 is the accompanying second sentence.

Example 4.2 Second sentence

Ond	cwympodd	hi	drwy'r	gwaelod
But CONJ	fell VERB	she *FEM3SG*	through PREP the DET	bottom NOUN (m)

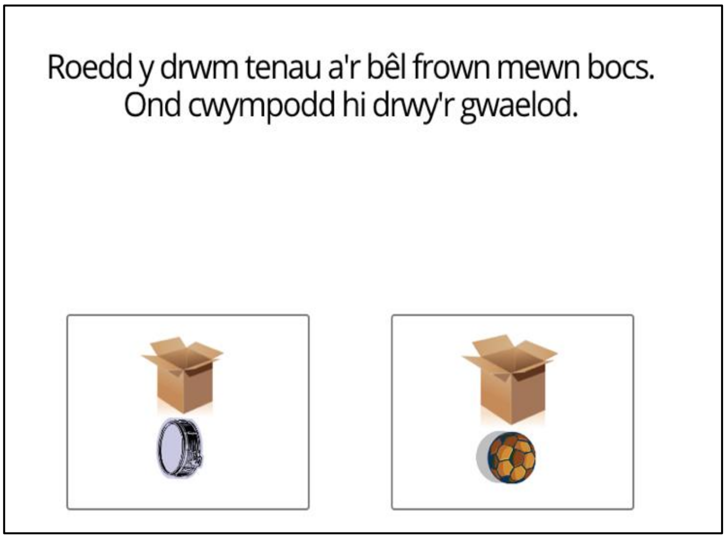
"But she fell through the bottom"

The second sentence included a non-local reference, either the possessive adjective form '*ei*' (third person singular masculine and feminine) or the anaphoric pronouns '*fe*' (3rd person singular masculine) or '*hi*' (3rd person singular feminine). These non-local references must agree with the gender of the antecedent noun in the first sentence (e.g., *drwm* [m] or *pêl* [f]), therefore showing non-local gender marking in the second sentence via the reference. The participants were asked to indicate whether the reference (possessive adjective or anaphoric pronoun) in the second sentence referred to the feminine or masculine noun in the first sentence. In Example 4.2, the anaphoric pronoun 'hi' [f] agrees in gender with the feminine noun in the first sentence and shows gender when it is independent of mutations.

The pairs of sentences were shown in conjunction with two images. The images were depictions of the two nouns named in the first sentence but conveyed the meaning of the second sentence. For instance, in the example provided above, ‘Roedd y drwm tenau a'r bêl frown mewn bocs. Ond cwympodd hi drwy'r gwaelod’, both nouns *drum* [m] and *ball* [f] were shown to have fallen through the bottom of the box. The participants were required to select one of the two images which corresponded to the item referred to in sentence two. In this example, as *bêl* (ball) is feminine and the pronoun in the second sentence is ‘hi’ (she) feminine, the participant should select the image with the ball falling through the box, not the *drum* [m]. Figure 4.1 is a screenshot of how the trial appeared in the task, as shown to the participant in Gorilla.

Figure 4.1

Receptive Task Trial Example



In Binks and Thomas' (2019) version, the task included 30 pairs of sentences – 18 trials included animate nouns (6 human; 12 animal) and 12 trials included inanimate nouns. An additional 6 inanimate trials were created and added to the task to ensure a balance of animate (18 trials) and inanimate trials (18 trials). Of the 30 original pairs of sentences, 8 nouns were changed in total, with six being changed to include noun forms typically used by Welsh speakers from South Wales, and the other two changing to balance gender across the conditions as some of the new noun forms changed in gender. This was done to ensure the task was suitable for a wider geographic spread of Welsh speakers, rather than only for Welsh speakers from North Wales. Of the 36 sentences, 10 targeted feminine nouns with the possessive adjective *ei* + AM, 10 targeted masculine nouns with the possessive adjective *ei* + SM, 8 targeted feminine nouns with the feminine anaphoric pronoun ‘hi’, and 8 targeted masculine nouns with the masculine anaphoric pronoun ‘fe’. The sentences can be found in appendix B.

The majority of the images used in the original task were included in this version, but not all. Four new images replaced original versions to convey better visualisations of the images and four new images were created to convey the meaning of the added inanimate noun trials. The new images were taken from *Vectorstock.com* and *Clipart-library.com,* and these were manipulated in *Word Paint*. All of the images were in colour and saved as *png* files. A total of 70 images were used in the test.

Following task instructions, the participant took between 5-7 minutes to complete the task. The trial order was fixed for all participants. The data were scored for correct (1 point) and incorrect (0 points) responses. The test was scored out of 36. The results from the comprehension of grammatical gender task will be used to address the research question – (1) ‘To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in comprehension when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?’. Assessing gender when it is independent of mutations and when it is encoded through mutations will contribute to our understanding as to whether Welsh-English bilingual adults show accurate use of Welsh grammatical gender in non-local contexts when reading for comprehension.

4.1.6 Elicited Production Task

The elicited production task was a close replication of Sharp's (2012) elicited production task. The task assessed local gender marking of nouns via three gendered numeral forms, the numbers two, three and four which have masculine and feminine forms (see section 2.4 Gender in Welsh for a detailed description). The different forms are given in in the Table 4.3.

Table 4.3

Gender Marked Numerals in Welsh

Numeral	Masculine Form	Feminine Form
Two	Dau	Dwy
Three	Tri	Tair
Four	Pedwar	Pedair

The task assessed whether these numerals were produced by adults in correct gendered forms and the relationship between gender and mutations on the following target nouns. As previously mentioned in section 2.4 Mutations and Gender in Welsh, the grammatical gender system and mutation system often work in conjunction with one another. The mutation rules following the numeral in the phrase are presented in the Table 4.4.

Table 4.4

Numeral Mutation Rules

Numeral	Mutation
Dwy ('2' F)	SM
Dau ('2' M)	SM
Tair ('3' F)	No mutation
Tri ('3' M)	AM
Pedair ('4' F)	No mutation
Pedwar ('4' M)	No mutation

The task was replicated as closely as possible; however, some changes were made to the target nouns. A list of these 12 changes can be seen in the Table 4.5 and the full list of nouns used in this task, including their gender and animacy, can be found in appendix C.

Table 4.5

Noun Changes

Original Noun	New / Replaced Noun	Translation	Reason for change
Taid	Tadcu [m] [human]	Grandad / Grandfather	North > South Wales
Llyffant	Broga [m] [animal]	Frog / Toad	North > South Wales
Prycop	Corryn [m] [animal]	Spider	North > South Wales
Dynes	Menyw [f] [human]	Woman	North > South Wales
Malwen	Malwoden [f] [animal]	Snail	North > South Wales
Teisen	Cacen [f] [inanimate]	Cake	North > South Wales
Twll	Tegell [m] [inanimate]	Hole > Kettle	Depictable image
Llinyn	Llyfr [m] [inanimate]	Line/String > Book	Depictable image
Brigyn	Beic [m] [inanimate]	Twig > Bike	Depictable image
Dinas	Diod [f] [inanimate]	City > Drink	Depictable image
Popty	Plât [m] [inanimate]	Oven > Plate	N > South + mutatable letter
Bodyn	Drych [m] [inanimate]	? > Mirror	Unfamiliar noun

There were three phases to each trial (48 trials). Each trial followed the same pattern with two pre-recorded sentences, followed by one elicited sentence produced by the participant. This is shown in the three figures below. The number five was used in the second phase as it is neither marked for gender nor triggers a mutation.

For example:

Mae	pum'	**cath**
Be-$_{3SG}$	five $_{NUEMRAL}$	cat $_{NOUN\ (f)}$

'There are five cats'

Mae	pum'	**ci**
Be-$_{3SG}$	five $_{NUEMRAL}$	dog $_{NOUN\ (m)}$

'There are five dogs'

Figure 4.2

Example Trial: Phase 1

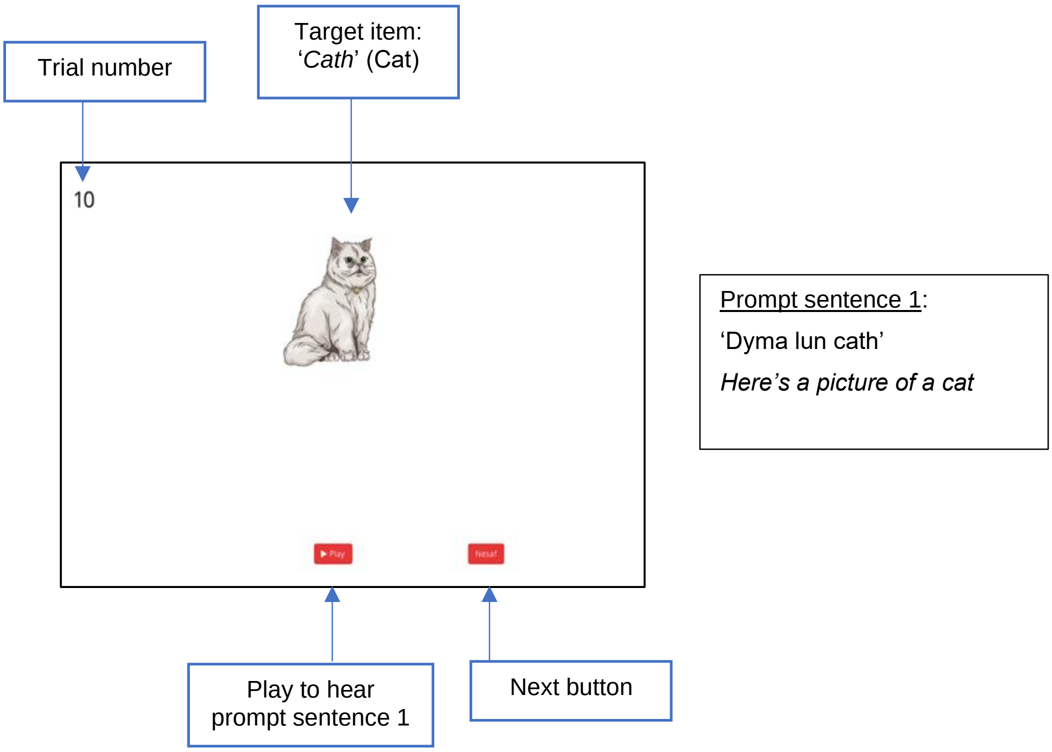


Figure 4.3

Example Trial: Phase 2

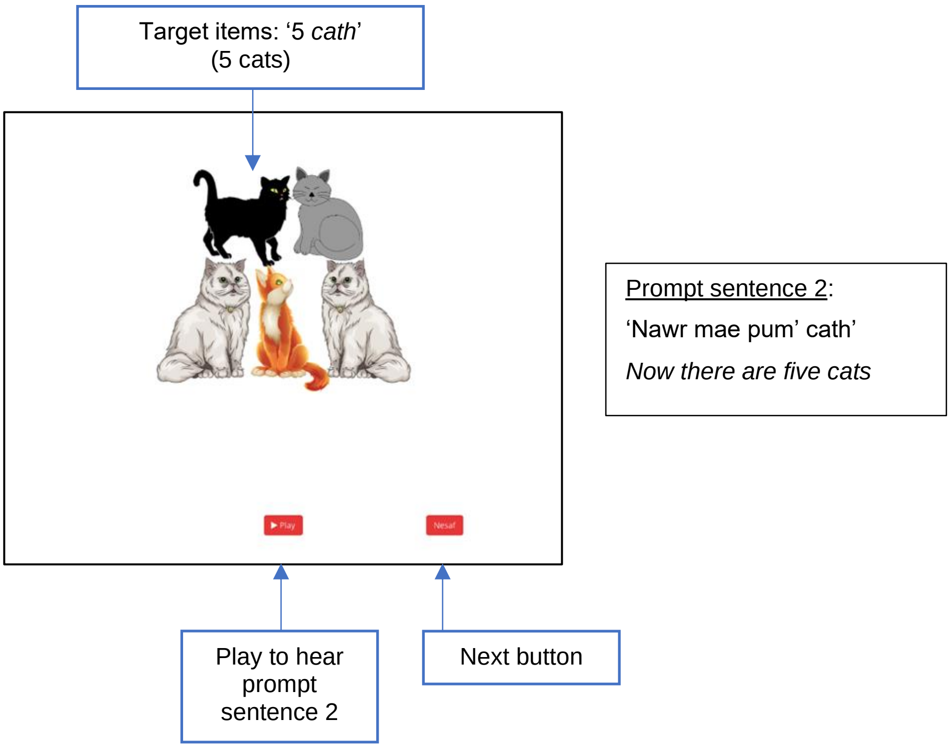
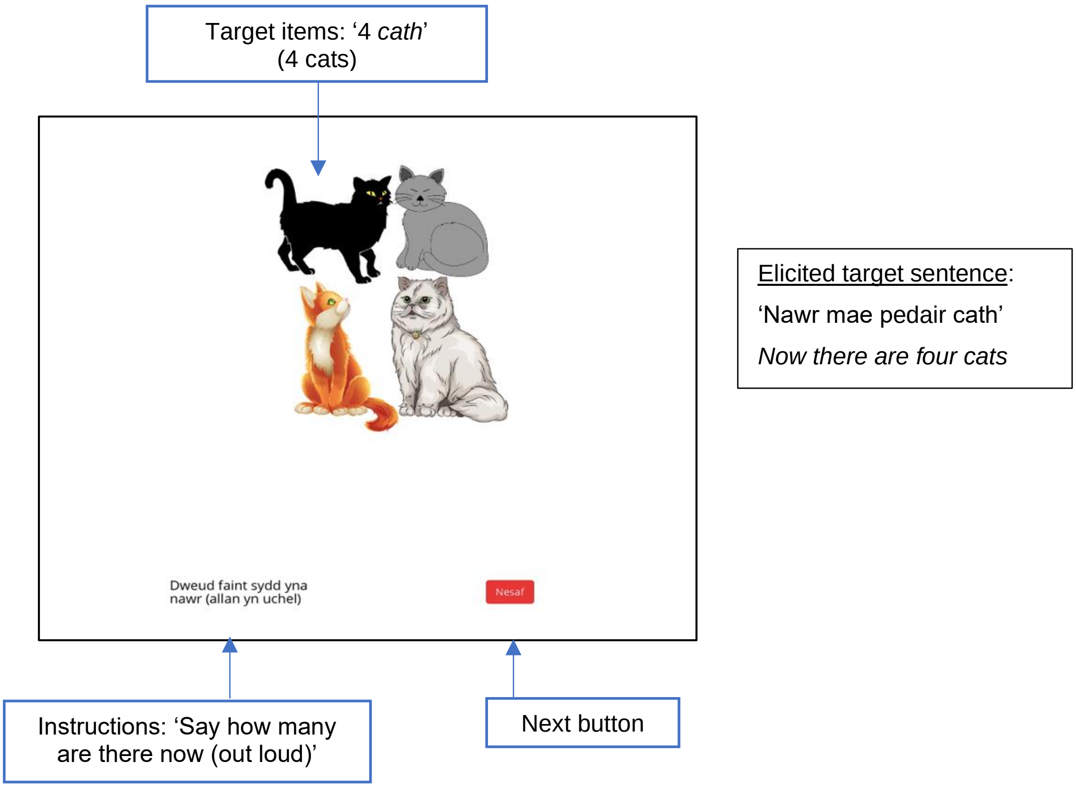


Figure 4.4

Example Trial: Phase 3



In Sharp's task, phase one included three different syntactic conditions: (1) no cues, (2) one syntactic cue, and (3) two syntactic cues. This element was removed from the task because it was a confounding factor, in the sense that it conflated the gender and mutation systems prior to the participants' sentence production. These syntactic cues were mutations which indicated the gender of the noun. This could prime the participants in identifying the gender of the noun, making the production of the gendered numeral easier. Therefore, it was removed from this task.

Following personal communication with Sharp, new images were used in the task. The new images were taken from *Vectorstock.com* and *Clipartlibrary.com*, which were all in colour and saved as *png* files. The images were shown in sets of three along with the three phases of each trial, totalling in 144 newly created images (48 sets of three, with 3 sets of images per trial).

The sentences for phase one and phase two were pre-recorded by a Welsh female speaker. They were recorded at a moderate pace with neutral intonation. The sentences were recorded, cropped, and cleaned in *Audacity* (version 2.4.2). Each sentence was saved as an individual sound file and uploaded into the task on

Gorilla.sc. The elicited sentence uttered by the participant in phase three was recorded through Zoom. This was because when using the recording device zone in Gorilla during trial runs by myself and the pilot participants, there was a lag when capturing the spoken sentences. Therefore, spoken stimuli were recorded through Zoom to ensure all spoken utterances were captured.

Following task instructions, the participants took 8-10 minutes to complete the task. The trial order was fixed for all participants (as in the previous experimental measure). The data were scored for correct numeral (1 point) and incorrect numeral (0 points) to give grammatical gender scores. This score was out of 48.

A second scoring system was established to explore the relationship between gender and mutation(s). The participants were given 1 point for correctly producing the noun in either its bare form following *tair, pedwar* and *pedair*, or 1 point for correctly mutating the noun following the numerals *dau, dwy* and *tri*. The participant was given 0 for incorrectly mutating the noun (e.g., SM in place of AM) or for not mutating the noun (e.g., no mutation in place of SM). Therefore, two sets of 48 points are available: (1) 48 points for correctly selecting the gendered numeral form, and (2) 48 points for correctly mutating or not mutating the noun. These scores are presented in section **Error! Reference source not found.** Elicited production task results.

The results from this elicited production task will be used to the address research question – (2) 'To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in production when gender is independent of mutations and in conjunction with mutations, in local contexts?'. Assessing gender in this way will contribute to research exploring whether Welsh-English bilingual adults show productive command of grammatical gender in Welsh, whilst gaining a better understanding of how Welsh speakers produce gendered numerals. It will also help the filed to understand the relationship between gender and mutations, specifically, when gender is independent of mutations and in conjunction with mutations in local contexts.

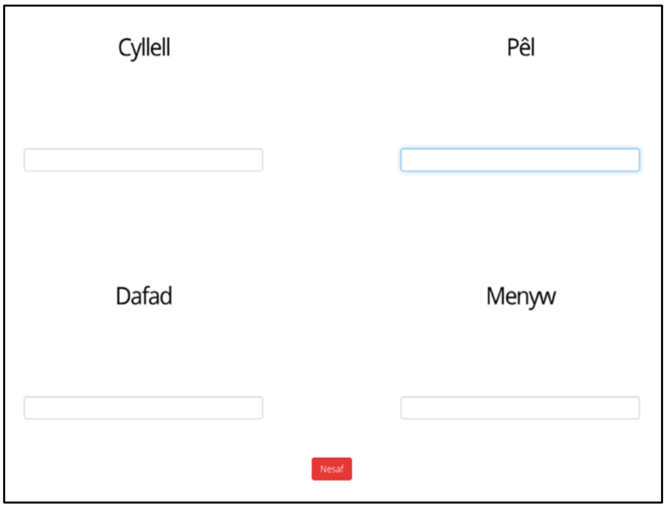
4.1.7 Productive Vocabulary Task

In order to assess how Welsh adult speakers comprehend and produce grammatical gender, it is useful to know whether the participants have knowledge of the target vocabulary included in the experimental tasks. Following previous methodologies (e.g., Arnaud, 1992; Fitzpatrick & Meara, 2004), participants' productive vocabulary knowledge was assessed using a straightforward translation task. The test is productive by nature, in the sense that the participant is required to generate a translated form of the target word (Laufer et al., 2004). The measure allows selected vocabulary items to be predetermined by the researcher (Jiménez Catalán & Moreno Espinosa, 2005), testing how well the participant knows the word (Pignot-Shahov, 2012). If the participant is not able to produce the correct translation of the target noun, it could be questioned whether they can be expected to comprehend or produce the gender of the target noun in the experimental measures of gender. All of the nouns from the two tasks were included in this measure. This productive vocabulary task was used to check whether the participants were familiar with the target nouns in the experimental tasks outlined above.

A total of 70 nouns were used in the experimental tasks. In Gorilla, the participants were presented with four words per screen and were required to type in the English translation in the space below the target noun. In total, there were 17 screens displaying four nouns and one screen displaying two nouns. The full list of nouns can be found in appendix D. Figure 4.5 is an example screenshot of the task.

Figure 4.5

Productive Vocabulary Measure Trial Example



Following instructions, the participant completed the task which typically took 4-5 minutes. This task was administered after the experimental tasks to avoid any priming to the target nouns. For each correct translation, participants were given a score of 1; for each incorrect translation, participants were given a score of 0. They were not scored for spelling. The target nouns with lexically acceptable alternatives were also considered correct and the target items with lexically acceptable alternatives are shown below.

Table 4.6

Target Vocabulary Synonyms

Welsh noun	Translated form (1)	Lexically acceptable
Tad-cu [m]	Grandad	Grandfather
Menyw [f]	Woman	Female
Pêl-droediwr [m]	Footballer	Football player
Prifathrawes [f]	Headmistress	Female headteacher
Mam [f]	Mum	Mother
Plismon [m]	Policeman	Police officer
Merch [f]	Daughter	Girls
Dysgl [f]	Dish	Plate
Cyfnither [f]	Cousin	Female cousin

Regional forms of the noun were also considered, as we assessed Welsh speakers from a wide geographic spread across Wales. There were two nouns with different possible translated forms based on different geographic locations in Wales. These are shown below.

Table 4.7

Regional Translated Forms

Welsh noun	South Walian form	North Walian form
Broga [m]	Frog	Toad
Gwasgod [m]	Waistcoat	Jacket

By testing the speakers’ productive vocabulary knowledge, the items that each participant translated incorrectly will be removed from the experimental data sets to ensure that the participants comprehension and production of grammatical gender is only being measured for the target nouns that they are familiar with. These results will also help me to decide which nouns should or should not be included in the follow up experiment investigating the processing of gender.

4.1.8 Cloze Tests

In section 3.5.2, I defined the construct of linguistic proficiency, described its role in the predicting different outcomes in bilingual performance and outlined the various tools available to capture linguistic proficiency in speakers. Cloze tests were selected as they are an independent measure of proficiency (Ayoun, 2019), they are often sufficiently integrative (e.g., it not only assesses morphosyntactic competence, but also lexical and discourse competence) and scores can discriminate between the bilingual participants of interest (Tremblay & Garrison, 2010). Moreover, cloze tests are easy to create, they can take a relatively short period time to complete, they are flexible in format and can be easy to score if clear scoring criteria are established. This makes using cloze test a practical tool for assessing proficiency in experimental research (Tremblay & Garrison, 2010). Even though it would be possible to discriminate between lower and higher proficiency speakers, the aim of administering the cloze tests is to include their proficiency scores as continuous predictors into the linear mixed effects modelling, rather than to classify the bilinguals according to proficiency levels (e.g., low vs. high).

Following previous studies (e.g., Gaillard & Tremblay, 2016; Tremblay, 2007, 2008, 2009; Tremblay & Garrison, 2010), Brown's (1980) cloze test was selected to measure English proficiency in the bilingual participants. The test was created for the Guangzhou English Language Centre in China, and was subsequently used at the University of Hawaii as part of their English Language Institute placement test (Tremblay & Garrison, 2010). It is considered to be a reliable and valid measure of proficiency in English. The task format requires the participant to supply words that have been removed from a passage. It is a 399-word passage discussing mankind’s evolution, with a 26-word lead in and a 29-word finish. A fixed ration-deletion method (ratio 1:7) was employed, with fifty words having been deleted, of which 26 were content words (i.e., open-class: nouns, adjectives, verbs, etc) and 24 were function words (i.e., closed-class: determiners, prepositions, pronouns, etc). A bank of exact and lexically acceptable answers was made available by Brown (1980). One point was given for each acceptable response, and no points were given for unacceptable responses.

While cloze tests are used to measure L2 Welsh adult learners Welsh speaking proficiency levels as part of the WJEC examination board, these cloze tests were considered unsuitable to administer as part of the current experiment given that the participants were not learners. Additionally, the WJEC cloze tests were 300-word passages which only included 16 gaps, with words deleted on a 1:18 ratio, which is an insufficient number of gaps. With no standardized Welsh proficiency measure available, a Welsh cloze test was created to measure proficiency in Welsh adult bilinguals.

In order to find a passage of suitable topic and length, *Golwg* (a Welsh language magazine) as well as the three levels of Welsh for adult's examination papers were scanned (foundation, intermediate, higher). However, passages often included conversations and were too long in length (500+ words). Also, the majority of resources were inaccessible. The L1 Welsh A-level exam papers were also scanned, with passages of suitable length, but included extremely complex vocabulary discussing complex topics. Finally, a passage was selected from an L1 Welsh GCSE exam paper. This was considered to be accessible in content and difficulty, suitable in topic and in length for the Welsh-English bilingual adults tested in this experiment. This was a non-academic passage on BBC1's Blue Planet.

Following previous experimental research (Tremblay & Garrison, 2010), the rational-deletion method was employed to elicit a balanced proportion of content and function words. With a 1:7 ratio deletion applied to Brown's English cloze test, a ratio deletion method of 1:6 to 1:8 was applied to the 358-word Welsh passage. In total, 44 words were deleted, of which 22 were content words and 22 were function words. A bank of acceptable answers was created on the basis of the participants' responses which were used for scoring the test. This was done because scoring cloze tests on the basis of acceptable answers has more face validity, as it is rarely the case that only one word is possible in any given context (Tremblay & Garrison, 2010). One point was given for each acceptable response, and no points were given for unacceptable responses.

At the time of data collection, it was not possible to administer the two cloze tests in Gorilla.sc. Therefore, *learningapps.org* was used to create and administer the tests. This website tool was selected as it was possible to apply the selected ratio deletion methods of choice and to include open-ended answers (i.e., not multiple-choice options). Each cloze test took between 5-15 minutes to complete. The participants completed the Welsh cloze test, followed by the English cloze test. The Welsh and English cloze tests, along with the bank of exact and lexically acceptable answers for both tests, can be found in appendix E.

To measure the internal consistency of the cloze tests, a Cronbach's alpha was performed for each test. For the Welsh cloze test, a test for internal consistency was found to demonstrate an acceptable level of internal consistency, $\alpha = .818$. There were 44 gaps in the cloze test, however, 11 component variables were removed from the analysis because all of the participants scored 100% on those 11 items (7 function words and 4 content words). For the English cloze test, the score was found to be quite low, $\alpha = .531$, which does not suggest that it is a reliable measure of English proficiency. There were 50 gaps in the cloze test, however, nine items were removed from the analysis because all of the participants scored 100% on those nine items (3 function words and 6 content words). Given the Cronbach's alpha scores, English linguistic proficiency is not considered as an individual difference variable nor is it included in any of the subsequent analyses, only the Welsh cloze test scores are considered moving forward.

The scores from the Welsh proficiency test will be used to address the research questions (1a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' comprehension of Welsh grammatical gender, when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?' and (2a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' production of Welsh grammatical gender, when gender is independent of mutations and in conjunction with mutations, in local contexts?'.

4.1.9 Pilot

The battery of tasks was pilot tested with two female Welsh speakers, one aged 22 and one aged 60. Pilot testing was carried out to ensure that (1) task instructions were clear, (2) the tasks elicited the target responses, (3) to see the length of time to complete the battery of tasks, and (4) to ensure that it was not too difficult to navigate for participants of different ages (e.g., using zoom and opening Gorilla online). Pilot testing was also done to see whether the tasks were too advanced for Welsh speakers who are simultaneous bilinguals (English dominant, aged 22) and later bilinguals (Welsh AoA=12, Welsh dominant, aged 60). These bilinguals had different lengths of exposure to Welsh and different acquisition/learning experiences. The battery was completed with understanding and ease, without any difficulties navigating the tasks online (via Gorilla.sc and learningapps.org). Upon inspection, the tasks elicited the target responses. Both participants completed the battery within 1 hour. The following sections present the results from this first experiment.

4.1.10 Data Analysis

In this experiment, there were two experimental tasks, one assessing the comprehension of gender and the second assessing the production of gender in Welsh-English bilingual adults. The responses from the tasks were scored in a binary fashion, in such a way that subject responses were coded as one for correct and zero for incorrect. The scores for each task are the two different dependent variables. After calculating descriptive statistics, all responses were analysed fitting generalised linear mixed-effects models.

Analysing data using mixed-effects has become increasingly popular within many areas of research (Magezi, 2015), as they have several advantages over common analyses of variance, such as standard or repeated measures ANOVA (Brown, 2021). First, ANOVAs work with aggregated means, reducing important variability in the data, whereas mixed-effects models consider each of the participants' observations separately, modelling data using single data points. Second, ANOVAs model item-level and participant-level variability, but this is not possible to do so simultaneously. Mixed-effects models, however, allow for all the factors under investigation to be considered at once, thereby obviating the need for separate by-participants and by-items analyses (Cunnings, 2012; Cunnings & Finlayson, 2015). they are able to account for random populations that share a nested relationship, are able to capture the crossed relationship between counterbalanced sets of linguistic stimuli presented to different subjects and analyse all of the individual data points in the data set (i.e., no computed means across groups) (Lo & Andrews, 2015). Another benefit of mixed-effects models is the treatment of missing observations and unequal sample sizes per group (Quené & Van Den Bergh, 2004). ANOVAs can considerably reduce sample size, as it would either delete the whole participant or the whole item where there are missing values. The final advantage of mixed-effects models is that it is possible to include both categorical and continuous predictors (i.e., independent variables), as well as analyse categorical dependent variables. Mixed-models measure how well an outcome variable can be predicted by fixed and random effects (Brown, 2021). Fixed effects model how the independent variable(s) affect(s) the outcome variable, while random effects model variance that can be attributed to other factors inherent in the sampling of the study (Cunnings, 2012).

To find the best fitting model for each of the outcome variables, which is a model that best explains the data, the models were evaluated via forward model comparison using the anova() command to perform the likelihood ratio test criterion (Bates, Mächler, Bolker, & Walker, 2015; Cunnings, 2012). For both tasks, generalized linear mixed-effects modelling *glmer* function was used and the additional parameter (link function) *family = binomial* was specified as the outcome variable is binary (Brown, 2021; Linck & Cunnings, 2015). To help the model(s) converge (i.e., to find a good fit for the data within a reasonable number of iterations of attempting to estimate model parameters), control parameters were added, including the argument *control = glmerControl*, the optimizer = *bobyqa* was specified and the number of iterations were specified using the argument *maxfun = 1000000*. Additionally, the argument *calc.derivs = FALSE* was specified to remove some of the derivative calculations that occur after the model has reached a solution (Brauer & Curtin, 2018; Brown, 2021). The random effects structure was maximally specified (Barr et al., 2013), but one that is theoretically motivated and supported by the data (e.g., Bates et al., 2015a; Bates et al., 2015b). The model structures are reported in the results subsections of the tasks.

All models were run in the R programme for statistical computing (R Core Team, 2022), using version 4.2.2 (2022.10.31) and version 1.1- 30 of the lme4 package (Bates et al., 2015). All models were fit with the Laplace Approximation for Maximum Likelihood using lme4 function (Bates et al., 2015).

4.2 Results: Overview

Section 4.2 reports the results of the tasks administered in this first experiment, described in the previous section (Section 4.1). The battery of tasks was completed by 40 Welsh-English adult bilinguals. The two main research questions in this first experiment are (1) ‘To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in comprehension when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?' and (2) 'To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in production when gender is independent of mutations and in conjunction with mutations, in local contexts?'. There is also a secondary interest to both research questions, investigating the roles of dominance and proficiency in Welsh-English adult bilinguals' comprehension and production of Welsh gender.

To address these research questions, the results are divided into three sections. The first section reports the results for the productive vocabulary task, followed by the comprehension of grammatical gender task results and the elicited production task results. Finally, section 4.3 discusses these results in light of the literature reviewed in chapter 3.

4.2.1 Productive Vocabulary Task

The productive vocabulary task assessed the participants' productive vocabulary knowledge using a straightforward translation task, where participants were given nouns in Welsh and were asked to translate them into English (see section 4.1.7 for methodology). The task was used to check whether the participants were familiar with the target nouns and whether they understand the target vocabulary included in the experimental tasks assessing the comprehension and production of gender. Following previous methodologies (Hopp, 2013, 2016b), the nouns which were incorrectly translated by the participants were removed from the subsequent analysis. Therefore, it will be possible to say that if and when the participants experience any difficulty comprehending or producing the gender of the target nouns in the experimental tasks, it is not due to a lack of familiarity with the target noun.

In total, 70 nouns were assessed, which were the target nouns from the two experimental tasks. The data were scored for correct translation (1 point) and incorrect translation (0 points) to give productive vocabulary scores. Lexically acceptable items were also considered correct (see section 4.1.7 for details and appendix E for the list of acceptable translations).

Table **4.8** shows the group scores.

Table 4.8

Productive Vocabulary Task: Group Descriptive Statistics

Group descriptives: Vocabulary	Raw	Percentage
Mean	68.5 / 70	97.9
SD	3.45	4.93
Range	55 - 70	77 – 100

Table **4.8** shows that the participants performed at ceiling. The overall group average was 97.9% (*SD*=4.93), with a range of 77% to 100%. The results revealed that 25 out of the 40 participants scored 100%, 11 participants scored between 95% - 99%, one scored between 90% - 94%, one participant scored between 85% - 89%, and another between 80% - 84%, while one participant scored between 75% - 79%. The participant who scored lowest (77%) incorrectly translated 16 of the 70 target nouns. The results also showed that 67 out of the 70 nouns were translated correctly by 90% or more of the participants, with 47 of the target nouns having been accurately translated by all 40 participants. This shows that the participants performed at ceiling and are therefore highly familiar with the majority of the target nouns included in the experimental tasks. Therefore, this indicates that there are no individual problematic items.

The nouns which were incorrectly translated by the participants were removed from the subsequent analysis. The data analysed was based on the accuracy of the translated nouns by the Welsh-English bilingual adult speakers and the number of items removed in each task are reported in the experimental task results sections.

4.2.2 Comprehension of grammatical gender task

This task assessed the comprehension of grammatical gender in non-local contexts, where gender was marked via possessive adjective forms (gender encoded through mutations) and anaphoric pronouns (gender independent of mutations). There were 36 trials in total which were balanced for gender (masculine and feminine) and animacy (animate and inanimate). Of the 18 animate trials, 12 included animal nouns and 6 included human nouns. The breakdown of items is shown in the table below.

Table 4.9

Comrehension task: Breakdown of Items

With / without mutations by gender	Object, n=18	Animal, n=12	Human, n=6
Masculine: gender -mutation (Fe)	5	3	-
Feminine: gender -mutation (Hi)	5	3	-
Masculine: gender +mutation (Ei+SM)	4	3	3
Feminine: gender +mutation (Ei+AM)	4	3	3

Participants were given a score of 1 for correct responses and a score of 0 for incorrect responses. Participant trials were excluded from the data set if they incorrectly translated the target noun in the productive vocabulary task. This was to ensure that the participants' comprehension of grammatical gender was only being measured for the target nouns that they were familiar with. In total, 13 trials were excluded from the data set and these items were excluded from seven participants, ranging from 1-5 items per participant. First, the overall task accuracy scores are presented and these are shown below.

Table 4.10

Comprehension Task Overall Performance

Comprehension task descriptives	Raw	Percentage
Mean	27.53/36	76.46
SD	6.66	18.51
Range	13 - 36	36 – 100

Table 4.10 shows that the overall group mean is 76.46% (*SD*=18.51%), where the participants scored on average 28 out of 36 on the task (*SD*=6.66). This suggests that as a group, they make use of grammatical gender in comprehension in non-local contexts. However, the standard deviation shows that there is individual variation within the group results and that there is a wide range in scores (36 – 100%). This indicates that some of the participants show accurate use of grammatical gender in comprehension, while other participants show less accurate use, of gender when it is marked via possessive adjective forms (gender encoded through mutations) and anaphoric pronouns (gender independent of mutations).

To better understand the results, the table below shows the scores according to the different linguistic conditions, namely, gender when it is marked via possessive adjective forms and anaphoric pronouns, and these are divided according to grammatical gender (masculine and feminine).

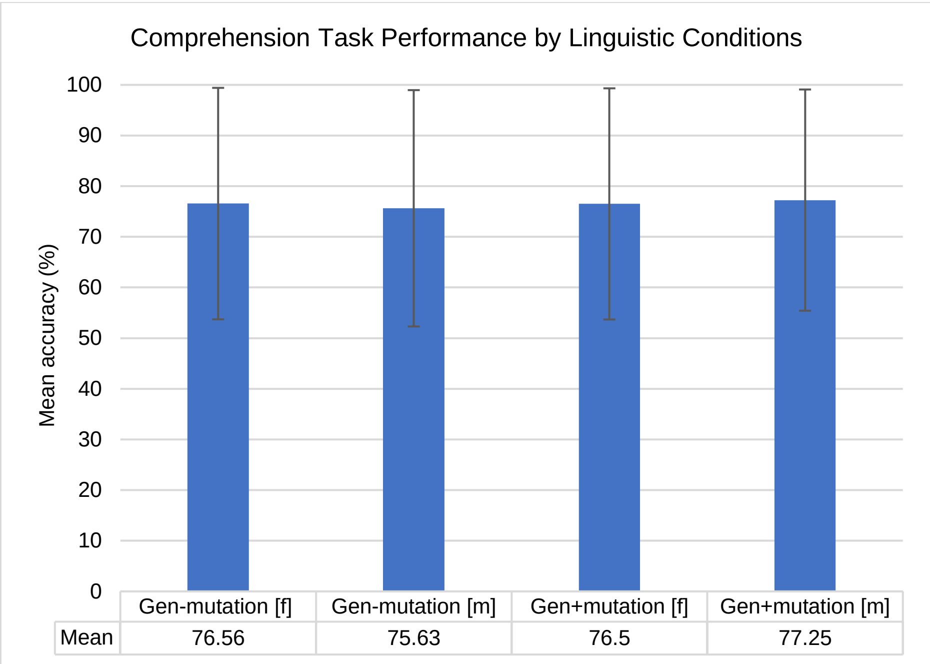
Table 4.11

Comprehension Task Performance by Linguistic Conditions

Comprehension	**Gen-mut [f]**		**Gen-mut [m]**		**Gen+mut [f]**		**Gen+mut [m]**	
task descriptives	Raw/8	%	Raw/8	%	Raw/10	%	Raw/10	%
Mean	6.13	76.56	6.05	75.63	7.65	76.50	7.73	77.25
SD	1.83	22.85	1.87	23.33	2.28	22.82	2.18	21.84
Range	2-8	25-100	2-8	25-100	3.10	30-100	1-10	10-100

Figure 4.6

Comprehension Task Performance by Linguistic Conditions



The results show that performance is similar across the two linguistic conditions for both genders. The participants scored highest on trials where gender is marked via possessive adjective forms for masculine nouns with SM (*M*=77.25, *SD*=21.48), next best on trials where gender is marked via anaphoric pronouns for feminine nouns (‘hi’ - no mutations involved) (*M*=76.56, *SD*=22.85), followed by trials where gender is marked via possessive adjective forms for feminine nouns with

AM (*M*=76.50, *SD*=22.82), then trials where gender is marked via anaphoric pronouns for masculine nouns (‘fe’ - no mutations involved) (*M*=75.63, *SD*=23.33). The standard deviation for the different conditions shows that there is a wide range in scores, from 25-100% for both masculine and feminine trials including anaphoric pronouns (gender -mutations), 30-100% for masculine trials including possessive adjectives (gender +mutations) and 10-100% for feminine trials including possessive adjective forms (gender +mutations).

The group results suggest that the participants identify the target nouns similarly across the different linguistic conditions and across both genders. The participants are overall scoring above chance for both linguistic conditions, indicating that they are able to show accurate use of gender is non-local contexts, when gender is independent of mutations and when it is encoded through mutations.

To understand how the participants responded to the items in the comprehension task, the best-fitting mixed effects model is found and reported. Gender (masculine, feminine) is included as a factor in the model, as well as pronoun type (possessive, anaphoric). To check the possible relationship between gender and mutations, model one included the fixed effects of gender and pronoun type (no interaction), and the second model included the interaction of gender and pronoun type. This allowed me to test whether including the interaction better explained the data. For the random effects structure, there was a by-participant random intercept as well as by-participant random slopes for gender and pronoun type, as well as associated correlation parameters. There was also a by-item random intercept. Deviation coding was set for gender and pronoun type (Schad, Vasishth, Hohenstein, & Kliegl, 2020).

The models evaluated via forward model comparison revealed that the model fit did not improve with the interaction of gender and pronoun type ($\chi^2(1) = 0.02$, $p = 0.88$), but with them only as main effects. Therefore, the most complex model justified by the data included the fixed effects of gender and pronoun type, but without the interaction between the two factors. The output from the best fitting model is below.

Table 4.12

Best-fitting model: Comprehension task

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	1.91	0.32	6.02	**<0.001**
Gender [masculine]	-0.39	0.37	-1.08	0.28
Pronoun type [possessive: gender + mutations]	-0.07	0.37	-0.19	0.85
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	2.62	1.61	0.90	0.95
Gender [masculine]	0.36	0.60	-	-
Pronoun type [possessive: gender + mutations]	0.53	0.72	-	-

The model's intercept is significant (β = 1.91, SE = 0.32, z = 6.02, p = <0.001), which represents the estimate of the average of condition means (gender and pronoun type). However, there was no effect of gender (β = -0.39, SE = 0.37, z = -1.08, p = 0.28) nor was there an effect of pronoun type (β = -0.07, SE = 0.37, z = -0.19, p = 0.85). To check the interactions between gender and pronoun type, pairwise comparisons were run using *emmeans* (see Lenth et al., 2022) with the adjust argument specifying 'Bonferroni' correction. The pairwise comparison results supported those found in the model, showing no statistically significant interactions between gender and pronoun type, for any of the possible interactions (*ps* > 1.000).

In summary, the best-fitting model further highlights the descriptive results, which show that the participants performed similarly across the two genders and the two linguistic conditions, with no significant difference at either level, or with no significant interactions between them. These results suggest that there is no obvious difference in performance when comprehending gender in non-local contexts, when gender is independent of mutations and when it is encoded through mutations.

A secondary interest are the roles of language dominance and Welsh linguistic proficiency in the comprehension of grammatical gender in non-local contexts, gender is independent of mutations and when it is encoded through mutations. First, the overall scores for language dominance and Welsh linguistic proficiency are provided, followed by the best-fitting generalised linear mixed effects model(s).

The responses from the BLP generated a quantitative score on a continuum, ranging from -218 to +218, with a positive score indicating English dominance, while a negative score indicates Welsh dominance, and a score close to 0 (+/- 20) suggested a 'balanced bilingual'. The table below presents the overall group scores.

Table 4.13

BLP scores

BLP descriptives	Group score (n=40)	Welsh dominant (n=16)	English dominant (n=17)	Balanced (n=7)
Mean	-1.21	-55.62	50.44	-2.29
SD	53.82	29.32	19.05	16.09
Range	-118.06 – 88.27	-118.06 – -30.25	23.16 – 88.27	-18.80 – 19.63

The group mean is -1.21 (*SD*=53.82), with scores ranging from -118.06 (Welsh dominant) to 88.27 (English dominant). The overall mean would suggest that the group were balanced bilinguals, however, when taking a closer look at individual scores, few of the participants could be identified as balanced bilinguals (n=7). Based on the scores (and at the time of testing), 16 of the participants are Welsh-dominant (*M*=- 55.62, *SD*=29.32, 13 females, mean age 33 years), 17 participants are English-dominant (*M*=50.44, *SD*=29.05, 13 females, 1 non-binary individual, mean age 34 years), and 7 participants are balanced bilinguals (*M*=-2.29, *SD*=16.09, 3 females, mean age 38 years). The scores suggest that the Welsh and English dominant speakers are comparably dominant in the two languages, however, there is a larger range in scores in the Welsh dominant speakers (118.06 – -30.25) than the English dominant speakers (23.16 – 88.27). The following table presents the scores according to the four subcomponents of the BLP.

Table 4.14

BLP Sub-Sections Scores

Sub-section		Welsh	English
History	*M*	43.04	41.00
	SD	10.21	10.98
Use	*M*	22.97	30.96
	SD	15.47	15.28
Proficiency	*M*	44.38	50.90
	SD	9.81	4.98
Attitudes	*M*	49.60	35.92
	SD	6.67	10.92
TOTAL		159.99	158.78

Table 4.14 shows that the participants scored similarly in the language history section, scoring a mean of 43.04 (*SD*=10.21) for Welsh and 41.00 (*SD*=10.98) for English, with a small difference of 2.04. For the language use section, the participants score lower for the use of Welsh section (*M*=22.97, *SD*=15.47) than the use of English (*M*=30.96, *SD*=15.28). The overall proficiency score is also lower for Welsh (*M*=44.38, *SD*=9.81) than it is for English (*M*=50.90, *SD*=4.98), with a difference of 6.53. However, scores were higher for attitudes towards Welsh (*M*=49.60, *SD*=6.67) than towards English (*M*=35.92, *SD*=10.92).

The global dominance score is considered as a continuous variable to explore the possible effect of language dominance in the comprehension of grammatical gender in non-local contexts when gender is independent of mutations and when it is encoded through mutations. Therefore, each participant's global dominance score is included as a continuous predictor in the model(s). The possible effect of Welsh linguistic proficiency is also considered. The overall group scores from the Welsh cloze test are given below.

Table 4.15

Cloze Test Scores

Cloze test descriptives	**Welsh cloze scores**	
	Raw/44	%
Mean	39.9	90.63
SD	4.03	9.17
range	28 - 44	64 – 100

The results revealed that the participants performed well on the cloze test, with a mean of 90.63% (*SD*=9.17). The results also show a relatively large range of scores (64 - 100%), with four participants scoring 100%. The results suggest that the participants are highly proficient speakers of Welsh as measured by the cloze test. This is not completely unexpected as the participants who voluntarily took part likely had a special interest in the Welsh language, as they were aware it was a study investigating how Welsh speakers understand sentences. Therefore, it could be suggested that only proficient Welsh speakers were interested in participating as they were confident speakers willing to contribute to research. This high score on the cloze test contrasts the speakers' self-assessed proficiency, as measured by the BLP (Welsh: *M*=44.38, in comparison to English: *M*=50.90). Extra-linguistic factors such as confidence - specifically, the lack of - is known to impact upon regular Welsh language use and likely influences their self-perceived levels of Welsh proficiency (Hodges, 2024; Newcombe, 2002), despite scoring well on a Welsh proficiency test, which demonstrates an acceptable level of internal consistency, $\alpha = .818$.

The cloze test score is considered as a continuous variable to explore the effect of Welsh linguistic proficiency on the comprehension of grammatical gender in non-local contexts when gender is independent of mutations and when it is encoded through mutations. Therefore, each participant's score is included as a continuous predictor in the model(s).

To find the model that best explains the data with the inclusion of dominance and proficiency, the best-fitting model reported in Table 4.12 is taken as the base model, which included fixed effects of gender and pronoun type, but without an interaction between the two. Two new models are computed to better understand the possible roles of the two individual difference variables. Dominance and proficiency scores were scaled to reduce collinearity (Fernandez et al., 2021). The first model included dominance and proficiency as fixed effects and the second model included the interaction between dominance and proficiency. The models evaluated via forward model comparison revealed that the model fit improved with the interaction of dominance and proficiency ($\chi^2(1) = 0.02$, $p = 0.08$). Therefore, the most complex model justified by the data included the fixed effects of gender, pronoun type, dominance and proficiency, and the interaction between the two individual difference variables. The output from the best fitting model is below.

Table 4.16

Best-fitting model: Comprehension task with IDs

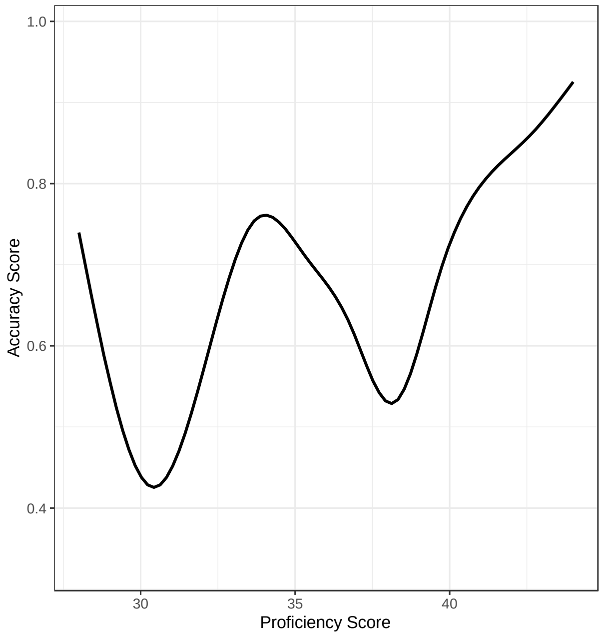
Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	1.51	0.28	5.41	**<0.001**
Gender [masculine]	-0.14	0.35	-0.40	0.69
Pronoun type [possessive: gender + mutations]	-0.09	0.37	-0.24	0.81
Dominance	0.001	0.004	0.11	0.91
Proficiency	0.33	0.07	4.59	**<0.001**
Dominance : Proficiency	-0.003	0.001	-2.90	**0.004**

Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	1.27	1.13	0.86	0.93
Gender [masculine]	0.22	0.46	-	-
Pronoun type [possessive: gender + mutations]	0.57	0.76	-	-

The model's intercept is significant (β = 1.51, SE = 0.28, z = 5.41, p = <0.001). As before, there was no effect of gender (β = -0.14, SE = 0.35, z = -0.40, p = 0.69) nor an effect of pronoun type (β = -0.09, SE = 0.37, z = -0.24, p = 0.81). Additionally, there was no effect of dominance (β = 0.001, SE = 0.004, z = 0.11, p = 0.91). However, there was an effect of proficiency (β = 0.33, SE = 0.07, z = 4.59, p = <0.001) and a significant interaction between dominance and proficiency (β = -0.003, SE = 0.001, z = -2.90, p = 0.004). Additional models were run checking the interactions between the different linguistic conditions (gender and pronoun type) and the individual difference variables (dominance and proficiency), but these models did not better explain the data than the best-fitting model presented in Table 4.16, nor were there any other significant interactions between the linguistic conditions and dominance and proficiency, different to the best fitting model presented above. The plot below shows the relationship between Welsh linguistic proficiency and the overall accuracy scores, followed by a plot for dominance.

Figure 4.7

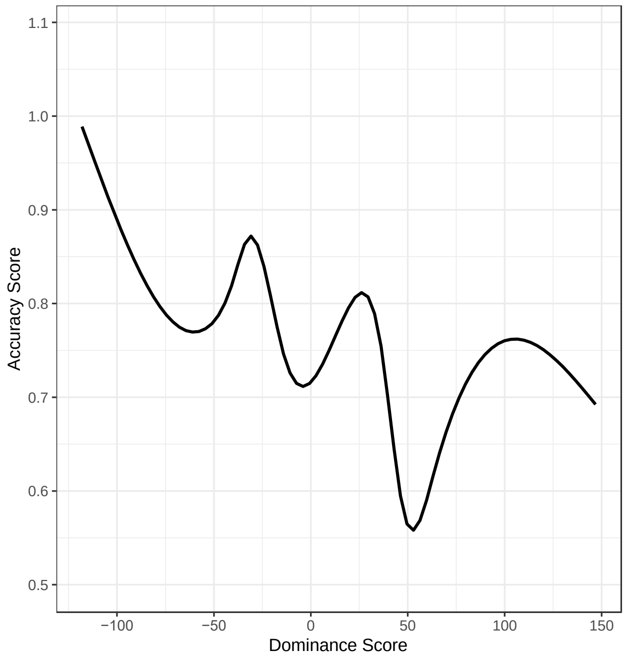
Comprehension: Welsh Linguistic Proficiency



The figure above shows the relationship between the proficiency and overall gender accuracy scores. This emerged as significant in the model (z = 4.59, p = <0.001), indicating that as proficiency scores increases, accuracy scores also increase. However, despite emerging as significant in the model, there is no obvious trend in Figure 4.7. The standard error is also larger toward lower proficiency scores (between 30-35). An additional model was run to check the interaction between proficiency, gender, and pronoun type, but there were no significant results or interactions. Therefore, no additional model is reported here. The plot below shows the relationship between language dominance and the overall accuracy scores.

Figure 4.8

Comprehension: Language Dominance



The figure above shows the relationship between dominance and overall gender accuracy scores. The plot shows a general downward trend, showing that as accuracy scores decrease, English dominance scores increase. This suggests that participants who are more Welsh dominant (below 0) had higher accuracy scores, whilst participants who are more English dominant (above 0) have lower accuracy scores. However, this did not emerge as significant in the model (z = 0.11, p = 0.91). An additional model was run to check the interaction between dominance, gender, and pronoun type, but there were no significant results or interactions. Therefore, no additional model is reported here.

The best-fitting model also revealed a significant interaction between dominance and proficiency (z = -2.90, p = 0.004). A correlation was run to assess the interaction. Given that the data are continuous, the data were not normally distributed (W = 0.947, p = <.001). Therefore, a Spearmen's correlation was performed and showed a negative correlation between the two variables (*rho* = -0.569, p = <.001). As English dominance increased, proficiency scores decreased. Inversely, more Welsh dominant participants showed increased proficiency scores.

To summarise, the results from this section are used to address the two research questions, (1) 'To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in comprehension when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?' and (1a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' comprehension of Welsh grammatical gender, when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?'.

The overall groups results indicate that the Welsh bilingual speakers show fairly accurate use of grammatical gender in comprehension when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts. However, there is individual variation within the group results, showing a wide range in scores. This indicates that some of the participants show accurate use of grammatical gender in comprehension, while other participants show less accurate use of gender when it is marked via possessive adjective forms (+mutations) and anaphoric pronouns (-mutations) in non-local contexts. The descriptive results indicated that performance was similar across the two linguistic conditions for both genders, and this was confirmed in the model. The model showed that neither gender nor pronoun type were significant and there were no significant interactions between them. These results indicate that there is no evidence to suggest that accuracy is higher for gender when it is independent of mutations than when gender is encoded through mutations, overall or for either gender specifically (masculine / feminine).

Concerning the individual difference variables, the results suggest that there is a relationship between Welsh linguistic proficiency and the overall accuracy scores, with higher proficiency participants showing higher accuracy scores. The results also revealed an interaction between proficiency and language dominance, with more Welsh dominant participants showing increased proficiency scores. However, there is no evidence to indicate that the participants' language dominance effected the overall accuracy scores, nor is there evidence to suggest any interactions between either of the individual difference variables and the different linguistic conditions (gender / pronoun type). Therefore, it is not possible to say that either variable affects gender accuracy differently when gender is independent of mutations than when gender is encoded through mutations non-local contexts[10]. These findings are discussed in light of the proposed research questions and the reviewed literature, in the discussion section of this thesis. The following section presents the second experimental measure – the elicited production task.

4.2.3 Elicited Production Task

This section reports the elicited production task results. The task assessed local gender marking of nouns via gendered numeral forms in production. Participants were required to produce either the number two, three or four in local gender contexts, following two prompt sentences. For example, the first prompt sentence stated "Dyma lun cath" (Here is a picture of a cat). The second prompt sentence stated "Nawr, mae pum' cath" (Now there are five cats). The participant produced the final sentence, who stated whether there were *two*, *three* or *four* cats. These three numerals have both masculine and feminine forms; thus, the participant produced the gendered numeral which agreed in gender with the target noun (i.e., cat).

There were 48 trials and there were two scoring systems (i.e., two sets of 48 points available). The first set of points scored for gender accuracy, where the participants were scored for producing the correct gendered numeral form (*two*, *three* or *four*). This assessed gender accuracy only, in other words, gender-marked numerals are not involved with the mutation system in these instances. Scoring the participants in this way informs us of their use of gender when it operates without mutations and also, this first scoring system serves as a basis for the second scoring system. The second set of points scored for gender-mutation accuracy, where the participants were scored for producing the mutated post-numeral noun correctly or for producing the post-numeral noun in its bare form correctly. However, participants were scored for +/- mutation accuracy production, based on the correct responses from the first scoring system. This ensured that the participants +/- mutation accuracy was based on the items they accurately assigned the gendered numeral to the target noun. The results from the first scoring system (gender

[10] A table with each individual's BLP and Welsh proficiency scores, per condition in this comprehension task, can be found in Appendix F.

accuracy) are reported, followed by the results from the second scoring system (gender-mutation accuracy).

Scoring System 1: Gender Accuracy

Participant trials were excluded from the data set if they incorrectly translated the target noun in the productive vocabulary task. This was to ensure that the participants' production of grammatical gender was only being measured for the target nouns that they were familiar with. In total, 16 trials were excluded from the whole data set, where participants received 1 point for producing the correct gendered numeral form, and this was replaced with 0 for incorrectly translating that target noun. These items were excluded from eight participants and the excluded trials ranged from 1-5 items per participant. The results after removing these 16 trials are presented below. The overall group score is presented below.

Table 4.17

Elicited Production Task: Group Descriptive Statistics

Group descriptives	Raw	Percentage
Mean	35.25	73.44
SD	7.36	15.33
Range	24 – 48	50 – 100

Table 4.17 shows that the overall group mean is 73.44% (*SD*=15.33%), where the participants scored on average 35 out of 48 (*SD*=7.36). This suggests that as a group, they show fairly accurate use of grammatical gender in production in local contexts. However, the standard deviation shows that there is individual variation within the group results and that there is a wide range in scores (50 – 100%). This suggests that some of the participants have productive command of Welsh grammatical gender in local contexts, when gender is marked via gendered numerals, while others do not.

To better understand the results, the table below shows the scores according to the different gendered numerals, specifically, the participants' accuracy according to masculine and feminine forms of the numerals two, three and four.

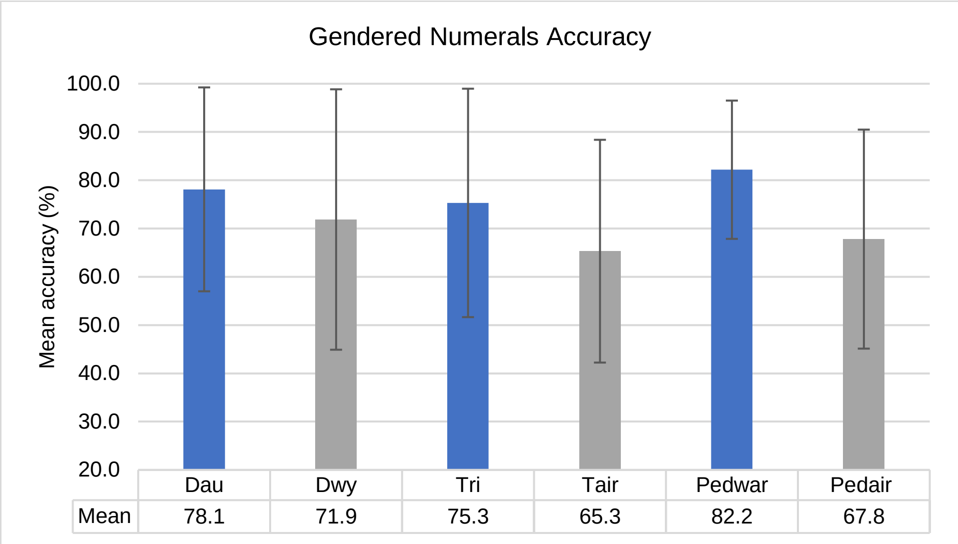
Table 4.18

Elicited Production Task: Group Descriptive Statistics by Gendered Numerals

Gender	Dau [m]		Dwy [f]		Tri [m]		Tair [f]		Pedwar [m]		Pedair [f]	
accuracy	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	6.25	78.1	5.75	71.9	6.03	75.31	5.23	65.31	6.58	82.2	5.43	70.3
SD	1.69	21.12	2.16	26.98	1.89	23.66	1.85	24.08	1.15	15.32	1.82	21.34
Range	0-6	0-100	0-6	0-100	0-6	0-100	0-6	0-100	3-6	38-100	0-6	0-100

Figure 4.9

Elicited Production Task: Group Descriptive Statistics Gendered Numerals



The table and figure above show that the participants performed best on the masculine number four ‘pedwar’ (*M*=82.2, *SD*=15.32), while accuracy was lowest on the feminine number three ‘tair’ (*M*=65.3, *SD*=24.08). There is a difference of 6.2% in the performance between the masculine [dau] and feminine [dwy] gendered numeral two forms, with participants scoring an average of 78% for ‘dau’ and 72% for ‘dwy’. There is a difference of 8% in the performance between the masculine [tri] and feminine [tair] gendered numeral three forms, with participants scoring an average of 75% for ‘tri’ and 65% for ‘tair’. There is a larger difference of 14% in the performance between the masculine [pedwar] and feminine [pedair] gendered numeral four forms, with participants scoring an average of 82% for ‘pedwar’ and 68% for ‘pedair’. The figure also shows that the participants performed best on the masculine numerals in comparison to their feminine counterparts. This suggests that the participants experienced more difficulty with assigning feminine gender when naming numerals for the target nouns, in comparison to masculine nouns.

To understand how the participants responded to the items in the production task, the best-fitting mixed effects model is found and reported. Gender (masculine, feminine) is included as a factor in the model, as well as gendered numeral form (two / three / four). To check the possible relationship between gender and numeral forms, model one included the fixed effects of gender and gendered numeral form (no interaction), and the second model included the interaction of gender and gendered numeral form. This allowed me to test whether including the interaction better explained the data. For the random effects structure, there was a by-participant random intercept as well as by-participant random slopes for gender and gendered numeral form, as well as associated correlation parameters. There was also a by-item random intercept. Deviation coding was set for gender and gendered numeral form (Schad et al., 2020).

The models evaluated via forward model comparison revealed that the model fit did not improve with the interaction of gender and gender numeral form ($\chi^2(2) = 0.21$, $p = 0.90$), but with them only as main effects. Therefore, the most complex model justified by the data included the fixed effects of gender and gendered numeral form, but without the interaction between the two factors. The output from the best fitting model is below.

Table 4.19

Best-fitting Model: Elicited Production Task Gender Accuracy

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	1.69	0.22	7.56	**<0.001**
Gender [masculine]	0.96	0.44	2.19	**0.03**
Gendered numeral form [number 2]	0.31	0.20	1.58	0.11
Gendered numeral form [number 3]	-0.34	0.20	-1.75	0.08
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	1.19	1.09	0.69	0.83
Gender [masculine]	4.45	2.11	-	-
Gendered numeral form [number 2]	0.05	0.21	-	-
Gendered numeral form [number 3]	-0.34	0.23	-	-

The model's intercept is significant (β = 1.69, SE = 0.22, z = 7.56, p = <0.001). There was an effect of gender (β = 0.96, SE = 0.44, z = 2.19, p = 0.03), with masculine items having a higher score than feminine items. However, there was no effect of gendered numeral form (β = 0.31, SE = 0.20, z = 1.58, p = 0.11; β = -0.34, SE = 0.20, z = -1.75, p = 0.08, respectively). To check the interactions between gender and gendered numeral form, pairwise comparisons were run using *emmeans* (see Lenth et al., 2022) with the adjust argument specifying 'Bonferroni' correction. The results revealed that there were no statistically significant interactions between gender and gendered numeral form, for any of the interactions (*ps* > 0.055). The only comparison approaching significance was masculine numeral two and the feminine numeral three (β = 1.61, SE = 0.55, z = 2.91, p = 0.055).

Collectively, the overall results show that the Welsh adult speakers have a good productive command of grammatical gender when gender is marked via gendered numerals, however, there is a great deal of individual variation in the data[11]. The model revealed that participants performed better on masculine items than feminine items, but there was no statistically significant difference in

[11] A table with each individual's BLP and Welsh proficiency scores, per gendered numeral in this production task, can be found in Appendix G.

performance between the numeral forms, nor any significant differences between the different forms of the two genders.

A secondary interest to the production of grammatical gender is the role of Welsh linguistic proficiency and language dominance. Two additional models with the inclusion of dominance and proficiency were run. The best-fitting model reported in Table 4.19 is taken as the base model, which included the fixed effects of gender and gendered numeral form, but without the interaction between the two factors and the addition of the two individual variables. Dominance and proficiency scores were scaled to reduce collinearity (Fernandez et al., 2021). The first model included dominance and proficiency as fixed effects and the second model included the interaction between dominance and proficiency.

The models evaluated via forward model comparison revealed that the model fit improved with the interaction of dominance and proficiency ($\chi^2(1) = 7.54$, $p = 0.006$). Therefore, the most complex model justified by the data included the fixed effects of gender, gendered numeral form (no interaction), dominance and proficiency, and the interaction between the two individual difference variables. The output from the best fitting model is below.

Table 4.20

Best-fitting Model: Elicited Production Task Gender Accuracy with IDs

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	1.38	0.20	6.79	**<0.001**
Gender [masculine]	0.97	0.44	2.23	**0.03**
Gendered numeral form [number 2]	0.23	0.19	1.19	0.24
Gendered numeral form [number 3]	-0.29	0.19	-1.52	0.13
Dominance	0.001	0.001	0.11	0.92
Proficiency	0.24	0.05	4.54	**<0.001**
Dominance : Proficiency	-0.002	0.001	-3.04	**0.002**

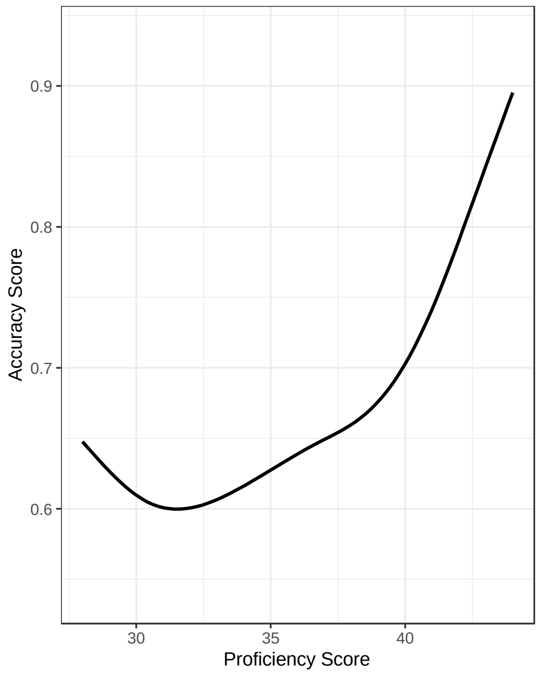
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	0.58	0.76	0.68	0.83
Gender [masculine]	4.37	2.09	-	-
Gendered numeral form [number 2]	0.02	0.07	-	-
Gendered numeral form [number 3]	0.04	0.21	-	-

The model's intercept is significant (β = 1.38, SE = 0.20, z = 6.79, p = <0.001). There is an effect of gender (β = 0.97, SE = 0.44, z = 2.23, p = 0.03), with masculine items having a higher score than feminine items. There was also an effect of proficiency (β = 0.24, SE = 0.05, z = 4.54, p = <0.001) and a significant interaction between dominance and proficiency (β = -0.002, SE = 0.001, z = -3.04, p = 0.002). However, there is no effect of gendered numeral form (β = 0.23, SE = 0.19, z = 1.19, p = 0.24; β = -0.29, SE = 0.19, z = -1.52, p = 0.13, respectively) and there was no effect of dominance (β = 0.001, SE = 0.001, z = 0.11, p = 0.92).

Additional models were run checking the interactions between the different linguistic conditions (gender and numeral form) and the individual difference variables (dominance and proficiency). None of these models better explained the data than the best-fitting model in Table 4.20, nor were there any other significant interactions between the linguistic conditions and dominance and proficiency, different to those presented in Table ***4.20***. The plot below shows the relationship between Welsh linguistic proficiency and the overall accuracy scores, followed by a plot for dominance.

Figure 4.10

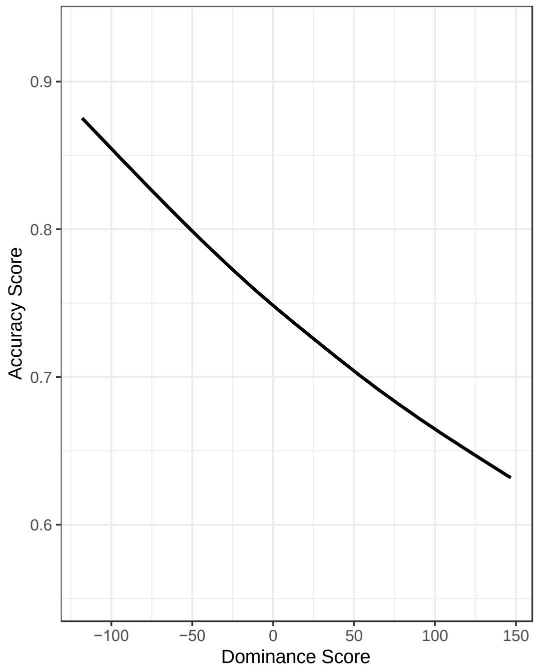
Production: Gender Accuracy and Proficiency



The figure above shows the relationship between the proficiency and overall gender accuracy scores. Figure 4.10 shows a general upward trend from approximately the scores of 31 / 32, as there is an initial drop in the scores from 28 / 29 to 30, before rising again upward to 44. This emerged as significant in the model (z = 4.54, p = <0.001), indicating that as proficiency scores increases, accuracy scores also increase. The second individual variable of interest was language dominance. The plot below shows the relationship between language dominance and the overall accuracy scores.

Figure 4.11

Production: Gender Accuracy and Dominance



The figure above shows the relationship between the dominance and overall gender accuracy scores. The plot shows a general downward trend, showing that as accuracy scores decrease, English dominance scores increase. This suggests that participants who are more Welsh dominant (below 0) had higher accuracy scores, whilst participants who are more English dominant (above 0) have lower accuracy scores. However, this did not emerge as significant in the model (z = 0.11, p = 0.92).

The best-fitting model also revealed a significant interaction between dominance and proficiency (z = -3.04, p = 0.002). A correlation was run to assess the interaction. Given that the data are continuous, the data were not normally distributed (W = 0.947, p = <.001). Therefore, a Spearmen's correlation was performed and showed a negative correlation between the two variables (*rho* = -0.569, p = <.001). As English dominance increased, proficiency scores decreased. Inversely, more Welsh dominant participants showed increased proficiency scores.

To summarise, the results suggest that there is a relationship between Welsh linguistic proficiency and the overall accuracy scores, with higher proficiency participants showing higher accuracy scores. The results also revealed an interaction between proficiency and language dominance, with more Welsh dominant participants showing increased proficiency scores. However, there is no evidence to indicate that the participants' language dominance effected the overall accuracy scores, nor is there evidence to suggest any interactions between the individual difference variables (dominance and proficiency) and the different linguistic conditions (gender / numeral forms). Therefore, it is not possible to say that either variable affects gender accuracy differently when gender is marked via these gendered numeral forms.

This first scoring system (gender only) sheds light on the production of gender when it is marked via gendered numeral forms, in a pure gender agreement process. However, there is no involvement with the mutation system. The results for the second scoring system are presented below which will also be used to address the research question regarding the production data. These results will provide an understanding of how gender operates independently of mutations and in conjunction with mutations, when gender is marked via gendered numerals in local contexts.

Scoring System 2: Mutation / No-mutation Accuracy

For this second scoring system, the participants were scored out of 48 (16 trials targeting the numeral 2, 3 and 4). The participant was awarded 1 point for correctly producing the noun in either its bare form, following '*tair'* 3 feminine, '*pedair*' 4 feminine and '*pedwar*' 4 masculine, or 1 point for correctly mutating the noun, following '*dau'* 2 masculine, *'dwy'* 2 feminine and '*tri'* 3 masculine. The participants were scored based on the correct responses from the first scoring system. In other words, if the participant scored zero for incorrectly assigning the gender of the target noun, it was not possible to receive a score of 1 for correctly mutating the noun or producing the noun in bare form, in this secondary analysis (e.g., producing 'dau **g**ath' in place of 'dwy gath', despite getting the soft mutation correct on the noun: cath > **g**ath). These items were removed from analysis. However, prior to removing these items, Table 4.21 shows how the Welsh bilinguals produced the target nouns following the gendered numerals, according to correct mutation, incorrect mutation, and bare form production, for the complete dataset.

Table 4.21

Elicited Production Task: Accuracy Score Breakdown

Mutation accuracy	Correct mutation	Incorrect mutation	No mutation
2 masculine	n=231	n=2 (double mutation)	n=87
(SM)	*M*=72.2% *SD*=32.9%	*M*=0.63% *SD*=2.8%	*M*=27.2% *SD*=33.4%
2 feminine	n=267	n=1 (AM)	n=52
(SM)	*M*=83.3% *SD*=26.0%	*M*=0.31% *SD*=2.0%	*M*=16.3% *SD*=26.1%
3 masculine	n=136	n=0	n=184
(AM)	*M*=42.5% *SD*=25.6%		*M*=57.5% *SD*=25.6%
3 feminine	N/A	n=11 (AM) *M*=3.4%, *SD*=6.9%	n=283
(no mutation)		n=26 (SM) *M*=8.1%, *SD*=12.5%	*M*=88.4% *SD*=17.1%
4 masculine	N/A	n=16 (SM)	n=304
(no mutation)		*M*=5.0% *SD*=12.3%	*M*=95.0% *SD*=12.3%
4 feminine	N/A	n=47 (SM)	n=273
(no mutation)		*M*=14.7% *SD*=24.2%	*M*=85.3% *SD*=24.2%

Table 4.21 shows that the participants were most accurate when producing the noun in bare form following the masculine numeral *four* (*M*=95.0% *SD*=12.3%), with very few instances of SM produced on the target noun (*M*=5.0% *SD*=12.3%). They performed next best when producing the noun in bare form following the feminine numeral *three* (*M*=88.4% *SD*=17.1%), however, some nouns were SM (*M*=8.1%, *SD*=12.5%) while others were AM (*M*=3.4%, *SD*=6.9%). Generally, participants were also accurate in producing the noun in bare form following the feminine numeral *four* (*M*=85.3% *SD*=24.2%), with some instances of SM produced on the target noun (*M*=14.7% *SD*=24.2%). The participants also performed well on SM nouns following the feminine numeral *two* (*M*=83.3% *SD*=26.0%), however, there were 52 instances where the nouns following the feminine numeral 2 were produced in bare form, while one was produced with AM.

The participants were less accurate in producing SM nouns following the masculine numeral *two* (*M*=72.2% *SD*=32.9%). There were two instances where two participants over-mutated the same masculine noun following the numeral 2, which suggests that they assumed that the base form /t/ was the SM form /d/ and then SM again /d/ > /dd/ instead of /t/ > /d/. For the masculine numeral 2, besides these two instances, participants produced the noun in its bare form 87 times and there were no instances of AM. The participants performed poorest on nouns following the masculine numeral *three* (*M*=42.5% *SD*=25.6%). The masculine numeral three triggers AM on following nouns starting with *p*, *t* and *c*. However, the majority of the nouns (*M*=57.5%, *SD*=25.6%) were produced in their bare form, with no mutation at all. This may be because AM after the numeral 3 [m] is less common than other instances of AM in the Welsh language. Notably, one participant only produced the masculine numeral *three* form ‘tri’ for all trials showing three items, therefore scoring 100% accuracy for the production of ‘tri’ and 0% for the production of ‘tair’. Whereas one other participant only produced the feminine numeral *three* form ‘tair’ for all trials showing three items, therefore scoring 100% accuracy for the production of ‘tair’ and 0% for the production of ‘tri’. This is interesting given the fact that there is a clear distinction between what they need to do with masculine nouns and feminine nouns following the numeral *three*.

Collectively, these results show that the participants were most accurate when there is no mutation involved via the masculine numeral *four*, then when mutation is involved, SM following the feminine numeral *two* is the most consistent. The participants are least accurate when producing AM after the masculine numeral *three*.

The information presented in Table 4.21 shows how the participants produced the nouns following the gendered numeral forms, however, these results are based on a full data set. Recall that the second scoring system scored the participants for no mutation and mutation accuracy production, based on the correct responses from the first scoring system. Therefore, 345 trials were removed from the figures presented in Table 4.21. The accuracy results according to gender and the different gendered numeral forms is presented below.

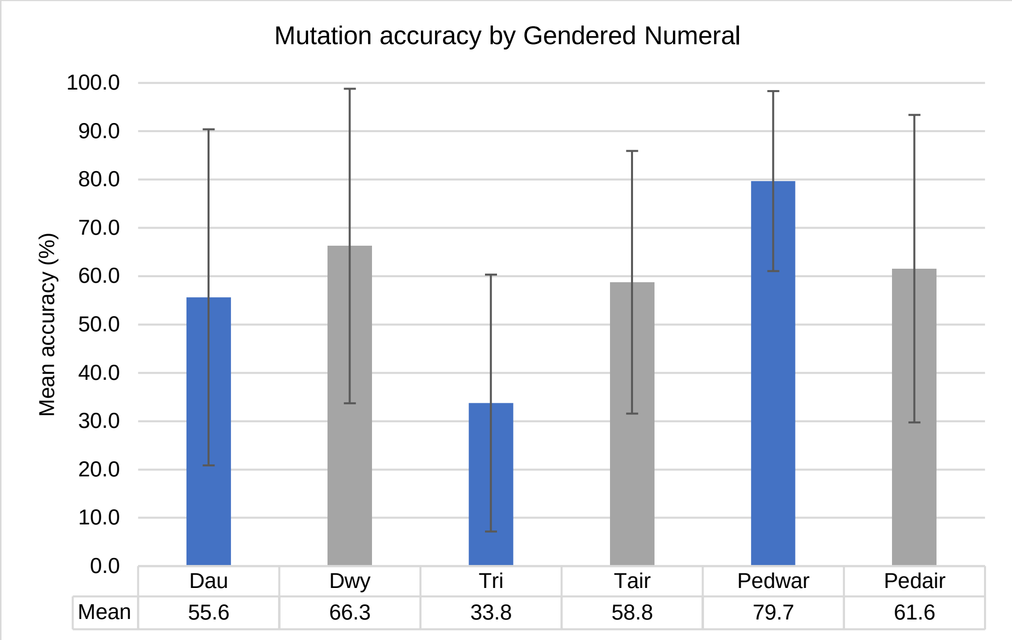
Table 4.22

Elicited Production Task: Group Descriptive Statistics Mutation accuracy

Mutation accuracy by gendered numeral	**Dau [m] (SM)**		**Dwy [f] (SM)**		**Tri [m] (AM)**		**Tair [f] (no mut)**		**Pedwar [m] (no mut)**		**Pedair [f] (no mut)**	
	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%	Raw	%
Mean	4.5	55.6	5.3	66.3	2.7	33.8	4.7	58.8	6.4	79.7	4.9	61.6
SD	2.8	34.8	2.6	32.6	2.1	26.6	2.2	27.2	1.5	18.6	2.55	31.8
Range	0-8	0-100	0-8	0-100	0-7	0-88	0-8	0-100	2-8	25-100	0-8	0-100

Figure 4.12

Elicited Production Task: Group Descriptive Statistics Mutation accuracy



The table and figure above show that the participants most accurately produced nouns in their bare form following the masculine numeral *four* [*pedwar*] (*M*=79.7%, *SD*=18.6%) and performance was next best when producing SM nouns following the feminine numeral *two* [*dwy*] (*M*=66.3%, *SD*=32.6%). The participants also performed relatively well when producing nouns in their bare form following the feminine numeral *four* [*pedair*] (*M*=61.6%, *SD*=31.8%), and performance was next best when producing nouns in their bare form following the feminine numeral *three* [*tair*] (*M*=58.8%, *SD*=27.2%), followed by producing SM nouns following the masculine numeral *two* [*dau*] (*M*=55.6%, *SD*=34.8%). Performance was weakest when producing AM nouns following the masculine numeral *three* [*tri*] (*M*=33.8%, *SD*=26.6%). This indicates that accuracy is best when masculine nouns are produced in their bare form and when feminine nouns are SM.

To understand how the participants responded to the items in the production task, for this second scoring system (gender-mutation accuracy), the best-fitting mixed effects model is found and reported. Similar to the generalised linear mixed effects model reported as part of the first scoring system, gender (masculine, feminine) is included as a factor in the model, as well as gendered numeral form (two, three and four). To check the possible relationship between gender and numeral forms, model one included the fixed effects of gender and gendered numeral form (no interaction), and the second model included the interaction of gender and gendered numeral form. This allowed me to test whether including the interaction better explained the data. For the random effects structure, there was a by-participant random intercept as well as by-participant random slopes for gender and gendered numeral form, as well as associated correlation parameters. There was also a by-item random intercept. Deviation coding was set for gender and gendered numeral form (Schad et al., 2020).

The models evaluated via forward model comparison revealed that the model fit improved with the interaction of gender and gender numeral form ($\chi^2(2) = 16.06$, $p = <0.001$). Therefore, the most complex model justified by the data included the fixed effects of gender and gendered numeral form, and the interaction between the two factors. The output from the best fitting model is below.

Table 4.23

Best-fitting Model: Elicited Production Task Mutation Accuracy

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	0.69	0.27	2.60	**0.009**
Gender [masculine]	-0.35	0.34	-1.05	0.30
Gendered numeral form [number 2]	0.19	0.22	0.87	0.39
Gendered numeral form [number 3]	-0.92	0.21	-4.36	**<0.001**
Gender [M] : Numeral [2]	-0.46	0.38	-1.21	0.23
Gender [M] : Numeral [3]	-1.15	0.38	-3.05	**0.002**
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	2.06	1.43	0.68	0.83
Gender [masculine]	1.61	1.27	-	-
Gendered numeral form [number 2]	0.54	0.74	-	-
Gendered numeral form [number 3]	0.34	0.58	-	-

The model's intercept is significant ($\beta = 0.69$, SE = 0.27, $z = 2.60$, $p = 0.009$). There was an effect of gendered numeral form ($\beta = -0.92$, SE = 0.21, $z = -4.36$, $p = <0.001$), with numeral *three* items (masculine nouns AM & feminine nouns not mutated) having lower accuracy than numeral *four* items (masculine and feminine nouns produced in bare form). There was also a significant interaction for masculine forms and numeral *three* forms (both masculine & feminine nouns) (β = -1.15, SE = 0.38, z = -3.05, p = 0.002). However, there was no effect of gender (β = -0.35, SE = 0.34, z = -1.05, p = 0.30) or gendered numeral form for numeral *two* items (masculine or feminine) (β = 0.19, SE = 0.22, z = 0.87, p = 0.39).

Pairwise comparisons were run using *emmeans* with the adjust argument specifying ‘Bonferroni’ correction. The results revealed that the numeral *two* condition had higher accuracy than the numeral *three* condition items (β = 1.11, SE = 0.38, z = 2.9, p = 0.01). However, there was no statistically significant difference in accuracy scores between numeral *two* and numeral *four* (β = -0.53, SE = 0.37, z = -1.42, p = 0.47). Additional pairwise comparisons were run to check the interactions between gender and gendered numeral forms and the results revealed six statistically significant interactions between gender and the gendered numeral forms. These are shown in the table below.

Table 4.24

Pairwise Comparisons – Elicited Production Task

Comparison	**Estimate**	**SE**	***z***	***p***
Masculine ‘2’ – Masculine ‘4’	-1.56	0.50	-3.13	0.03
Feminine ‘2’ – Masculine ‘3’	2.27	0.59	3.83	0.002
Masculine ‘3’– Feminine ‘3’	-1.50	0.50	-3.00	0.04
Masculine ‘3’ – Masculine ‘4’	-3.02	0.49	-6.16	<.0001
Masculine ‘3’ – Feminine ‘4’	-1.77	0.53	-3.30	0.01
Feminine ‘3’ – Masculine ‘4’	-1.52	0.51	-3.01	0.04

The pairwise comparisons showed that accuracy was lower for SM nouns following *two* [m] than nouns produced in bare form following *four* [m] (p = 0.03), higher for SM nouns following *two* [f] than AM nouns following *three* [m] (p = 0.002), lower for AM nouns following *three* [m] than nouns produced in bare form following *three* [f] (p = 0.04), lower for AM nouns following *three* [m] than nouns produced in bare form following *four* [m] (p = <.0001), lower for AM nouns following *three* [m] than nouns produced in bare form following *four* [f] (p = 0.01), and lower for nouns produced in bare form following *three* [f] than nouns produced in bare form following *four* [m] (p = 0.04). Four of the six comparisons revealed that accuracy is better when nouns are produced in their bare form compared to when they are mutated, while one comparison showed that mutation accuracy is higher when it is SM [f] compared to AM [m], and the other showed that no-mutation accuracy is higher for masculine nouns (4 [m]) compared to feminine nouns (3 [f]). This indicates that accuracy is highest when nouns are not mutated in comparison to when they are mutated, and when nouns are mutated, accuracy is higher for SM nouns [f] in comparison to AM nouns [m].

A secondary interest are the roles of language dominance and Welsh linguistic proficiency in the production of grammatical gender in local contexts when gender is independent of mutations and in conjunction with mutations. Recall that descriptive results for dominance and proficiency were presented previously as part of the comprehension task results (see Table 4.13, Table 4.14 and Table 4.15). Each participant's dominance and proficiency scores are included as continuous predictors in this model(s).

To find the model that best explains the data with the inclusion of dominance and proficiency, the best-fitting model reported in Table 4.23 is taken as the base model, which included the fixed effects of gender and gendered numeral form, and the interaction between the two factors, for the gender-mutation accuracy scores. Four new models were computed to better understand the possible roles of the two individual difference variables. Dominance and proficiency scores were scaled to reduce collinearity (Fernandez et al., 2021). The first model included dominance and proficiency as fixed effects, the second included in the interaction between dominance and proficiency, the third included the interaction between gendered numeral form and dominance, and the inclusion of proficiency as a fixed effect, and the final model included the interactions between gender, gendered numeral form, dominance, and proficiency.

The models evaluated via forward model comparison revealed that the most complex model justified by the data included the fixed effects of gender and gendered numeral form (and their interaction), with the addition of both dominance and proficiency as fixed effects, as well as an interaction between the two variables, but no interaction between the variables and gender and gendered numeral form (χ^2 (1) = 8.42, p = 0.004). The output from the best fitting model is below.

Table 4.25

Best-fitting model: Elicited Production Task with IDs

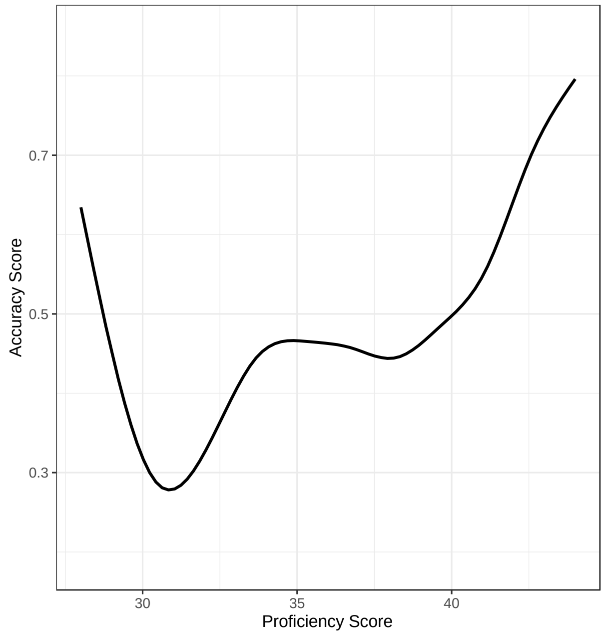
Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	0.31	0.23	1.28	0.20
Gender [masculine]	-0.33	0.34	-0.97	0.33
Gendered numeral form [number 2]	0.14	0.22	0.63	0.53
Gendered numeral form [number 3]	-0.90	0.21	-4.26	**<0.001**
Dominance	0.002	0.004	0.49	0.62
Proficiency	0.29	0.06	4.73	**<0.001**
Gender [M] : Numeral [2]	-0.42	0.38	-1.11	0.27
Gender [M] : Numeral [3]	-1.21	0.38	-3.16	**0.002**
Dominance : Proficiency	-0.003	0.001	-3.21	**0.001**
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	0.97	0.99	0.70	0.84
Gender [masculine]	1.60	1.26	-	-
Gendered numeral form [number 2]	0.46	0.68	-	-
Gendered numeral form [number 3]	0.31	0.55	-	-

The results from the best-fitting model showed that similarly to the model without the individual variables, there was an effect of gendered numeral form (β = -0.90, SE = 0.21, z = -4.26, p = <0.001), with numeral *three* items (masculine nouns AM & feminine nouns not mutated) having lower accuracy than numeral *four* items (masculine and feminine nouns produced in bare form). There was also a significant interaction for masculine forms and numeral *three* forms (both masculine & feminine nouns) (β = -1.21, SE = 0.38, z = -3.16, p = 0.002). Additionally, the output shows there was an effect of proficiency (β = 0.29, SE = 0.06, z = 4.73, p = <0.001) and a significant interaction between dominance and proficiency (β = -0.003, SE = 0.001, z = -3.21, p = 0.001). However, the model's intercept was not significant (β = 0.31 SE = 0.23, z = 1.28, p = 0.20). There was no effect of gender (β = -0.33, SE = 0.34, z = -0.97, p = 0.33) nor an effect of gendered numeral form for numeral *two* items (β = 0.14, SE = 0.22, z = 0.63, p = 0.53). There was also no effect of dominance (β = 0.002, SE = 0.004, z = 0.49, p = 0.62).

Of the two individual variables, one emerged as significant in the model, which was Welsh proficiency. The plot below shows the relationship between Welsh linguistic proficiency and the overall accuracy scores.

Figure 4.13

Elicited Production: Welsh Linguistic Proficiency

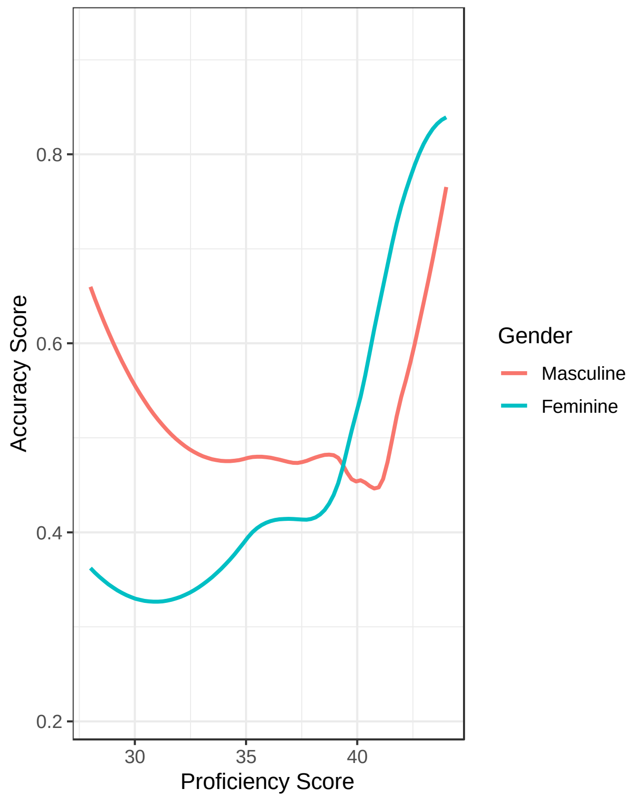


The figure above shows the relationship between proficiency and overall accuracy scores. This emerged as significant in the model (z = 4.73, p = <0.001). The plot shows a general upward trend from participants who scored approximately 32 and above on the cloze test. There is a clear dip in the plot, where participants who scored 30-31 on the cloze test have a noticeably lower accuracy score. The standard error is also larger for participants who scored between 30 – 35.

A separate model was run including the effect of proficiency only and the interaction between proficiency, gender-mutation and gendered numeral form. The model revealed a significant interaction was between gender-mutation and proficiency (β = -0.16, SE = 0.05, z = -3.12, p = 0.002), with higher proficiency speakers showing increased accuracy for feminine nouns in comparison to masculine nouns. This is shown in the plot below.

Figure 4.14

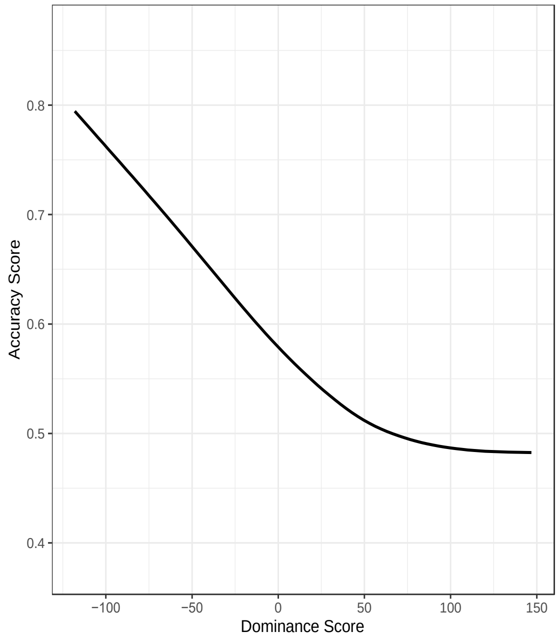
Elicited Production: Welsh Linguistic Proficiency By Gender



The plot shows that participants with lower proficiency scores are more accurate on masculine nouns, however, as proficiency increases, accuracy is higher for feminine nouns. There were no other statistically significant interactions in the additional model, between proficiency, gender, and gendered numeral form (*p*s > 0.09). The second individual variable of interest was language dominance. The plot below shows the relationship between language dominance and the overall accuracy scores.

Figure 4.15

Elicited Production: Language Dominance



The figure above shows the relationship between the dominance and overall accuracy scores. This did not emerge as significant in the model (z = 0.49, p = 0.62). The plot shows a general downward trend, showing that as accuracy scores decrease, dominance scores increase. This suggests that participants who are more Welsh dominant had higher accuracy scores, whilst participants who are more English dominant have lower accuracy scores.

The best-fitting model with the Welsh proficiency and language dominance also revealed a significant interaction between dominance and proficiency (β = -0.003, SE = 0.001, z = -3.21, p = 0.001). A correlation was run to assess the interaction. Given that the data are continuous, the data were not normally distributed (W = 0.947, p = <.001). Therefore, a Spearmen's correlation was performed and as the comprehension task data, the correlation showed a negative correlation between the two variables (*rho* = -0.569, p = <.001). As English dominance increased, proficiency scores decreased. Inversely, more Welsh dominant participants showed increased proficiency scores.

To summarise, the results from this section are used to address the two research questions, (2) 'To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in production when gender is independent of mutations and in conjunction with mutations, in local contexts?' and (2a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' production of Welsh grammatical gender, when gender is independent of mutations and in conjunction with mutations, in local contexts?'. The results from the first scoring system (gender accuracy) have shown that overall, the Welsh speakers have a good productive command of grammatical gender when gender is marked via gendered numerals, however, there is notable individual variation in the data[12]. The model revealed that participants performed better on masculine items than feminine items, but there was no statistically significant difference in performance between numeral forms, nor any significant differences between the different numeral forms of the two genders. The results from the second scoring system (gender-mutation accuracy), revealed no effect of gender, but several significant pairwise comparisons. Specifically, the results revealed that accuracy is better when nouns are produced in bare form compared to when they are mutated. The pairwise comparisons also showed that no-mutation accuracy is higher for masculine nouns (4) compared to feminine nouns (3 & 4). Also, mutation accuracy is higher when it is SM (2 [f]) compared to AM (3 [m]). Collectively, this suggests that accuracy is highest when nouns are produced in bare form in comparison to when they are mutated, and when nouns are mutated, accuracy is higher for SM nouns (feminine only) in comparison to AM nouns (masculine).

Concerning the individual difference variables, the results suggest that there is a relationship between Welsh linguistic proficiency and the overall accuracy scores. Further analysis revealed an interaction between gender and proficiency, with lower proficiency participants scoring more accurately on masculine nouns, while higher proficiency participants scored more accurately on feminine nouns. The results also revealed an interaction between proficiency and language dominance, with Welsh dominant participants showing increased proficiency scores. However, similar to the comprehension data, there is no evidence to indicate that the participants' language dominance effects the overall accuracy scores. Therefore, it is not possible to say that dominance effects gender-mutation accuracy differently when gender is independent of mutations and in conjunction with mutations, in these

[12] A table with each individual's BLP and Welsh proficiency scores, per gendered numeral in this production task, can be found in Appendix G.

local contexts. These findings are discussed in light of the proposed research questions and the reviewed literature, in the discussion section of this thesis.

4.3 Discussion

The aim of this first experiment was to investigate the comprehension (non-local contexts) and production (local contexts) of grammatical gender in Welsh-English adult bilinguals, in a more geographically diverse Welsh adult population than those previously tested. I set out to investigate whether the variable use of gender and mutations in Welsh adult speakers is due to a lack of gender knowledge or whether it is an indication of the adults' lack of ability and consistency in comprehending / producing gender when it is involved with mutations. A final goal was to investigate the possible roles of two individual difference variables, namely, language dominance and Welsh linguistic proficiency, in understanding whether these factors may predict higher performance on measures of grammatical gender in Welsh-English bilingual adults.

When considering the findings from the two experimental tasks together, it is possible to discuss the results in three parts. First, the results are discussed in light of the *Interpretability Hypothesis*. Second, the results are discussed in light of the *Missing Surface Inflection Hypothesis*. Third, the investigation of individual difference variables is discussed in light of previous findings. However, before discussing the findings in light of the predictions made and the two theoretical approaches reviewed in chapter 3, the following section summarises the results and compares the present findings to those from previous studies exploring the comprehension and production of grammatical gender in adult Welsh speakers. This is done prior to discussing the results in light of the two theoretical approaches, as the previous Welsh gender findings have not yet been discussed under theoretical frameworks of second language acquisition. Therefore, direct comparisons to previous studies are first made to contextualise and situate the findings within the existing body of Welsh gender literature. The individual difference variables are not considered in the following section, but are discussed in a later section, given that this is the first time they have been examined in relation to Welsh adults' comprehension and production of grammatical gender.

4.3.1 Results Summary and Welsh data comparison

The first experimental task assessed the comprehension of grammatical gender in non-local contexts, where gender was marked via possessive adjective forms (gender encoded through mutations) and anaphoric pronouns (gender independent of mutations). The results revealed an overall group mean of 76.46% (*SD*=18.51%). When the results were divided according to linguistic conditions and gender (masculine, feminine), the results showed that when gender agreement is via the third person possessive adjective 'ei', with AM on immediately following nouns signalling feminine gender and SM signalling masculine gender to agree with the antecedent noun (i.e., gender encoded through mutations), overall accuracy is quite high for both genders (AM [f], *M*=76.50, *SD*=22.82; SM [m], *M*=77.25, *SD*=21.84). Similarly, the results showed that when gender agreement is via third person singular anaphoric pronouns ('hi' [f], 'fe' [m]) (i.e., no involvement with mutations), overall accuracy is also quite high for both genders ('hi' [f], *M*=76.56, *SD*=22.85; 'fe' [m], *M*=75.63, *SD*=23.33). Overall, performance is also similar between the two linguistic conditions and notably, performance is similar for masculine and feminine items, both within and across the two linguistic conditions. These findings were further supported in the best-fitting model, which showed no significant effect of gender (masculine, feminine) ($p = 0.28$) and no significant effect of pronoun type (i.e., 'fe/hi' vs 'ei' ($p = 0.85$). Further interactions were checked but there were no statistically significant interactions between gender and pronoun type, for any of the possible interactions (*ps* > 1.000).

Taken together, the results indicate that overall, the participants show accurate use of grammatical gender in comprehension, while other participants show less accurate use of gender when it is marked via possessive adjective forms (gender encoded through mutations) and anaphoric pronouns (gender independent of mutations). The results also suggest that there is no statistically significant difference in performance when comprehending gender in non-local contexts, when gender is independent of mutations ('fe/hi') and when gender is encoded through mutations ('ei' +SM [m] / AM [f]), and that performance is similar for both genders (within and across both linguistic conditions).

The overall comprehension task group result can be compared to those reported by Binks (2017). In their study, 2L1 bilinguals scored an average of 77.41%

(*SE*=5.21), L1 bilinguals scored an average of 85.23% (*SE*=2.08%) and L2 bilinguals scored an average of 85% (*SE*=5.77%). The current findings are most comparable to the 2L1 bilinguals in Binks’ study. This similarity is interesting given the differences between the Welsh speakers. For instance, the bilinguals in Binks' (2017) study were categorised according to the language(s) spoken at home. The L1 bilinguals only had Welsh at home, the 2L1 bilinguals had both Welsh and English at home, while the L2 bilinguals only had English at home, while all attended Welsh medium primary and secondary schools. Although the bilinguals in the current experiment have similar home language profiles, they differ in terms of their education, resulting in a range of different bilingual profiles. Of the forty participants in the current experiment, 11 acquired both Welsh and English from birth, 16 acquired Welsh first followed by English (before the age of 6), 7 acquired English first followed by Welsh (before the age of 6), 3 acquired English first and began learning Welsh at the age of 11 and 3 acquired Welsh from birth and began learning English in an instructional setting at 7/8. However, only 26 of the participants had Welsh medium education at both primary and secondary level, while 14 did not. These 14 varied in their bilingual education combinations (e.g., bilingual (dual track) primary and secondary education, and English at primary and secondary education (Welsh taught as a subject)). Therefore, in light of the differences between the Welsh speakers in this experiment and those in Binks’ (2017) study, accuracy rates are comparable on this task assessing the comprehension of gender in non-local contexts, in spite of the differences between the Welsh speakers.

No further results were reported in Binks’ study concerning the adult data; however, they reported findings emerging from teenager data. Binks found that performance was better on feminine items (72.86%) than masculine items (66.02%) and that the teenagers correctly chose *ei* + AM (associated with a feminine possessor) (75.40%) more often than they chose *ei* + SM (associated with masculine items) (68.24%) (*p*=.001). Similarly, performance on feminine-marked pronouns (69.11%) was significantly stronger in comparison to masculine-marked pronouns (61.45%) (*p*=.000). However, the effect of gender (masculine, feminine) and linguistic condition (possessive adjective forms and anaphoric pronouns) were not found in the current adult data, as the results showed similar performance for both linguistic conditions and for both genders across the two linguistic conditions.

The absence of an effect of gender in the current results is also in contrast to the results in Thomas and Gathercole's (2005) study investigating receptive abilities in adult Welsh speakers. Their results revealed that accuracy was higher for interpreting the "feminine" possessor in sentences involving feminine referents (85%) than the "masculine" possessor of sentences involving masculine referents (79.4%). They attributed the gender effect to the fact that feminine referents are likely to be more salient because of the AM after 'ei' as signalling feminine possession, than the SM signalling masculine possession. However, this explanation is only applicable to one of the two linguistic conditions assessed in this task (i.e., not when gender was marked via third person singular anaphoric pronouns ('hi' [f], 'fe' [m])). Furthermore, this explanation is not extendable to the current set of results, given the absence of a gender effect (for both linguistic conditions).

The absence of a gender effect and the similar performance between the two conditions may be accounted for by the following reasons. First, it may be that identifying co-referential pronouns in non-local contexts when gender is encoded through mutations (via possessive adjective forms), is as established as when gender is independent of mutations and is marked via anaphoric pronouns. It may be that the two instantiations of gender marking in non-local contexts are comparably robust, where one instance is not more salient than the other. This argument would be in contrast to those put forward by Thomas and Gathercole (2005) and Binks (2017). Second, as the group of participants in the current experiment are highly proficient Welsh speakers, it is possible that the task is more suitable and discriminates better with a wider range of proficiency levels (despite having a variety of bilingual profiles and not all being Welsh dominant speakers). Third, there may be an insufficient number of items to capture the data required to find evidence of a difference between the two instantiations of gender marking in non-local contexts.

The second experimental task assessed the production of grammatical gender in local contexts, when gender was marked via gendered numerals (2, 3 and 4). Two scoring systems were established, first, to assess gender accuracy regarding the production of the gendered numeral forms, and second, to understand the relationship between gender and mutations in numeral-noun contexts. This secondary scoring system establishes how nouns are produced following the different numerals, as here, gender operates independently of mutations (following 4 [m] & [f], and 3 [f]) and in conjunction with mutations (following 2 [m] & [f], following 3 [m]).

The results from the first scoring system showed an overall group mean of 73.44% (*SD*=15.33%), with a wide range is scores (50 – 100%). When divided according to numerals, the descriptive results revealed that performance was best for the gendered numeral 'pedwar' (4 [m]) (*M*=82.2), followed by 'dau' (2 [m]) (*M*=78.1), 'tri' (3 [m]) (*M*=75.3), 'dwy' (2 [f]) (*M*=71.9), 'pedair' (4 [f]) (*M*=70.3) and then 'tair' (3 [f]) (*M*=65.3). The results from the best-fitting model showed an effect of gender (p = 0.03), with masculine numerals having a higher score than feminine numerals. However, there was no numeral effect (as suggested by the descriptive results). Pairwise comparisons showed that there were no statistically significant interactions between gender and numeral form(s) (ps > 0.055). These results indicate that there is a production accuracy hierarchy for gendered numerals in the Welsh adult speakers, however, this is not a statistically significant finding. Yet, masculine numeral forms are produced more accurately than their feminine counterparts. Overall, group accuracy is relatively high, however, there is variability in the data.

The elicited production task overall gender accuracy result (first scoring system) is 73.44%. This result cannot be compared to any previous findings, as this task is a replication of Sharp's (2012), and they did not report an overall accuracy score. However, this result itself is briefly discussed here. The score of 73.44% is relatively high, sitting comfortably above chance, however, it may be considered below expected 'target-like' levels (i.e., 90+% accuracy) given that the participants are highly proficient Welsh speakers. Even though the masculine numeral forms are more frequent in the input than the feminine forms (due to the Welsh counting system), one may wonder whether producing and hearing gendered numeral forms in the input is an infrequent co-occurrence between numerals and nouns. Previous research has found that speakers often rely on co-occurrence relations between gender marked modifiers and nouns, and this co-occurrence relation between numeral and noun is likely to be less frequent than hearing other sequences in the input, such as Det-Noun or Det-Adj-Noun sequences (Grüter et al., 2012). Therefore, accuracy may never be at ceiling levels given the frequency of exposure to these co-occurrences in the input.

The effect of gender in the current results (scoring system one), with masculine items having a higher score than feminine items, is in line with Sharp's (2012) findings. They reported an effect of gender, with the masculine gendered numeral forms produced more accurately than feminine gendered numeral forms. In agreement with suggestions made by Sharp, the higher accuracy of masculine forms in comparison to the feminine forms may stem from the fact that there is a higher percentage of masculine nouns in Welsh than feminine nouns (approx. 69% [m], 27% [f] and 4% indeterminate, Hammond, 2016). Additionally, the masculine forms of the numerals 2, 3 and 4 are taught as part of the Welsh counting system (e.g., un (one), dau [m] (two), tri [m] (three), pedwar [m] (four), pump (five) and so forth) and the feminine forms are not.

The elicited production task second scoring system excluded incorrect items based on the first scoring system, thus, reducing the number of items considerably. When the results were divided according to the numeral forms, gender-mutation performance was best for nouns following ‘pedwar’ (4 [m]) (M=79.9), then ‘dwy’ (2 [f]) (M=66.3), ‘pedair’ (4 [f]) (M=61.6), ‘tair’ (3 [f]) (M=58.8), ‘dau’ (2 [m]) (M=55.6) and then ‘tri’ (3 [m]) (M=33.8). The results from the best-fitting model revealed that there was no gender effect (p = 0.30), however, mutation accuracy for nouns following numeral *three* had significantly lower accuracy than nouns following numeral *four* (p = 0.002). Pairwise comparisons revealed six statistically significant results (see Table 4.24 for details). Importantly, what emerged from the results was that accuracy for 4/6 comparisons was higher for nouns produced in their bare form compared to mutated nouns (SM following 2 feminine, AM following 3 masculine). One comparison showed that mutation accuracy was higher when it is SM [f] compared to AM [m], and the other showed that no-mutation accuracy is higher for masculine nouns (4 [m]) compared to feminine nouns (3 [f]). This indicates that accuracy is highest when nouns are not mutated in comparison to when they are mutated, and when nouns are mutated, accuracy is higher for SM [f] nouns in comparison to AM [m] nouns, in numeral-noun contexts.

The second scoring system assessed participants for noun mutation / no-mutation accuracy, based on the correct gendered numeral responses. It is not possible to compare these results to Sharp’s results, as they did not report further results to those beyond numeral form and the gender effect. However, it is possible to discuss these numeral-noun phrase results in light of previous mutation / no-mutation accuracy production data for nouns following the determiner (‘y’ = ‘the’). Recall that for this second analysis, there was no effect of gender (p = 0.30). This lack of gender effect is in line with Sharp (2012), who assessed the production of

Det-N sequences in 15 L1 Welsh speaking adults (from North Wales). The results revealed that the adults performed close to ceiling levels for both feminine nouns (SM) (85%) and masculine nouns (bare form) (82%) following the determiner, but there was no effect of gender. The results from the current experiment and from Sharp's study, however, are in contrast to those found by Thomas (2001). Thomas examined the production of Det-N sequences in 30 Welsh speaking adults (from North Wales) and found an effect of gender, with errorless production of masculine nouns in their bare form (i.e., 100%), compared to 88% accuracy for SM feminine nouns. Even though gender was not examined via Det-N sequences in the current experiment, gender was examined in local contexts via gendered numerals and similar patterns have emerged. For instance, accuracy was highest when nouns are not mutated following numerals in comparison to when they are mutated, and when nouns are mutated, accuracy is higher for SM nouns [f] in comparison to AM nouns [m]. This pattern emulates those found by Thomas (2001); however, numeral-noun accuracy scores are generally lower and show more variability than those found for Det-N sequences in previous studies. This is likely because these nouns are being assessed following gendered numerals, and this co-occurrence is less frequent in the input (Grüter et al., 2012).

The results from the first task are used to address research question one, (1) 'To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in comprehension when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?' and the second experimental task results are used to address research question two, (2) 'To what extent do Welsh-English adult bilinguals show accurate use of grammatical gender in production when gender is independent of mutations and in conjunction with mutations, in local contexts?'. The following sections discuss these results in light of the two theoretical approaches reviewed in chapter 3, beginning with the *Interpretability Hypothesis*, followed by the *Missing Surface inflection Hypothesis.*

4.3.2 Interpretability Hypothesis

This section discusses the results in light of the *Interpretability Hypothesis* (Tsimpli, 2003; Tsimpli & Dimitrakopoulou, 2007). First, the comprehension task results are discussed, followed by the production task results.

Under the IH, it was predicted that for the comprehension of grammatical gender, when gender is independent of mutations (i.e., via anaphoric pronouns 'hi' and 'fe'), this should not present itself as a problematic area for the Welsh speakers, whereas when gender is encoded through mutations (i.e., via possessive adjective forms 'ei'+AM/SM), this may constitute a problematic area for the Welsh bilingual speakers. Therefore, evidence to support these predictions under the IH would be higher accuracy scores for the condition when gender is independent of mutations than when gender is encoded through mutations. This difference in scores between the two conditions should be evident in the descriptive results and thus (potentially) emerge as a significant effect in the best fitting model. However, this is not the case. The descriptive results showed that the Welsh speakers performed similarly across the two linguistic conditions, scoring an average of 76.10% for anaphoric pronouns and 76.88% for possessive adjective forms, with little difference between the two conditions. This is further supported by the best-fitting model, which revealed no significant effect of condition ($p = 0.85$). This suggests that there is no evidence indicating that the Welsh speakers experience any more problems or difficulty with gender when it is encoded through mutations than when it is independent of mutations, during comprehension in non-local contexts.

The results were also looked at in more depth, by dividing the conditions according to grammatical gender, to see if one gender was more problematic than the other, in one or both of the two linguistic conditions. The results showed that accuracy scores were similar for both genders within and across the conditions. Specifically, the participants scored an average of 76.56% for masculine items and 75.63% for feminine items when gender was marked via anaphoric pronouns (gender without mutations(s)), while they scored an average of 76.5% for masculine items and 77.25% for feminine items when gender was marked via the possessive adjective form (gender with mutations(s)). The best fitting model also revealed that there was no overall effect of gender ($p = 0.28$) and that there were no statistically significant interactions between gender and pronoun type ($ps > 1.000$). This indicates that no particular gender presented more difficulty in one linguistic condition than the other, with similar performance for masculine and feminine items in both conditions.

Taken together, it is possible to say that the results from the comprehension task are incompatible with the predictions made by the IH. This lack of evidence to support the IH is in contrast with previous findings. For example, beyond the domain of grammatical gender, support for the IH comes from Prentza and Tsimpli (2013), who examined the comprehension of null and postverbal subjects in L2 English intermediate and advanced learners of L1 Greek. The results revealed that although the advanced learners made fewer errors than the intermediate learners, the results showed that both groups produced a significant number of errors compared to the L1 English speakers, suggesting evidence of optionality in the syntactic properties of null and postverbal subjects in the learners. They argued that this is because interpretable features are more readily available than uninterpretable features as a means of analysing L2 input. The authors concluded that their findings lend support to the IH as the evidence suggests that null and postverbal subjects cause prolonged learnability problems in L2 learners of English, because of their different interpretable feature specification between the Greek / English interlanguage.

The lack of evidence to suggest that the comprehension task findings in the current experiment show support for the IH, may be accounted for by the following reasons. First, the IH argues that uninterpretable features will constitute a problematic area for L2 learners, as the unavailability of uninterpretable features in the L1 renders them unattainable for L2 learners. It is not possible to argue that the comprehension task results are evidence of this, which is in line with findings from Solaimani, Myles and Lawyer (2023). For instance, Solaimani and colleagues investigated the acquisition of relative clauses (RCs) in adult L2 learners of English (L1 Persian and L1 French) and found evidence to suggest that the L2 speakers have access to uninterpretable features in the acquisition of English RCs. Therefore, they argued that they were capable of acquiring the morphosyntactic phenomena. This conclusion can be extended to the current Welsh bilingual data, by suggesting that uninterpretable features are accessible to the Welsh-English bilingual adults. There is no evidence to suggest that the uninterpretable features of grammatical gender (when gender is encoded through mutations) are vulnerable, while those bearing interpretable features are not.

Second, the predictions made for the Welsh bilinguals are based on the theoretical framework set out by the IH, which explicitly compares accessibility to interpretable and uninterpretable features between L1 speakers and L2 learners. It argues that uninterpretable features are susceptible to critical period effects, in that they cannot be acquired after late childhood unless they are instantiated in the L2 learner's L1. Notably, the Welsh bilinguals in the current experiment are highly proficient (as measured by a Welsh cloze test, *M*=90.63%) simultaneous (n=11), early sequential (n=23) and later (n=6) bilingual speakers of Welsh and English. The bilinguals have not acquired Welsh as a second language during adulthood. Of the 23 early sequential bilinguals, only seven acquired English first, followed by Welsh (between the ages of 3-6) and of the six later-bilinguals, only three acquired English first, followed by Welsh (between the ages of 7-11). Therefore, it may be that there is no evidence to suggest that uninterpretable features (i.e., gender when it is encoded through mutations) are unattainable, vulnerable, or problematic for the Welsh adult speakers, as they are not L2 adult learners of Welsh, and all participants started to learn Welsh by the age of 11.

Third, the IH claims a difference in the availability and accessibility of interpretable and uninterpretable features but does not necessarily claim that there will be a difference between gender mutation and no-mutation contexts. This distinction between gender in contexts with and without mutations has yet to be explored and the present discussion is the first attempt to extend this theoretical approach to such Welsh gender findings. The results in this experiment are discussed in light of the IH as it provides theoretical grounding most suitable to discuss the phenomena under consideration. However, it may be that the IH is not suitable to explain such data, when gender in Welsh is encoded through mutations or when it is independent of mutations, given the complexity and intricacy of the gender and mutation systems. These three suggestions are revisited and discussed further following the production task results, as the production task presents more data to be scrutinised under the IH.

Under the IH, it was predicted that for the production of grammatical gender, when gender is marked via gendered numeral forms ([m] & [f]) (i.e., scoring system one), the production of gender should not be problematic, due to the semantic content, indicating a specific numerical value. When gender is examined in relation to mutations (i.e., scoring system two), the production of gender when it is independent of mutations (i.e., 4 'pedwar' [m] / 4 'pedair' [f] / 3 'tair' [f] + no mutation) should not present itself as a problematic area for the Welsh speakers, likely because it is a relatively straightforward acquisition process for the Welsh speaker. Similarly, when gender operates in conjunction with mutations (i.e., 2 'dau' [m] / 2 'dwy' [f] + SM / 3 'tri' [m] +AM), this should not be the locus of difficulty for the bilingual speakers as these are consistent and predictable grammatical rules in Welsh. However, if there is evidence to suggest that the bilinguals experience difficulty with 'tri' + AM, this may be because 'tri' [m] and 'tair' [f] trigger different instances of mutation ('tri' + AM) / no mutation ('tair' + no mutation) on post-numeral nouns, indicating that the presence or absence of a mutation may reflect an uninterpretable feature. Therefore, evidence to support these predictions under the IH would be high accuracy scores for the production of gender marked numerals, high accuracy scores for the production of nouns in bare form (i.e., no mutation) following 'pedwar', 'pedair' and 'tair', as well as high accuracy scores for the production of nouns in SM form following 'dau' and 'dwy' and nouns in AM form following 'tri'. However, variability in scores for nouns following 'tri' may be attributed to the tentative suggestion that it does reflect an uninterpretable feature. Additionally, it may be that performance will be best when scored for gender numeral form accuracy, as it is a pure gender agreement process and is not involved with the mutation system. These patterns should be evident in the descriptive results and thus (potentially) emerge as significant effect and/or interactions in the best fitting model(s). Interestingly, the patterns emerging from the data are somewhat complex, while parts of these predictions are upheld, but not all.

The first prediction, hypothesising that the production of grammatical gender should not be problematic when gender is marked via gendered numeral forms, is arguably met. The overall group score for the production of gendered numeral forms is 73.44%. This suggests that some of the participants have productive command of Welsh grammatical gender, when gender is marked via gendered numerals, while others do not (i.e., due to the wide range in scores 50-100%). The reason that this first prediction is arguably met, is because 73.44% is a relatively high accuracy score and does not clearly indicate that this is a problematic area for the Welsh-English bilinguals. Indeed, it is notably lower than what one would assume 'target-like' or 'ceiling' levels to be given that the participants are highly proficient Welsh speakers (i.e., 90+% accuracy), however, as mentioned previously, hearing and producing numeral-noun phrases is likely to be an infrequent co-occurrence in the input. These computation of co-occurrence relations have been shown to be a key mechanism in infants' early language learning and these co-occurrence relations are often relied upon to detect a noun's gender (Grüter et al., 2012).

In addition to this, it is suggested that the idea of 'target-like' or 'ceiling' levels of accuracy (i.e., 90+%) are adjusted, given that all adult Welsh speakers, those in this experiment and those beyond this research, are all bilingual speakers of Welsh with English. Therefore, there is no 'monolingual' or 'native' baseline to make comparisons to. Researchers must consider how the Welsh adult speakers range in their bilingual profiles, with some having learnt Welsh from birth alongside English (i.e., simultaneous bilinguals), or learnt English later (i.e., 3-11 years), or alternatively, acquired English from birth and learnt Welsh later (i.e., 3-11 years). As a result, it is argued that this first prediction is met, with no clear evidence to suggest that the Welsh speakers, as a group, experience problems with the production of gender when it is marked via gender marked numerals.

The afore-discussed results concerned the production of gender when it is marked via gender numeral forms, which is a pure gender agreement process and is not involved with the mutation system. The participants were also scored for noun mutation / no-mutation accuracy, which explores the production of gender when it operates with and without mutations on nouns. Interestingly, under the IH, it was predicted that when gender was independent of mutations and when it operates in conjunction with mutations, neither should be a locus of difficulty for the bilingual speakers, given the interpretability of the no-mutation / mutation following the gendered numeral forms. Therefore, one would expect first to see high accuracy scores across both conditions and to secondly see comparably similar scores between the two conditions (i.e., independent vs in conjunction). However, these patterns have not clearly emerged from the data, therefore, the predictions are only partially met.

The descriptive results revealed that overall accuracy scores for nouns produced in bare form were on average high (*M*=66.77%), following 'pedwar' (*M*=79.9%) 'pedair' (*M*=61.6%) and 'tair' (*M*=58.8%), whereas the mutated nouns were lower overall (*M*=52.13%), following 'dwy' (*M*=66.3%), 'dau' (*M*=55.6%) (both SM) and 'tri' (*M*=33.8%) (AM). However, these results do not present a clear picture which suggests that nouns are produced more accurately when they are not mutated, than when they are mutated, as the accuracy score for SM nouns following 'dwy' [f] is second highest out of the six results. Pairwise comparisons revealed six statistically significant differences, with four out of the six showing that accuracy is better when nouns are produced in their bare form compared to when they are mutated. However, no significant difference was revealed between nouns following ‘pedwar’ (*M*=79.9%) and ‘dwy’ (*M*=66.3%). Consequently, it is possible to say that the results paint a complex picture.

The prediction stating that when gender operates independently of mutations, this should not present itself as a problematic area for the bilingual speakers, is partially met, as there is evidence to suggest that this is applicable for non-mutated nouns following ‘pedwar’ [m] but less so for non-mutated nouns following ‘pedair’ [f] and ‘tair’ [f], given the reduced accuracy scores. Therefore, it is not possible to say that it is an overall unproblematic area, given the variability in the scores following the three different gendered numerals. However, it could be argued that this feature is stable when following the masculine numeral form ‘pedwar’, which is likely because masculine nouns following the determiner ‘y’ are produced in bare form and this is reflected in this particular numeral-noun sequence.

Similarly, the prediction stating that when gender operates in conjunction with mutations, this should not be the locus of difficulty for the bilingual speakers, is also partially met, as there is evidence that tentatively suggests that this is applicable for soft-mutated nouns following ‘dwy’ [f] but less so for soft-mutated nouns following ‘dau’ [m] and aspirate-mutated nouns following ‘tri’ [m], given the reduced accuracy scores. Even so, the accuracy score for soft-mutated nouns following ‘dwy’ is only relatively high and the aspirate-mutated nouns following ‘tri’ [m] are notably low. Therefore, it is not possible to say that it is an overall unproblematic area, given the variability in the scores following the three different gendered numerals. Rather, one could argue that this area is potentially problematic for Welsh bilinguals, notably so for AM nouns following ‘tri’, in comparison to SM nouns following ‘dwy’ and ‘dau’. It is possible to wonder why AM nouns appear to present more difficulty than SM nouns. As previously mentioned, one possible explanation could be because the gendered numeral forms ‘tri’ [m] and ‘tair’ [f] trigger different instances of mutation (‘tri’ + AM) / no mutation (‘tair’ + no mutation) on post-numeral nouns. It could be argued that the presence or absence of the mutation reflects an uninterpretable feature. Alternatively, the low accuracy scores for AM nouns may be attributed to the fact that this feature (i.e., 3 [m] + AM) is vulnerable. It is possible that the low accuracy score of 33.8% for AM nouns after ‘tri’ is indicative of disuse and/or obsolescence, where the feature is being lost in the Welsh language (Thomas & Gathercole, 2005).

The findings from the production task are complex and do not paint a clear picture, as some evidence lends support to the IH, whereas other evidence does not. However, taken together, it could be argued that the predictions made by the IH are not the most suitable to explain the patterns emerging from the Welsh gender production data. This lack of clear support for the predictions set out by the IH is in contrast to Tsimpli and Mastropavlou (2007), who argued that their findings could be explained by the IH. They investigated the acquisition of articles in Greek L2 child and adult learners, and claimed that the difference in lower accuracy scores for definite articles compared to indefinite articles served as evidence to show support for the interpretability distinction set out by the IH. They argued that the evidence of a priority for indefinite over the definite article in both child and adult data, suggests that definite articles are more problematic due to their uninterpretable features.

In contrast, the lack of clear support for the predictions made by IH was found by Chondrogianni (2008). They also investigated the acquisition of definite articles and direct object clitics in sequential bilingual L1-Turkish-L2 Greek children of differing proficiency levels. The results revealed that with increasing proficiency, the children began to produce definite articles more accurately (low - 26.95%; lower intermediate - 83.79%; upper intermediate - 95.98%; high - 99.52%), showing a fast development after the initial stage. For the direct object clitics, the results revealed that the low proficiency learners did not produce clitics, however, the production of clitics increased as proficiency increased (lower: 55.82%, upper: 80.4%, high: 96.78%). They argued that these results do not support the IH, as it would predict articles and clitics to be equally omitted by the learners at initial stages, given their uninterpretable features. They explained that, albeit at low levels, the low proficiency learners produce articles, whereas they do not produce direct object clitics. Therefore, the predictions made by the IH are not best suited to explain the definite articles and direct object clitics findings in the sequential bilingual children.

Taken together, it is possible to say that some assumptions made by the IH that can be used to explain some of the patterns, however, this is not the case for all of the production data. Therefore, the IH is not considered a particularly suitable theoretical approach to explain the patterns emerging from the Welsh gender production data. Three reasons were mentioned previously as to why the IH may not be compatible with the comprehension task findings. These reasons included the fact that the IH explicitly compares accessibility to interpretable and uninterpretable features between L1 speakers and L2 learners, the assumption that uninterpretable features are problematic for L2 learners, and that it does not necessarily claim a difference between gender mutation and no-mutation contexts. These three reasons can be extended to the production data, in considering why the results do not lend clear support to the IH. The final suggestion, that the IH does not necessarily claim a difference between gender mutation and no-mutation contexts, may be the clearest explanation as to why this is. For example, the relationship between gender and mutations is complex and intricate, with several researchers claiming the Welsh gender system to be opaque as a result its function with and without mutations (e.g., Gathercole, Thomas, & Laporte, 2001; Thomas, 2001; Thomas & Gathercole, 2007). It may be that the intricacies of the two systems working in conjunction and independent of one another is beyond the scope or capabilities of the predictions made by the IH. It is an interesting theoretical approach that merits empirical testing and allows itself to be scrutinised in light of the Welsh gender mutation / no-mutation data, however, it is possible that more data is needed to examine the interpretability distinction proposed by the IH.

Some final considerations for the IH are that although it makes predictions based on the interpretability distinction and that it compares accessibility to interpretable and uninterpretable features between L1 speakers and L2 learners, the IH does not explicitly differentiate between local and non-local gender marking, nor does it explicitly differentiate between comprehension and production performance (Chondrogianni, 2024). Therefore, it is possible that the predictions made by the *Missing Surface Inflection Hypothesis* (Prévost & White, 2000) are more appropriate to explain the patterns emerging from this Welsh gender data, in particular the production data.

4.3.3 Missing Surface Inflection Hypothesis

This section discusses the results in light of the *Missing Surface Inflection Hypothesis* (Prévost & White, 2000). First, the production task results are discussed, followed by the comprehension task results. The production results are discussed before the comprehension results given that the MSIH does not make explicit predictions for comprehension, but only production.

Under the MSIH, it was predicted that for the production of grammatical gender, when gender is marked via gendered numeral forms ([m] & [f]) (i.e., scoring system one), the Welsh bilinguals should not experience any difficulty in producing these forms, given that the MSIH assumes a dissociation between syntax (grammatical gender) and morphology (mutations), and gender marking via numerals only involves the domain of syntax. Evidence to support this prediction under the MSIH would be high accuracy scores for the production of gender marked numerals. This first prediction is arguably met. As noted previously, the overall group score is 73.44%, suggesting that some of the participants have productive command of Welsh grammatical gender, when gender is marked via gendered numerals, while others do not (wide range in scores 50-100%). Echoing the afore-mentioned explanation as to how and why this prediction is met (under the IH also), given that the score of 73.44% is notably lower than what one would assume 'target-like' or 'ceiling' levels to be, is the fact that it is a relatively high score and the perceived 'target-like' or 'ceiling' levels should be adjusted for the Welsh adult bilinguals. This suggestion is put forward in light of the fact that hearing and producing numeral-noun phrases is likely to be an infrequent co-occurrence in the input, and the fact that there is no 'monolingual' or 'native' baseline to make comparisons to, therefore, one should consider how the Welsh adult speakers range in their bilingual profiles. Consequently, it is argued that under the MSIH, this prediction is met, given that gender marked in this way only involves the domain of syntax.

As part of this idea that the MSIH assumes a dissociation between syntax (grammatical gender) and morphology (mutations), it was predicted that accuracy will be higher when gender is examined independent of mutations than when it is in conjunction with mutations. Evidence to support this prediction would be higher accuracy scores for the production of nouns in bare form (i.e., no mutation) following 'pedwar', 'pedair' and 'tair', while accuracy would be lower when nouns are mutated following 'dau', 'dwy' and 'tri'. As discussed in light of the IH, the descriptive results revealed that overall accuracy scores for nouns produced in bare form were on average high (M=66.77%), following 'pedwar' (M=79.9%) 'pedair' (M=61.6%) and 'tair' (M=58.8%), whereas the mutated nouns were lower overall (M=52.13%), following 'dwy' (M=66.3%), 'dau' (M=55.6%) (both SM) and 'tri'

(*M*=33.8%) (AM). Pairwise comparisons revealed six statistically significant differences, with four out of six showing that accuracy is better when nouns are produced in their bare form compared to when they are mutated. However, no significant difference was revealed between nouns following ‘pedwar’ (*M*=79.9%) and ‘dwy’ (*M*=66.3%). Consequently, it is not possible to say that this prediction is met overall, rather, that part of this prediction is met, because the picture is relatively complex.

What the MSIH predicts further to the IH, are the patterns beyond the general dissociation between syntax (grammatical gender) and morphology (mutations). There are two additional predictions made under the MSIH which aim to explain the patterns in the results. First, the MSIH can be used to predict that when the Welsh speakers need to produce a soft mutated form, they may produce a default form (i.e., no mutation) in place of the more specified form (i.e., soft mutation). Second, when the Welsh speakers need to produce a mutated form, they may produce a less specified form (i.e., soft mutation) in place of a more specified form (i.e., aspirate mutation). Prior to removing items that the speakers incorrectly produced the gendered numeral form, there is evidence of the Welsh speakers resorting to the default form (i.e., no mutation) than the more specified form (SM). Nouns following the feminine numeral two ‘dwy’ should be SM, this was done correctly 83.3% of the time, while 16.3% of the nouns were produced with no mutation and one instance (0.4%) AM. This is in line with Thomas (2001), who found that errors with SM accuracy for feminine nouns were attributed to feminine nouns being produced with no mutation instead of the SM, but no evidence of feminine nouns being AM. Nouns following the masculine numeral two ‘dau’ should be SM and this was done correctly 72.2% of the time, while 27.2% of the nouns were produced with no mutation and two instances of nouns being over-mutated (0.8%) (i.e., double SM – ‘t’ > ‘d’ SM, and again ‘d’ > ‘dd’). This finding that Welsh adult speakers resorted to the default form (no mutation) where a more specified form (SM) is required, is arguably compatible with the predictions of the MSIH. This may be because the adults fall short of specifying certain nouns with the target gender feature due to its entanglement with mutations (Unsworth, 2013).

After the masculine numeral three ‘tri’, masculine nouns AM. In place of producing the most specified form, i.e., AM, it could be that nouns are produced in a less specified form, i.e., SM. The accuracy for AM on the target nouns was low

(33.8%) and this was even the case prior to removing items that the speakers incorrectly produced the gendered numeral form (i.e., producing 'tair' [f] in place of 'tri' [m]). Prior to excluding these items, only 42.5% of the nouns were accurately AM following 'tri'. The errors were a result of the participants producing the nouns in their bare form, not their AM form and there were no instances of SM. Therefore, it is not possible to say that the Welsh speakers resorted to a less specified form than the more specified form, but instead, resorted to the default form (i.e., no mutation) in place of the most specified form (AM). This finding is in line with Unsworth (2013), who suggested that the Dutch-English bilingual children resorted to using the default common form *de* in place of using the more specified determiner form *het* for neuter nouns. Therefore, it is not possible to say that the Welsh speakers are using a less specified form where the more specified form is required, for nouns following the masculine numeral three 'tri'. As a result, the second prediction is not met, as there is evidence to suggest that they produce a default form (i.e., bare form) where the most specified form (i.e., AM) is required.

Interestingly, there were also instances of mutated nouns when nouns should be produced in bare form. For example, 5% of nouns following the masculine numeral four 'pedwar' were SM instead of no mutation, and 14.7% of nouns following the feminine numeral four 'pedair' were SM instead of no mutation. Nouns following the feminine numeral three 'tair' should be produced in bare form, this was done correctly 88.4% of the time, while 8.1% of the nouns were SM and 3.4% of the nouns were AM. This demonstrates an interesting interplay between the mutations when examining with the production of these three gendered numeral forms.

Taken together, the predictions made by the MSIH can partly explain the data. Similar to the IH, the MSIH can be used to explain the gendered numeral production scores (scoring system one). The MSIH can also explain some of the mutation / no-mutation patterns found in this production task, but not the whole data set. It is possible to say, however, that the predictions made by the MSIH are largely suitable to describe the findings from this production task. Even though the predictions do not fully explain the patterns emerging form the data, it can be considered the most suitable of the two theoretical approaches.

Noted previously is the fact that the MSIH does not make explicit predictions for comprehension, only production. It does, however, propose that performance is poorer on production measures than comprehension measures as a by-product of communication pressure, because production happens during real-time, requiring fast, spontaneous and automatic processing, while comprehension tasks do not typically involve such pressures (Alarcón, 2011; Grüter et al., 2012; Hopp, 2013, 2016). When comparing the accuracy scores for gender when it is with mutations and without mutations on the comprehension task and the production task, the production scores are overall lower, for both linguistic conditions. The participants scored an average of 76.10% for anaphoric pronouns (gender without mutations) and 76.88% for possessive adjective forms (with mutations) on the comprehension task, while on the production task, they scored an average of 66.77% for non-mutated nouns following 'pedwar', 'pedair' and 'tair', and an average of 52.13% for mutated nouns following 'dwy' (SM), 'dau' (SM) and 'tri' (AM). Therefore, the descriptive results may serve as evidence to suggest that performance is poorer on production than comprehension, when looking at the relationship between gender and mutations.

This finding is arguably in line with previous research. For example, Alarcón (2011) examined Spanish Det-N-Adj gender agreement and found better performance on comprehension measures that production measures in L2 Spanish speakers. Alarcón (2011) claimed that the differences between L2 comprehension and production performance is due to a mapping problem, in that errors stem from difficulties in accessing the abstract features and mapping them onto their surface forms. They concluded that the learners have the gender features in their underlying grammars despite their difficulties in mapping the abstract gender feature to its appropriate form. Similar results were found by Montrul (2011) who also examined Spanish Det-N-Adj gender agreement in a production task and two written tasks in L2 Spanish learners and found performance to be better on both comprehension tasks compared to the production task. Montrul (2011) suggested that the L2 learners have acquired the abstract gender feature but have difficulty in realizing gender morphology in production, due to a production performance problem, which is consistent with the MSIH.

One linguistic property that has not been discussed at length but is considered briefly here, is the role of locality (i.e., local vs non-local gender agreement). If there was evidence suggesting that performance was better on the production task than the comprehension task in this experiment, this may be due to the fact that the production task measured gender marking in local contexts and the comprehension task measured gender marking in non-local contexts. This could be suggested under the assumption that accuracy may be lower for gender agreement in non-local contexts, because of a processing burden imposed by linear distance (e.g., Keating, 2010). However, there is some evidence to suggest that performance is poorer on the production task than the comprehension task, rather than the other way around. The overall lower scores in production may be because the Welsh bilinguals speakers fall short of mapping the appropriate gender forms to features in the given grammatical context, reflecting a problem with spelling out gender-marked forms in production, and not due to a representational deficit or impairment (Black & Tararova, 2020; Hopp, 2013, 2016b). The MSIH does not explicitly distinguish between local and non-local gender marking, however, the MSIH provides a fruitful testing ground to examine the production of gender with and without mutations, in both local and non-local contexts. However, this is beyond the scope of this experiment.

In conclusion, it is possible to say that the MSIH is a suitable theoretical approach to discuss the production of gender and mutations in Welsh-English adult bilinguals. Notably, it does not fully explain all of the patterns emerging from the data, yet it can explain the data in some parts. The MSIH does not, however, make explicit predictions for comprehension. Therefore, it is not considered an appropriate theoretical approach to try and explain the patterns emerging from the comprehension data. If there was evidence of a dissociation between grammatical gender and mutations in the comprehension task, it would be sensible to try and extend the MSIH to that data. However, the results revealed that the Welsh speakers performed similarly across the two linguistic conditions, with one condition marking gender with mutations and the other marking gender without mutations.

In addition to the two main research questions posed, investigating the comprehension of gender in non-local contexts and the production of gender in local contexts, this experiment also explores the roles of two individual difference variables - language dominance and Welsh linguistic proficiency - in the production and comprehension of gender. These two variables are discussed in the section below.

4.3.4 Individual difference variables

A secondary interest to both the comprehension and production tasks, are the possible roles of language dominance and Welsh linguistic proficiency. Two sub research questions were posed to explore this, these were (1a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' comprehension of Welsh grammatical gender, when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts?' and (2a) 'What are the roles of dominance and proficiency in Welsh-English adult bilinguals' production of Welsh grammatical gender, when gender is independent of mutations and in conjunction with mutations, in local contexts?'. The results regarding these individual differences are briefly summarised before discussing them in light of previous studies.

Regarding the roles of these variables in the comprehension of grammatical gender in non-local contexts, the results revealed a statistically significant effect of proficiency (p = <0.001) but no effect of dominance (p = 0.91). Additionally, there was a significant interaction between dominance and proficiency (p = 0.004). Additional models were run checking the interactions between the different linguistic conditions and the individual difference variables, but there were no other significant interactions, different to the two significant results emerging from the best-fitting model. As a consequence, it is not possible to explain how proficiency and dominance affect the two linguistic conditions or gender (masculine, feminine) specifically, other than (1) as proficiency scores increased, overall accuracy scores also increased, (2) there was no effect of dominance, and finally, (3) participants who are more Welsh dominant showed increased proficiency scores.

Similarly, regarding the roles of these variables in the production of grammatical gender in local contexts when gender was marked via gendered numeral forms (i.e., first scoring system – gender only), the results revealed a statistically significant effect of proficiency (p = <0.001) but no effect of dominance (p = 0.92). Additionally, there was a significant interaction between dominance and proficiency (p = 0.002). Additional models were run checking the interactions between the different linguistic conditions and the individual difference variables, but there were no other significant interactions, different to the two significant results emerging from the best-fitting model. Therefore, it is not possible to explain how proficiency and dominance affect the different gendered numeral forms or gender specifically, other than (1) as proficiency scores increased, overall accuracy scores also increased,

(2) there was no effect of dominance, and finally, (3) participants who are more Welsh dominant showed increased proficiency scores.

As part of the second scoring system (i.e., second scoring system – gender-mutation), the results revealed a statistically significant effect of proficiency (p = <0.001) but no effect of dominance (p = 0.62). Additionally, there was a significant interaction between dominance and proficiency (p = 0.001). An additional model was run checking the interactions between the linguistic conditions, gender, and proficiency, and it revealed a significant interaction with gender (p = 0.002), with higher proficiency speakers showing increased accuracy for feminine nouns in comparison to masculine nouns. Furthermore, similar to the comprehension data and the gender only data in the production task, the interaction between dominance and proficiency is a result of participants who are more Welsh dominant showing increased proficiency scores. Therefore, the results indicate that (1) as proficiency scores increased, overall accuracy scores also increased, (2) higher proficiency Welsh speakers produced mutation / no-mutation on feminine nouns more accurately than masculine nouns, (3) there was no effect of dominance, and finally, (4) participants who are more Welsh dominant showed increased proficiency scores.

There were two overarching predictions made in light of previous research. The first predicted that dominance will play a role in both the comprehension and the production of grammatical gender in the Welsh-English bilinguals. The second predicted that both Welsh and English proficiency will play a role in both the comprehension and the production of grammatical gender in the Welsh-English bilinguals. Notably, English proficiency was not explored, given that the test for internal consistency for the English proficiency measure was found to be quite low. Therefore, only Welsh linguistic proficiency was investigated as part of this experiment. The results summarised above show that language dominance did not play a role in either the comprehension or production of grammatical gender tasks. Therefore, this prediction is not met. In contrast, the results showed that Welsh linguistic proficiency played a role in both the comprehension and production of grammatical gender tasks. Therefore, this prediction is met.

The lack of evidence to suggest that language dominance predicts outcomes in the comprehension and production of grammatical gender in Welsh, is not in line with previous research. Previous studies have found language dominance to be a predictive factor in linguistic outcomes in various different phenomena. For instance, Amengual (2016b) found that language dominance in the target language of assessment, predicted outcomes. The results from a picture naming task examining the production of the Catalan back mid-vowel contrast showed that simultaneous Catalan-Spanish bilinguals who were Catalan dominant had higher accuracy rates than the bilinguals who were Spanish dominant. Amengual interpreted this finding to suggest that language dominance was a strong predictor of performance in the production of the Catalan back mid-vowel contrast, as operationalised and measured by the BLP. Perpiñán (2017) also found language dominance to play a role in the production of the expression of Catalan clitics in simultaneous bilinguals. The results from an elicited production task showed that Catalan-Spanish bilinguals who were Spanish dominant, exhibited higher errors rates than Catalan dominant speakers. Similar results were found by Bonvin, Brugger and Berthele (2021) who investigated the relationship between language dominance and a lexical test (LexTale) in French / German bilinguals. The results showed a strong linear association between BLP and LexTale, suggesting a relationship between dominance and vocabulary proficiency.

In light of previous findings, it was predicted that dominance would play a role in the comprehension and production of Welsh gender. However, there is no evidence to suggest this for either task. There are two possible reasons for these findings, largely explained by methodological considerations. First, the BLP assesses language dominance through self-reports. Given that the individual completes the questionnaire themselves, it is possible that there is an intricate relationship between their self-rated ability and extra-linguistic factors, such as confidence. Previous research has found confidence - specifically, the lack of - to impact upon regular Welsh language use in speakers (e.g., Hodges, 2024; Morris, 2014; Newcombe, 2002). For example, Morris (2014) explored the relationship between Welsh speakers' backgrounds and their self-confidence in their language skills and use of Welsh, in 27 bilinguals from a Welsh-dominant area of North West Wales and an English-dominant area of North East Wales. They found that those from English-speaking homes felt less confident in their Welsh from both areas. It is possible that the Welsh speakers in the current experiment lack confidence in their language skills and use of Welsh, therefore scoring themselves lower in the proficiency, use and attitudes' sub-sections of the BLP, particularly for those who come from English speaking homes. Therefore, it may be that the scores generated by the BLP are not a wholly accurate representation of the Welsh speakers' language dominance. This is albeit a tentative suggestion, however, it could provide grounds for future research to consider administering detailed questionnaires, such as the BLP to capture the bilingual's language dominance, as well as an additional questionnaire capturing the demographic background of the Welsh speaker and their language use and their attitudes, such as the one administered by Morris in their study.

Second, it is possible to wonder whether the BLP captures the depth of detail needed to produce a continuous dominance score that represents the bilingual speaker. The BLP posits that it assesses the general bilingual profile and considers a variety of linguistic variables. This is accurate; however, it is possible that the BLP misses aspects of a bilingual speaker's profile in order to provide a detailed insight into the Welsh speaker's contact with the Welsh language, in terms of literacy skills, including reading, speaking, and writing in Welsh. For instance, in the sub-section of language use, it considers the speaker's percentage of use of the two languages when they are with friends, family, at school/work, the language they use speak to themselves and the language they count in. The section does not include questions relating to the amount of Welsh while reading, writing, on social media, watching tv, listening to the radio or listening to music, for example. It is possible that the BLP does not capture the bilinguals use sensitively enough of the two languages in their day to day lives. This methodological consideration is revisited in the limitations section of this thesis.

While there is no evidence to suggest that language dominance plays a role in the comprehension and the production of grammatical gender in the Welsh-English bilinguals, there is evidence indicating that Welsh linguistic proficiency plays a role in the comprehension and the production of grammatical gender in the bilinguals. This is in line with previous findings, who have found an effect of proficiency on task scores. For instance, Montrul et al (2008) found more target-like performance of Spanish gender agreement in oral production among heritage speakers of Spanish than the L2 Spanish learners. Montrul attributed the difference in performance to proficiency levels, with the L2 Spanish learners producing more gender agreement errors. The authors suggested that this was due to lower proficiency levels compared to the heritage speakers and that the higher accuracy rates in the heritage speaker data are more representative of fast, implicit, and automatically processed knowledge, which is typically acquired during childhood.

Previous research also found that L1-like processing of gender can be attained by L2 learners, but that this depends on proficiency. In a study by Sagarra and Herschensohn (2010), results from a SPR task revealed that intermediate L2 learners and L1 speakers were sensitive to the gender agreement violations, but the L2 beginners were not. The authors suggested that L2 learners with a certain proficiency level can develop processing patterns that are qualitatively similar to those of L1 speakers. Similar results were obtained by Foucart and Frenck-Mestre (2011) who examined gender agreement violations in French L1 speakers and advanced L2 French learners (L1 German). The results revealed that both groups evidenced a P600 response. This led the authors to suggest that high-proficiency L2 learners who receive enough exposure to their L2 can process gender in a similar way to L1 speakers. Therefore, they claimed that syntactic processing in L2 is affected by proficiency levels in the target language.

It is therefore possible to argue that these findings are in line with previous research, showing that the Welsh linguistic proficiency levels in Welsh-English bilingual adults modulated the speakers' performance on measures of grammatical gender. Although previous research has explored the effect of proficiency on measures in L2 populations and that the Welsh bilinguals are not considered L2 speakers of Welsh, it is possible that Welsh proficiency modulated the comprehension and production scores because of the notable variability in the data. Even though the Welsh bilinguals as a group, performed at ceiling on the Welsh cloze test, these ceiling effects are not evident in the experimental measures. Therefore, it may be that the proficiency can explain some of the variability given the wide range in scores, across both experimental tasks.

An additional note here, is the possible fact that the patterns observed in the results are due to other individual difference variables which have not been explored in this experiment. For instance, some factors found in previous studies to be important include those such as the age of onset, memory, and other environmental factors (e.g., minority language use in the home, availability of a speech community beyond the home, and the use of the minority language in formal settings, among many others). Even though some of these factors and covariates are beyond the scope of this first experiment, some of these individual difference variables are considered in the follow-up experiment.

In conclusion, the findings discussed in this section contribute to the call for further research to investigate how Welsh–English bilingual adults' lives influence their language proficiency in more detail, in order to establish which factors predict and influence higher performance on measures of morphology. Future research should aim to disentangle which additional variables can account for meaningful variability within bilinguals, adding to Welsh linguistic proficiency, which was found to be a relevant predictor of differential outcomes in this experiment.

4.4 Chapter summary

This first experiment set out to explore four things:

- To tease apart the gender and mutation systems, addressing the point put forward by Thomas (2001) questioning whether the variable use of gender and mutations in Welsh adult speakers is due to a lack of gender knowledge or rather, is it an indication of adults' lack of ability and consistency with producing mutations in general.
- To examine the comprehension and production of grammatical gender in a more geographically diverse Welsh adult population than those previously tested.
- To consider two individual difference variables, namely, language dominance and linguistic proficiency, in order to investigate whether these factors influence higher performance on measures of morphology in Welsh-English bilingual adults.
- To see if the patterns emerging from the comprehension and production data could be explained by the predictions made by either the IH or MSIH.

This section (4.3) discussed the results presented in section 4.2 of this chapter. I have attempted to extend the predictions made by the IH and the MSIH to explain the patterns emerging from the two experimental tasks. It is possible to say that the MSIH is a suitable theoretical approach to discuss the production of gender and mutations in Welsh-English bilinguals. Notably, it does not fully explain all of the patterns emerging from the data, yet it can explain the data in some parts. The MSIH does not, however, make explicit predictions for comprehension and is therefore not considered a suitable theoretical approach to try and explain the comprehension data. However, the predictions made by the IH are arguably incompatible with the comprehension data.

Discussing the results in light of these two approaches is enlightening and informative for the field, to understand these findings under the IH and MSIH, given that these Welsh-English adult bilinguals speakers use Welsh in minority language conditions, with English as the dominant societal language. It is possible that there is another theoretical approach of L2 ultimate attainment that is more appropriate to explain the comprehension task data, however, this exploration is beyond the scope of this thesis.

Experiment two (presented in Chapter 6) extends the findings from comprehension and production to the processing of Welsh gender, investigating Welsh speakers' representation of gender during real-time. It explores whether the patterns emerged from the production data are replicated in real time processing. It also explores the role of individual difference variables, including language dominance and Welsh proficiency, in addition to some additional factors, to try and best understand what influences the processing of Welsh gender in Welsh-English bilingual adults. The following chapter discusses the processing of grammatical gender in adult L1 and L2 bilingual speakers, paying attention to theoretical approaches attempting to account for the variability observed in the data and the possible individual differences which may influence the processing of gender in these speakers.

Chapter 5
The Processing of Grammatical Gender

Chapter 4 presented the first experiment of this thesis, exploring the production and comprehension of grammatical gender in Welsh-English adult bilinguals. This second experiment extends the findings to the processing of Welsh gender, investigating Welsh speakers' representation of gender during real-time and considers the possible role of additional individual difference variables. It explores whether the patterns emerged from the first experiment, particularly the production data, is replicated in real time processing (Slabakova, Leal, Dudley, & Stack, 2020).

This fifth chapter now turns to the processing of grammatical gender in adult speakers. Previous research has investigated how L2 users (typically adult L2 learners who have learned an L2 later in life) differ from L1 speakers in regard to real-time comprehension. Research has suggested two main approaches: that L2 processing is qualitatively different from L1 processing, or alternatively, that processing in the L2 is quantitatively different and affected by individual differences (Cunnings, 2017; Hopp, 2022).

Various theoretical approaches attempt to explain how and why L2 adult learners differ from L1 monolinguals in their processing patterns. The Welsh bilinguals in the present thesis are not adult L2-like, nor are they like L1-monolingual speakers as they have more than one language, however, the Welsh-English bilingual speakers have two languages in their minds. It is therefore possible that they may resemble behaviour similar to both L1-monolinguals and L2 adult learners, or processing patterns somewhere along that spectrum. For Welsh-English adult bilinguals, processing patterns during real-time comprehension is unknown. Therefore, this is the focus of the second experiment, which will be reported in Chapter 6. But first, to understand processing behaviours in L1 and L2 adult speakers, three major overarching accounts of L2 processing are reviewed, including the Shallow Structure Hypothesis (Clahsen & Felser, 2006), the Declarative Procedural model of lexicon and grammar (Ullman, 2001, 2020) and cognitive capacity accounts (Hopp, 2010; McDonald, 2006). These theoretical models may be extendable and relevant to the current Welsh-English bilingual adult data.

5.1 Overview

In the following sections, I will outline the Shallow Structure Hypothesis (SSH) (Clahsen & Felser, 2006), followed by the Declarative Procedural model of lexicon and grammar (Ullman, 2001, 2020), before turning to examine cognitive capacity accounts (Hopp, 2010; McDonald, 2006), discussing support for each of these. I will also briefly review some of the approaches arguing for qualitative similarity between L1-L2 processing, but attribute any differences to a variety of reasons, such as the *Full Parse Hypothesis* (Swanson & Dekydtspotter, 2022), the lexical bottleneck hypothesis (Hopp, 2018) and the memory interference account (Cunnings, 2017). In reviewing these approaches, I hope to provide a well-rounded picture of the current state of the L2 processing literature which may be relevant to the Welsh bilinguals in this second experiment. The theoretical approaches are discussed in light of grammatical gender processing, the phenomenon under investigation in this thesis.

5.2 Shallow Structure Hypothesis (SSH)

One of the three major theoretical accounts which deserves attention is the Shallow Structure Hypothesis (SSH) (Clahsen & Felser, 2006). The SSH builds on existing L1 processing findings, and suggests two routes are available during L2 sentence processing (e.g., Ferreira, Bailey, & Ferraro, 2002). One involves the computation of a fully specified syntactic structure (termed “the grammatical route” in Clahsen & Felser, 2018, p. 698), while the other relies on lexical-semantic and surface information. This second is involved in less detailed representation of the input (termed “the heuristic route”). Both processing routes can also operate in parallel and neither L1 nor L2 speakers are proposed to be restricted to a particular processing route (Clahsen & Felser, 2018, p. 697). The assumption is that both L1 and L2 speakers have the same parsing mechanisms available. However, L1 and L2 processing systems may differ in terms of the underlying grammatical representations and the weighting of grammatical constraints. In L2 processing, the heuristic route may be used more than the grammatical processing route, as the L2 processing system relies more strongly on non-grammatical than on grammatical information.

In this way, the SSH discriminates between the knowledge / grammar and the parser. It is possible that L2 speakers may lack knowledge of a particular grammatical rule or be in possession of incorrect knowledge regarding that rule, or that they have L1-like representations and prioritizations of grammatical information, but these information sources may not be relied upon to guide real-time processing (Clahsen & Felser, 2006). The observed differences between L1 and L2 processing can be explained by assuming that the syntactic representations adult L2 users compute during comprehension are shallower and less detailed than those of L1 speakers (Clahsen & Felser, 2006). In other words, L2 speakers compute grammatical representations that lack complex hierarchical structure and more abstract elements of syntactic structure, where they build local (i.e., short-distance, linear) morphosyntactic dependencies but not complex ones (i.e., non-local, hierarchical). The L2 parses are said to be flat and shallow compared to the more complex parses of L1 speakers, suggesting that the differences between L1 and L2 speakers in reliance on these two routes are not assumed to be only quantitative, but rather qualitative (Clahsen & Felser, 2006).

The proposal of the SSH was inspired by a number of processing studies using different methodologies and linguistic structures. These studies arguably showed that L2 speakers experience difficulty resolving relative clause attachment ambiguities (e.g., Felser, Roberts, Marinis, & Gross, 2003), and that the processing of gender agreement in L2 is not L1-like (e.g., Sabourin, 2003). For instance, Felser et al (2003) examined whether two groups of advanced L2 users of English (L1: German & Greek) utilise the same parsing principles during syntactic ambiguity resolution as L1 speakers. A SPR task was administered, examining how the users resolve relative clause attachment ambiguities in two-site contexts (*of* vs. *with* NPs). The sentences were temporarily ambiguous and contained a relative clause modifying either the head of the overall object NP (NP1 attachment) or the embedded noun phrase (NP2 attachment). The results revealed that the L2ers showed an NP2 attachment preference for the preposition *with* but showed no clear preference for sentences containing the preposition *of*. This is in contrast to the L1 English speakers' results, who showed an NP2 preference for sentences containing both *of* and *with*. The authors indicated that the L2ers do not process ambiguous sentences of this type in the same way as adult L1 speakers of English, with L2ers relying more on non-structural information in parsing ambiguous sentences.

The results from a study examining gender agreement were also taken as evidence for the shallow parsing approach. For example, Sabourin (2003) examined the processing of gender agreement violations in L2 learners of Dutch (L1s: German, English, French, Italian, Portuguese & Spanish). While event-related potentials (ERPs[13]) were recorded, the participants read and judged sentences which contained gender agreement violations in three different contexts, within local (between a definite determiner and an adjacent noun, and between an indefinite noun and an adjacent adjective) and non-local domains (between a noun and an adjacent relative pronoun). The ERPs data showed a clear P600 effect[14] and a late frontal negativity in the L1 German-L2 Dutch speakers, indicating sensitivity to the agreement violations. However, the L1 speakers of Romance languages [+gender] and English [-gender] did not exhibit target-like P600 effects in response to gender errors in any condition. Sabourin suggested that similarity in gender-marking systems between L1 and L2 entails greater convergence on L2 processing of nominal agreement and attributed the lack of the P600 effect to representational deficits. Clahsen and Felser (2006) interpreted these findings to show that the L2 learners construct underspecified syntactic representations and that they do not reflect early automatic structure-building processes.

5.2.1 Support for SSH

Since its proposal in 2006, support for the *SSH* has come from studies exploring the processing of morphologically complex words (e.g., Silva & Clahsen, 2008) and inflected words (e.g., Neubauer & Clahsen, 2009). For instance, Silva and Clahsen (2008) investigated the use of morphological structure in processing inflected and derived word forms. L1 English speakers and advanced L2 learners of English (L1 Chinese and German) participated in masked priming experiments examining regular past-tense forms and deadjectival nominalizations. The results showed that the L1

---

[13] ERPs reflect the real-time electrophysiological brain activity of cognitive processes that are time-locked to the presentation of target stimuli (Morgan-Short et al., 2010).

[14] *P600s* reflect syntactic word-order and morphosyntactic processing difficulties (Morgan-Short et al., 2010). P600s occur approximately 600 -1000ms post-stimulus, eliciting late centro-parietal positivities (Tanner, Goldshtein, & Weissman, 2018). It reflects an attempt at repair or reanalysis of the violated morphosyntactic structure (Tanner et al., 2018). Evidence suggests that it involves the basal ganglia circuits and that it depends on the procedural memory system (Ullman, 2004, 2006).

speakers demonstrated efficient priming for both inflected and derived word forms. However, despite the fact that the L2 learners showed repetition-priming effects (like the L1 group), they showed reduced priming for derived word forms and no priming effects for inflected word forms. Silva and Clahsen suggested that L2 learners rely more on lexical storage and less on combinatorial processing of morphologically complex words than native speakers. Additionally, Neubauer and Clahsen (2009) examined the processing of irregular and regular participles in L1 German speakers and L2 German learners (L1 Polish). The results from a masked priming experiment showed facilitative priming effects for irregular and regular participles in the L1 group, however, L2 learners only showed priming effects for the irregular participles and not the regular participles. The authors interpreted this to indicate that L2 learners do not segment inflectional affixes from their stems during processing and that they rely more on lexical storage than on morphological parsing during processing. Additionally, they suggested that the later in life words are acquired, the more their processing will rely on direct lexical retrieval instead of grammatical computation.

Further support for the SSH comes from research examining the processing of L2 gender agreement. For instance, Keating (2009) investigated the processing of gender agreement in L1 Spanish speakers and beginner, intermediate and advanced L2 learners of Spanish (L1 English). The participants completed a GJT while eye-movement was recorded. The sentences contained N-Adj agreement at three syntactic domains, (1) within the DP (immediately following the noun), (2) in the VP (preceded by *es* "is" and the adverb *bastante* "quite, very"), and (c) in a subordinate clause (introduced by *si* "if" or *cuando* "when"). The results revealed that the L1 speakers were sensitive to gender agreement violations in all three domains, however, only the advanced learners showed sensitivity to agreement errors in the local domain (within the DP) and not errors located outside of the DP. The intermediate and beginner learners did not show sensitivity to gender agreement errors in any of the three domains. Keating suggested that learners can develop abstract gender in their L2 grammars and resemble L1-like processing for features that do not exist in the learners' L1. However, this is only within the DP, which is also subject to proficiency levels. Keating also claimed that distance is a key determinant of difficulty in noun-adjective gender agreement processing, positing that L2 learners experience difficulty computing gender agreement the more distant the adjective is from the head noun (even for highly proficient learners). Keating concluded that the findings support the SSH, stating that L2 learners have trouble computing syntactic dependencies across distant but not local domains due to deficits in processing, in turn suggesting that they rely on shallow processing strategies.

Despite the evidence presented in favour of the SSH, it has been subject to criticism. One criticism was that it was not clear how grammatical properties of the L1 might affect L2 processing. Steinhauer (2006) suggested that the findings from Sabourin (2003) provided evidence for L1 transfer rather than support for the SSH, because Sabourin did not claim that the L2 learners did not compute full syntactic analyses during sentence comprehension like L1 speakers. Therefore, Steinhauer believed that the study provided support for representational deficits in terms of transfer (e.g., Representational Deficit Hypothesis, Hawkins & Chan, 2005). Clahsen and Felser (2018) have since clarified that the tendency to underuse grammatical information is independent of a learner’s L1, but, particularly for L2 learners from typologically different L1 backgrounds, as these L1 effects are often absent.

Another criticism emerged from Dowens and Carreiras (2006), who commented on the absence of a discussion surrounding early proficient bilinguals or even balanced bilinguals, and what the SSH meant for these speakers (despite the relevance of these speakers to understanding L1/L2 processing differences). They stated that it would be necessary to know why the sentence parsing options available in the L2 are restricted to shallow strategies, under what circumstances these may occur, and what prevents adult L2 learners from eventually achieving L1-like parsing strategies. This was not directly addressed in Clahsen and Felser's commentary (2018), however, they proposed that exploring the SSH in relation to early and successive bilinguals, across the age of acquisition scale, is a promising avenue for further developing the hypothesis.

In their commentary, Clahsen and Felser (2018) addressed another criticism concerning the use of terminology to explain the differences between L1 and L2 processing. They discussed their use of labelling, where they spoke of “fundamental” and “qualitative” differences between L1 and L2 processing, maintaining that the representations that the adult L2 learners compute are shallower and less detailed than L1 speakers (Clahsen & Felser, 2006). However, they felt that these labels had been misinterpreted, where they acknowledged that the choice of these words was “unfortunate” (Clahsen & Felser, 2018, p. 695). Therefore, in its updated version, the

SSH emphasizes that differences between L1 and L2 processing are gradual, rather than qualitative (Clahsen & Felser, 2018).

More recent support for the SSH comes from studies examining L2 processing of gender agreement (e.g., Klassen, Ferreira, & Schwieter, 2021). For instance, Klassen et al (2021) interpreted their findings to show support for an adapted version of the SSH. They investigated gender agreement processing in L2 beginner learners of Spanish (L1 English), pre- and post-study abroad (SA) experience. The participants completed a SPR task examining D-N agreement violations. The results prior to the SA experience revealed that the participants were sensitive to gender agreement errors but only when the noun was masculine. In other words, reading times (RTs) were slower when the article was feminine (mismatched) than when the article was masculine (matched). However, this effect was no longer present in the results post-SA experience. This result indicates a shift away from integrating the syntactic details of gender agreement, resulting in the elimination of effect of gender post-SA experience. The results also revealed that the participants' RTs in the post-SA session were faster than the pre-SA session, showing an effect for time spent abroad. The authors interpreted these results to support the SSH, but with a few caveats. First, they suggested that post-SA experience, the participants were no longer affected by gender agreement mismatches, indicating that they were processing in a shallow manner. Second, they proposed that the learners integrated less information in their ongoing interpretation after spending time abroad, suggesting that (post-SA) the participants managed their shallow processing strategies and that they begin to maximize processing efficiency in their L2. The authors, however, do not suggest that these results indicate that L2 learners are processing in a deficient manner, rather, that the learning context demands shallow processing, where they may develop a greater reliance on certain cues (i.e., nongrammatical / lexical / contextual) for processing. They propose that there is a shift in their parsing strategies to be more communicatively focused as a result of the learning context.

The SSH was proposed to account for differences between L1-L2 processing that could not be accounted for by a shortage of computational resources or a lack of automatization in L2 processing (Clahsen & Felser, 2006, 2018). It is assumed that L1 and L2 speakers have the same processing architecture and metal processing mechanisms available, however, (even highly proficient) L2 speakers experience problems building and / or manipulating abstract syntactic representations. The SSH also holds that L2 speakers are guided more strongly than L1 speakers by semantic and / or surface-level information. Studies have shown support for the SSH, showing that L2 speakers rely more on lexical storage than on morphological parsing during processing (e.g., Klassen et al., 2021; Neubauer & Clahsen, 2009; Silva & Clahsen, 2008) and that L2 processing is qualitatively different from L1 processing (e.g., Keating, 2009).

In a similar vein, the Declarative Procedural (DP) model of lexicon and grammar (Ullman, 2001, 2004, 2005, 2015, 2020) posits a lack of L1-like grammatical processing in L2 speakers, who instead rely on different neurocognitive systems, in particular, those that involve lexical and semantic knowledge and processes. However, the two theoretical approaches, the SSH and the DP model, crucially differ in that the DP model holds that experience with the L2 eventually leads to the proceduralization of grammar, resulting in L1-like grammatical processing. The SSH does not propose such a qualitative change over time. The following section discusses the DP model proposed by Ullman (2001, 2004, 2005, 2015, 2020).

5.3 Declarative Procedural (DP) model

The second major over-arching theoretical approach to L2 processing is the Declarative Procedural (DP) model of lexicon and grammar (Ullman, 2001, 2004, 2005, 2015, 2020). The model posits that language learning, storage and use critically depend on two domain-general, cognitive, long-term memory systems - declarative and procedural memory (Ullman & Lovelett, 2018).

Declarative memory (DM), a temporal-lobe-based system (medial-temporal-lobe (MTL) and the hippocampus), subserves the learning, representation and processing of the mental lexicon as well as non-linguistic semantic knowledge (knowledge about facts) and episodic knowledge (knowledge about personal events/experiences), such as what a bicycle is and what you ate for lunch yesterday (Buffington & Morgan-Short, 2018; Ullman, 2001, 2005, 2020; Ullman & Lovelett, 2018). Information can be learned rapidly, with as little as a single exposure to the stimulus, yet additional and repeated exposures improve learning and retention

(Ullman, 2004, 2015, 2020). Knowledge in this memory system is mainly acquired explicitly, that is, available to conscious awareness, however, DM also underlies implicit (nonconscious) knowledge (Ullman, 2015), with research showing that the system is extremely flexible in what it can learn (Morgan-Short, Hamrick, & Ullman, 2022; Ullman & Lovelett, 2018).

In both L1 and L2, DM is considered important for learning and representing idiosyncratic linguistic knowledge and arbitrary associations (Ullman, 2015, 2020; Ullman & Lovelett, 2018). It is responsible for the learning of simple content words such as *cat*, including its phonological form, meaning, sub-categorization frames and the mappings between them (e.g., the sound-meaning mappings). Since DM is flexible in what it can learn, it may also be available for learning non-idiosyncratic, rule-governed aspects of language (Ullman, 2015, 2020). Therefore, just like simple words, one may be able to memorize complex forms such as *walked* or *the cat* as chunks together with their meanings, while, simultaneously, the underlying compositional rules should gradually be learned (Ullman, 2015; Ullman & Lovelett, 2018). For second language acquisition, language learning that depends on DM seems to improve during childhood, plateau in adolescence and early adulthood, after which it declines (Ullman, 2015, 2020; Ullman & Lovelett, 2018). Therefore, an older child or young adult may be better at learning in this system than a young child (Ullman, 2015; Ullman & Lovelett, 2018).

Procedural memory (PM) is less well understood than declarative memory, however, the workings of this memory system are becoming clearer (Ullman, 2015; Ullman & Lovelett, 2018). While DM underlies the mental lexicon, the PM system subserves aspects of mental grammar and is rooted in the frontal/basal-ganglia and its associated circuitry (Ullman, 2006, 2015, 2020). Procedural memory only underlies implicit knowledge (that is, unavailable to conscious awareness), where it enables the gradual implicit learning and processing of a wide range of activities and functions, including sequencing, rules, navigation and probabilistic categorization (Quam, Wang, Maddox, Golisch, & Lotto, 2018; Ullman & Lovelett, 2018). Learning in the PM system occurs gradually through repeated exposure, thus, typically slower than DM. However, what is eventually learned tends to become automatized, showing excellent retention, and seems to be processed more rapidly than knowledge in the DM (Ullman, 2020; Ullman & Lovelett, 2018).

While both L1 and L2 largely depend on DM for learning lexical and semantic knowledge, the mechanisms underlying the learning, representation and processing of grammar initially differ between L1 and L2 (Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Ullman, 2015). The DP model posits that in the L1, aspects of grammar, in particular rule-governed structure building, are generally subserved by PM. However, in the L2, particularly at lower levels of exposure and corresponding proficiency, these learners rely on lexical / ßsemantic processes (in DM) for the same functions (Hamrick, Lum, & Ullman, 2018; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; Stefaniak, Baltazart, & Declercq, 2021; Ullman, 2001, 2015). With increasing L2 exposure, experience and proficiency, this can lead to the proceduralization of grammar, where aspects of grammar may come to rely more on the same grammatical (procedural) memory system as those that underlie L1 grammar (Ullman, 2005; Ullman & Lovelett, 2018). Therefore, grammar should depend more on DM during early stages of L2 learning, but more on PM at later stages (Morgan-Short et al., 2022). Thus, grammar should tend to rely more heavily on PM for early-learned languages (L1 or L2) and more on DM for later learned languages (Morgan-Short et al., 2022). In sum, Ullman's DP model predicts that learners may display nativelike representations and processing of both the mental lexicon and grammar (Ullman, 2005, 2020; Ullman & Lovelett, 2018). However, the changes in the reliance from DM to PM is not due to some sort of "transformation", rather, it is a result of the gradual acquisition of grammatical knowledge in PM, which is increasingly relied on, with the accompanying decrease in dependence on any grammatical knowledge that was learned in the DM (Ullman, 2020).

The declarative and procedural memory systems interact with each other (Ullman, 2015, 2020). They are believed to subserve language learning and processing in a complementary fashion (Stefaniak et al., 2021). For L2 learning, the two systems can acquire the same or analogous knowledge, including knowledge of sequences and rules (Ullman, 2015). The two systems can also interact competitively, with Ullman (2015) describing it as a *seesaw effect*, where the dysfunction or attenuation of one system may lead not only to an increased dependence on the other system, but the enhanced functioning of this other system. There is also evidence suggesting that the learning of declarative knowledge may inhibit the learning of analogous knowledge in PM (Ullman, 2015).

In the framework of the DP model, DM is involved in the learning of lexical and semantic knowledge, and the activation of meaning. On the other hand, PM applies to grammar, which should hold across linguistic sub-domains such as phonology, morphophonology, morphosyntax and syntax (Morgan-Short et al., 2022; Ullman, 2015, 2020). Since PM should hold across linguistic subdomains, including syntax and morphosyntax, this system should be important for the processing of gender agreement violations (Morgan-Short et al., 2010).

5.3.1 Support for the DP model

Evidence supporting the DP model has largely come from studies using ERPs for exploring the neurocognition of language processing (Morgan-Short et al., 2010). Language-related ERP research often employs a violation paradigm for presenting linguistic stimuli, where the ERPs to a linguistic violation (e.g., lexical, syntactic, morphosyntactic) is compared to the ERPs response to a matched control word or structure (i.e., correct form vs. violated form) (Morgan-Short et al., 2010). Different types of violations, such as difficulties, disruptions and anomalies have been shown to produce particular ERP components in L1 and L2 speakers (Morgan-Short et al., 2010).

Research has shown that the DP model can account for the individual differences in L1, L2 and/or bilingual language processing, where it has been used to explain findings within morphosyntactic processing, examining features such as verb and phrase structure (e.g., Esfandiari, Nilipour, Nejati, Maftoon, & Khosrowabadi, 2020). For instance, Esfandiari et al (2020) examined verb processing in L1 French speakers and accounted for their findings under the DP model. In experiment one, 93 participants (children and adults) completed a GJT, where they were given verbs which evoked an action, followed by either typical or less typical options, e.g., the verb *to water* with *flowers* or *sister*. Half of the sentences included grammatical violations, including either an inversion or an agreement error. Experiment two employed similar methods and materials but included either typical or unusual (i.e., completely meaningless) options following the verb. The results showed that the participants relied on their PM for processing typical words following the verb and for grammatical violations, while the participants were seen to depend on their DM when the words following were less typical and unusual, detecting that these options were sometimes completely meaningless. They argued that when their grammar was explicitly tested, they needed to rely on their DM. When the grammar is violated or when the thematic role is not respected (i.e., the option is meaningless), the sentence is rejected by PM and DM must explicitly examine it. The authors posited that the DP model can be used to explain what is stored in each memory system and how these memory systems interact when one language is in use.

Studies have employed the ERP approach to examine L1 and L2 processing of gender agreement in noun phrases, however, few studies have directly accounted for their findings under the DP framework. Studies focusing on grammatical gender processing in L1 speakers only have reported a P600 effect, which in some cases is preceded by a LAN response (Barber & Carreiras, 2005; Caffarra, Janssen, & Barber, 2014; Hagoort & Brown, 1999). For instance, Hagoort and Brown (1999) investigated article-noun gender disagreement in L1 Dutch speakers. The violation of gender agreement was found to result in a P600 response, suggesting that the online processing of gender agreement information is a syntactic-form driven process, and not a conceptual/semantic process. Barber and Carreiras (2005) examined article-noun gender agreement violations in L1 Spanish speakers and found a LAN effect[15], suggesting that the speakers detected a mismatch between the morphosyntactic features when the nominal-phrase disagreed in gender. This was followed by evidence of reanalysis and repair processes which are reflected in the P600 component. Caffarra et al (2014) also examined article-noun disagreement in L1 Spanish speakers with results showing sensitivity to gender agreement violations. The findings indicated that the speakers detected a mismatch in gender in the nominal-phrase.

Research has also examined L2 gender agreement, focusing on article-noun agreement. For instance, Tokowicz and MacWhinney (2005) tested gender agreement violations in L2 Spanish learners (L1 English). The participants showed

---

[15] *LANs (*left anterior negativities) are interpreted as the early and initial anomaly detection which occurs around 300-500ms after encountering the violation. It is linked to rule-based automatic structure-building computations and is primarily over left hemisphere and/or anterior electrodes (Tanner et al., 2018). *LANs* depend on the grammatical/procedural memory system (Ullman, 2001, 2004, 2005).

P600s in response to article-noun gender agreement violations, even though they did not perform above chance on grammaticality judgments of the same sentences. The authors indicated that the L2ers were implicitly sensitive to these agreement violations. Sabourin and Stowe (2008) examined gender agreement violations in L1 Dutch speakers and L2 Dutch L1 German and L1 Romance language speakers. The authors violated the definite determiner in the NP and found that the L1 Dutch speakers evidenced a P600 effect. The L1 German speakers also showed evidence of the P600 effect, whereas the L1 Romance speakers, whose agreement systems differ from Dutch (i.e., binary gender system), did not. Sabourin and Stowe explained their results in terms of transfer effects, suggesting that when there is a degree of syntactic similarity between the L1 and L2, similar neural processing is possible in the two languages. Dowens, Guo, Guo, Barber and Carreiras (2011) investigated Spanish article-noun gender agreement violations in highly proficient L2 Spanish learners (L1 Chinese) and found that the L2ers evidenced a P600 effect. They argued that linguistic proficiency plays a role in L2 sentence processing, meaning that grammatical features that are not present in L1, such as gender, can be acquired at higher stages of proficiency in the L2. Foucart and Frenck-Mestre (2011) found similar results. They examined gender agreement violations in French L1 speakers and advanced L2 French learners (L1 German). The determiner in the DP was violated and the results revealed that both the L1 French speakers and L2 French learners evidenced a P600 response. The authors inferred that syntactic processing in L2 is affected by proficiency levels in the target language, in turn suggesting that once learners become more proficient, it may be possible for them to apply the rules of the French system procedurally and thus show evidence of online processing that is qualitatively similar to that of native speakers.

Few studies exploring grammatical gender agreement violations in L1 and L2 populations have directly accounted for their results under the DP model (Ullman, 2001, 2015). However, the DP model mainly focuses on the role of linguistic proficiency in L2 processing, and the role of L2 proficiency is highlighted in some of the afore-mentioned studies. It could be suggested that these results provide evidence to support the DP model, i.e., with increasing L2 proficiency, it can lead to the proceduralization of grammar, where aspects of grammar may come to rely more on the same procedural mechanisms as those that underlie L1 grammar (Ullman & Lovelett, 2018).

A study by Morgan-Short et al (2010) has accounted for its findings under the DP model. They tested noun-article agreement and N-Adj agreement using an artificial language learning paradigm, which was structurally similar to Romance languages (language: Brocanto2), and recorded ERPs. The L1 English adult participants learned to speak and comprehend the artificial language under either explicit or implicit training conditions. The participants fell into low and high proficient learners. The results revealed that the low proficient learners, who were trained implicitly, yielded N400s for both adjective and article gender agreement violations. The low proficient learners who were trained explicitly only showed an N400 effect for the N-Adj agreement violations. The authors suggested that the low proficient learners do not depend on processes that are used in L1 gender agreement processing, which is evidenced by the N400 and the absence of a P600 effect. It is likely that they rely on lexical / semantic processes and depend on DM for both adjective and article gender agreement processing at lower levels of proficiency (Ullman, 2001, 2005). However, the fact that the low proficient learners (who were trained explicitly) only showed N400s for N-Adj agreement violations may be because adjectives have richer semantic content than articles, therefore they may be easier for DM to process, resulting in more reliable N400s for noun-adjective than noun-article violations (Morgan-Short et al., 2010).

The high proficient learners, who were also trained implicitly and explicitly, evidenced N400s for noun-adjective agreement but showed a P600 effect for noun-article agreement violations. The authors suggested that noun-adjective agreement processing relies on lexical / semantic processing as shown by the N400s, while noun-article agreement processing depends on similar L1 processing mechanisms, at least to some extent, as evidenced by the P600 effect. The evidence of a P600 effect for noun-article but not noun-adjective violations could possibly be explained by the fact that articles are arguably more important grammatical roles than adjectives and other content words during online gender processing. The findings also suggest that at higher levels of L2 proficiency there may be an increased dependence on L1 grammatical mechanisms, including the reliance on procedural mechanisms to process gender agreement violations (Ullman, 2001, 2005). The evidence of a P600 effect for noun-article agreement in the high proficient learners is in line with the L2 processing literature, which has reported P600 responses in highly proficient L2 learners of Spanish (Dowens et al., 2011; Tokowicz & MacWhinney, 2005), French

(Foucart & Frenck-Mestre, 2011) and Dutch (Sabourin & Stowe, 2008). However, the findings for L2 learners are not entirely consistent, which is explained by L2 proficiency and/or L1-L2 similarity (Sabourin & Stowe, 2008).

The L1 gender processing literature, however, is consistent in reporting a P600 effect for noun-article and N-Adj gender agreement violations in different languages, including Spanish (Barber & Carreiras, 2005; Caffarra et al., 2014;), French (Foucart & Frenck-Mestre, 2011), Dutch (Hagoort & Brown, 1999; Sabourin & Stowe, 2008) and Chinese (Dowens et al., 2011). Under the DP model, consistent evidence of a P600 effect in L1 speakers reflects syntactic word-order and morphosyntactic processing difficulties, with these processes relying on the PM system to attempt repair or reanalysis of the gender agreement that has been violated (Ullman, 2004, 2006).

The DP model is a theoretical framework which is sometimes understood to make similar claims to the SSH (Clahsen & Felser, 2006, 2018), in that they both posit a lack of L1-like mechanisms for processing grammar in L2 speakers. However, this is only true to an extent, as Ullman maintains that this is only for initial stages of L2 processing (Ullman, 2001, 2004, 2005, 2015). Ullman (2001, 2004, 2015, 2020) claims that with increasing L2 exposure, experience and proficiency, L2 learners may become to rely on the same grammatical (procedural) memory system and mechanisms as L1ers to process grammar, while the *SSH* denies such a qualitative change over time (Ullman, 2006).

In contrast to the SSH and the DP model, identity-oriented approaches posit that the linguistic representations are fully engaged by L1 and L2 users alike, however, L1–L2 differences are related to performance differences or effects from the L1 (Hopp, 2022). Some argue that L2 users may not be able to fully compute and complete all processing steps as a result of the generally higher cognitive demands of processing an L2 (e.g., Hopp, 2010; McDonald, 2006). These approaches are outlined in the following section(s).

5.4 Cognitive processes in language processing

The third major over-arching approaches central to L2 processing, are the computational/cognitive capacity accounts (e.g., Hopp, 2010; McDonald, 2006).

Unlike the SSH (Clahsen & Felser, 2006) and the DP model (Ullman, 2001, 2004, 2015, 2020), the cognitive accounts advocated by Hopp (2010) and McDonald (2006) argue for quantitative differences, mainly attributed to cognitive resource limitations. Hopp (2010) and McDonald (2006) hold that processing in the L2 is inherently more taxing for memory resources than L1 processing, with increased cognitive demands hindering access to working memory and other attentional and decoding abilities. The non-target-like performance by advanced L2ers reduces to the overburdening of a L1-like processing system (Hopp, 2014), therefore reflecting quantitatively non-target-like processing in the L2 due to capacity differences, rather than fundamental differences in grammar for the parser (as per advocated by the SSH).

Capacity effects, such as working memory constraints, have been considered a factor in L2 processing, in that lower working memory may mask target-like processing among L2ers (Hopp, 2010; Reichle, Tremblay, & Coughlin, 2016). Working memory (WM) can be defined as the cognitive system(s) responsible for the control, regulation and active maintenance of information, in the face of potential distraction (Linck, Osthus, Koeth, & Bunting, 2014; Monnier, Boiché, Armandon, Baudoin, & Bellocchi, 2022). In the real world, it is one's ability to keep a small amount of information readily available for activities and decisions, including making statements, keeping track of conversations, navigating creative thinking and problem-solving (Logie, Camos, & Cowan, 2020).

Various theoretical models exist to explain its operation (Baddeley, 2000; Cowan, 2005; Unsworth & Engle, 2007), but its function remains similar across models: WM orders, stores, manages and processes immediate information in cognitive tasks, whilst ignoring irrelevant information in order to guide behaviour (Monnier et al., 2022). One of the most prevalent WM models was first proposed by Baddeley and Hitch in 1974 and has continuously developed since (Baddeley, Hitch, & Allen, 2020). The proposed and revised multi-component model holds a "non-unitary" view of WM with specialized processing and storage components, comprising an alliance of subsystems, each of which can be further fractioned (Baddeley, 2010; Baddeley et al., 2020; Serafini & Sanz, 2016). The components comprise of a supervisory attentional control system, the central executive (CE). The CE is responsible for various executive functions, such as inhibiting, controlling and allocating attentional resources among the cognitive process implicated in higher-level cognition, and retrieving information (Baddeley, 2010; Serafini & Sanz, 2016). It has the capacity to allocate resources, switch retrieval strategies and selectively attend to stimulus during the simultaneous execution of two tasks. It binds information from a number of various sources into a coherent episode by synchronising the working of the different stores (Baddeley, 2010).

The CE is aided by three subsidiary storage systems, the phonological loop, the visuo-spatial sketchpad, and the episodic buffer. The phonological loop handles phonological and verbal information. It has two major features, the first is a store in which speech-like memory traces are registered and spontaneously fade within about two seconds, and the second is a process where such traces can be refreshed via verbal or subvocal articulatory rehearsal process (Baddeley, 2003, 2010; Baddeley et al., 2020; Wen, 2016). The visuo-spatial sketchpad is capable of temporarily maintaining and manipulating visual and spatial non-verbal information in short-term memory (Baddeley et al., 2020). It plays a role in spatial orientation and in the solution of visuospatial problems, forming an interface between visual and spatial information which is accessed through the senses or from long term-memory (Baddeley et al., 2020). The episodic buffer is capable of holding multidimensional episodes or chunks, which may combine visual and auditory information (Baddeley, 2010; Baddeley et al., 2020). It acts as a buffer in that it provides a temporary store in which the various components of WM can interact and can interface with perception and long-term memory (LTM), with limited capacity of about four chunks or episodes (Baddeley, 2010).

The graphically presented model was first proposed by Baddeley and Hitch in 1974 and has since evolved (Baddeley et al., 2020). The stages of development of the multicomponent model are shown below (Figure 5.1: 1974 & Figure 5.2: 2012). The current version of the model, Figure 5.2, shows a speculative view of information from perception to WM (Baddeley et al., 2020).

Figure 5.1

WM model: 1974 (Baddeley et al., 2020, p.31, figure 2.2(a))

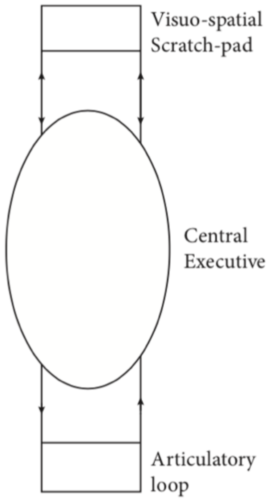
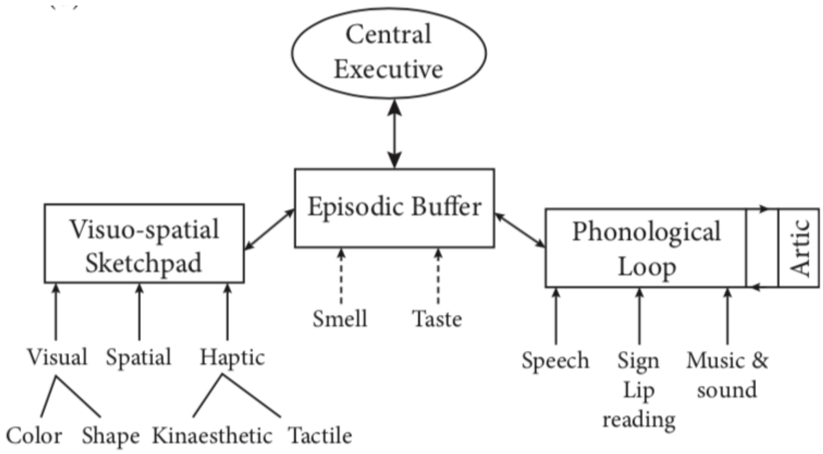


Figure 5.2

WM model: 2012 (Baddeley et al., 2020, p.31, figure 2.2(d))



This multi-component model has been a useful and powerful framework for addressing a range of questions on higher-level human cognitive activities, such as comprehension, complex learning, planning and reasoning (Wen, 2016). Working memory involves both the storage and processing of information, thus, several tasks and tools have been developed to measure WM capacity (WMC) and tap into both of these aspects, namely, simple and complex span tasks (Kim & Webb, 2022; Wen, 2016). Simple span tasks tap into the storage component of WM, measuring an individual's ability to store and rehearse information, while complex span tasks measure an individual's ability to store information while faced with additional processing tasks (Linck et al., 2014). Simple span tasks measuring phonological working memory (PWM) include non-verbal digit span (backwards and forwards), counting and letter rotation tasks, and verbal word, letter and non-word repetition tasks (Linck et al., 2014; Wen, 2016). Complex span tasks typically measure the CE component of WM, and these include verbal reading span, listening span and speaking span tasks, and non-verbal operation span, math span and N-back tasks (Linck et al., 2014; Wen, 2016).

The most popular complex span tasks used in language processing research include the reading span and the operation span tasks (Wen, 2016). Daneman and Carpenter's (1980) reading span task (RST) requires participants to read sets of sentences and remember an element for each sentence (e.g., a word of the sentence, a letter, or a numeral). As a secondary processing task, participants are required to provide judgments of the semantic acceptability or grammaticality of the sentence. This processing component obstructs the rehearsal of the elements to be recalled in memory, preventing the RST from being a measure of short-term memory (STM) alone. The operation span task (OST) (Turner & Engle, 1989), although similar to the RST, is a non-verbal measure of the CE component (Linck et al., 2014). Participants make judgements about simple arithmetic problems (i.e., solve maths equations) while storing words or letters, and are required to recall the final word(s) / letter(s) of each arithmetic equation in their presentation order. Complex span measures have been identified as more appropriate tools for assessing WMC than simple span tasks (Monnier et al., 2022), and are regarded stronger predictors of L1 comprehension (Daneman & Merikle, 1996) and L2 outcomes (Linck et al., 2014), with the latter suggesting that the executive control component of WM plays a larger role than short term memory when using an L2 (Linck et al., 2014).

5.4.1 Support for computational/cognitive capacity accounts

Empirical support for the role of WM in language comprehension comes from studies showing that speakers with stronger WMC are better able to learn vocabulary (in both L1 and L2), write more proficiently and have better L1 reading and listening comprehension (see Linck et al., 2014 for further details). Research also suggests that learners with larger WMC perform better in a range of L2 domains (e.g., oral production and processing morphosyntax) (Serafini, 2017), with some researchers maintaining that WMC correlates with syntactic processing ability (Antonicelli & Rastelli, 2022).

In previous studies, L1-like processing has typically been measured via sensitivity to agreement errors (Sagarra, 2019). WM has been used to predict sensitivity to agreement violations among L2 learners, with a view to identifying a subgroup of high-span L2ers that demonstrate target-like processing (Antonicelli & Rastelli, 2022; Reichle et al., 2016). In a study by Hopp (2010), they investigated the effects of proficiency and processing efficiency on L2 processing of structurally ambiguous input, disambiguated by means of case marking or verbal agreement, in advanced and near-native L2 learners of German (L1 English, Russian and Dutch). In two experiments, including a SPR task and a speeded GJT, the near-native group patterned with the L1 speakers in showing sensitivity at the points of agreement errors and inflectional disambiguation. However, the advanced learners did not to use morphosyntactic information during incremental parsing. They did not slow down upon encountering verbal agreement or case markers that ruled out the initial analysis of the sentence. Taken together, the computational demands of L2 processing leave insufficient resources for a complete parse. Therefore, L2 learners fail to integrate multiple information types as efficiently as monolinguals in real-time sentence processing, however, with increased proficiency, highly proficient L2 speakers are able to deploy morphosyntactic information to guide the parsing of structurally ambiguous input during real time. Increased cognitive capacity and proficiency reduces the computational demands of L2 processing, in turn facilitating the computation of fully syntactic parses. The findings from this study lead Hopp (2010) to propose that although L2 processing is cognitively more demanding than L1 processing, it is not qualitatively different, in turn disagreeing with the claims outlined by the *SSH*.

Coughlin and Tremblay (2013) investigated WM in L2 French learners' and found it played a role in the processing of local and non-local number agreement dependencies. The L2 learners of French (L1 English) and L1 French speakers completed a SPR task and a French RST. The results showed that the L1 speakers and highly proficient L2 learners were sensitive to the number agreement violations for both local and non-local dependencies, but not the lower-level proficiency L2 learners. The results also revealed that L2 learners with higher WM were more sensitive to number agreement violations in sentences containing a plural clitic in the short-distance condition and in sentences containing a singular clitic in the non-local condition, than L2 learners with lower WM.

Similar results were found by Rattanasak, Pongpairoj and Christianson (2022) who examined the role of WM in the processing of non-local subject-verb number agreement dependencies, in L2 learners of English (L1 Thai) and L1 English speakers. Results from a SPR task revealed that shorter distance agreement errors were easier to process for the L2 learners, but they had difficulty repairing non-local agreement dependencies, while the L1 speakers showed sensitivity to the agreement errors in both local and non-local conditions. The results also showed that WM, as measured by an RST, modulated the L2 learners' sensitivity to agreement violations. In other words, the learners did not show evidence of sensitivity to non-local agreement dependencies because they did not have the ability to maintain and hold this sensitivity when they were more distant.

There is also support for the role of WM in the processing of L2 gender agreement errors in online processing (e.g., Gabriele et al., 2021; Keating, 2010; Sagarra & Herschensohn, 2010). Sagarra and Herschensohn (2010) investigated L2 learners' sensitivity to gender agreement errors, violating the agreement between the noun and the adjective, in beginner and intermediate L2 learners of Spanish (L1 English), and L1 monolingual Spanish speakers. Results from an RST and a SPR task revealed that the intermediate L2 learners and L1 speakers were sensitive to the gender agreement violations, but the L2 beginners were not. The results also showed that WMC influenced sensitivity to gender agreement violations in the intermediate learners, revealing that L2 learners with higher WMC were more sensitive to gender agreement errors than those with lower WMC. However, there were no WM effects in the L1 speakers as a result of ceiling effects. The authors suggested that L2 learners with a certain proficiency level can develop processing patterns that are qualitatively similar to those of L1 speakers (contra to the *SSH*). They also suggested WM can be a significant factor and modulate gender agreement processing at the intermediate L2 level, but not the beginner L2 or L1 level, because WM effects may become irrelevant at ceiling and floor (low) levels. Taken together, WMC can facilitate gaining sensitivity to noun-adjective gender agreement in Spanish.

Working memory effects were also found by Keating (2010). Using eye-tracking technology, Keating examined the role of WM and linear distance on the processing of gender agreement in advanced L2 learners of Spanish (L1 English) and L1 Spanish speakers. Noun-adjective gender agreement was violated, manipulating the distance between the noun and adjective – placing either one, four or seven intervening words between the agreeing elements. Results revealed that distance influenced speakers' sensitivity to gender agreement errors, albeit at different points; with neither group showing sensitivity to the violations in the seven-word condition, only the L1 speakers showing sensitivity in the four-word condition, and both groups showing sensitivity in the 1-word condition. RST results also revealed higher-span L2 learners showed increased sensitivity to the gender errors, indicating that higher-span learners spent more time reading ungrammatical adjectives relative to grammatical controls. This arguably reflects a relationship between WM span and the detection of gender agreement violations across structural distance, suggesting that as linear distance increases, processing resources are consumed. These results also indicate that L2 learners can show L1-like processing, yet it is modulated by linear distance and thus, WM, as the information about the noun's gender could be "dumped" from WM by the time the adjective is encountered when the noun-adjective are not adjacent (Keating, 2010).

Similar results were found in a study by Gabriele et al (2021), who investigated number and gender agreement violations using ERP methodology in two groups of L2 Spanish novice learners (L1 English). The participants were grouped according to class instruction, with one group receiving basic instruction and the other at slightly more advanced levels of proficiency. Using an RSVP paradigm, L2 learners read sentences and judged their grammaticality. To test number and gender agreement, agreement was violated between the noun-adjective for number and for gender. The results revealed that both groups showed significant P600s for N-Adj number violations, however, only the more advanced group showed a marginally significant positivity to N-Adj gender violations. This suggests a trend toward emerging sensitivity to gender agreement violations but notable variability within the group, where sensitivity may be present in some individuals but not others. As a measure of WM, the participants also completed the Letter Number Sequencing subtest of the Wechsler Adult Intelligence Scale IV. The results revealed that learners with better WM showed significantly larger positivities for number violations, however, this was smaller for the gender violations, showing a close to significant effect (p = 0.071). Gabriele and colleagues posited that WM plays a role when linguistic features need to be encoded and maintained during online processing, which involves the ability to retrieve the relevant linguistic features from memory and the online deployment of knowledge of agreement.

In contrast to the afore-mentioned studies Foote (2011) did not find similar results. A SPR task testing N-Adj gender agreement was administered to early and late advanced Spanish-English and L1 Spanish speakers, investigating adjectival distance from the target noun (i.e., adjacent and separated). The results revealed that all participants showed sensitivity to N-Adj agreement errors, however, sensitivity was increased when the agreement error was adjacent, compared with when it was separated. There was also no evidence of WM effects, as measured by the RST (only to the late bilinguals). WMC did not influence sensitivity to gender agreement errors in the late bilingual group, at either distance, suggesting that gender agreement processing is not modulated by WMC in late Spanish bilinguals. Foote concluded that L2 processing is not qualitatively different from L1 processing, rather, advanced late bilinguals can develop the same types of knowledge in L2 as they have in their L1, however, WMC does not play a role.

Studies have shown evidence for the role of WM capacity in the processing of agreement in both L1 and L2 learners, for subject-verb (Hopp, 2010; Rattanasak et al., 2022), number (Coughlin & Tremblay, 2013; Gabriele et al., 2021) and noun-adjective gender agreement (Gabriele et al., 2021; Keating, 2010; Sagarra & Herschensohn, 2010). Collectively, these findings do not suggest qualitatively different processing mechanisms between L1 and L2 speakers, rather, they suggest that L2 learners can develop processing patterns that are qualitatively similar to those of L1 speakers, however, there are quantitative differences between L1-L2 processing, mainly attributed to cognitive resource limitations (Hopp, 2010; McDonald, 2006).

Within a similar vein, various other approaches have been proposed to counter the ideas argued by the SSH, that there are fundamental differences between L1-L2 processing. These include the *Full Transfer/Full Access/Full Parse Hypothesis* (Dekydtspotter, Schwartz, & Sprouse, 2006), the Lexical Bottleneck Hypothesis (Hopp, 2014, 2018) and the memory interference approach (Cunnings, 2017).

5.4.2 Additional L2 processing approaches countering qualitative differences

Unlike the SSH, which infers qualitative differences between L1 and L2 processing, Dekydtspotter et al (2006) proposed the 'Full Transfer/Full Access/Full Parse Hypothesis' (FTFAFP), which has since been updated to the *Full Parse Hypothesis* (Swanson & Dekydtspotter, 2022). The hypothesis posits that the mechanisms of L1 and L2 processing are the same, but the observed differences are due to extraneous factors to syntactic processing (i.e., different sources) and do not involve fundamentally different parsing mechanisms. It argues that L2 speakers can compute complete syntactic representations, but they may fail to complete these representations quickly or successfully due to an increased upstream processing load (Swanson & Dekydtspotter, 2022). To account for these differences, four potential sources of differences between L1 and L2 processing were proposed (Dekydtspotter et al., 2006).

The first potential source was non-target-like *prosody*. It was suggested that L2 speakers do not have target-like prosodic systems, in that their patterns of rhythm, sound, intonation, stress and tone differ to those of L1 speakers. Therefore, L2 speakers may impose non-target-like prosodies, yielding non-target-like outcomes. The second potential source involves *lexical access*. Dekydtspotter et al (2006) proposed that L2 speakers have under-learned lexical access routines. Evidence for this comes from eye-tracking studies where even advanced L2 speakers continue to re-read stimuli (Frenck-Mestre, 2002, 2005), therefore, indicating that L2 speakers lack automaticity and are therefore generally slower and less efficient processors than L1 speakers. The third potential source relates to *L2 lexicon*. Dekydtspotter and colleagues argued that L2 lexicon is not built from scratch from the target language input, rather, L2 lexical items first develop through the relabelling of L1 items with perceived phonetic strings from the TL input. It is only through error-driven reanalysis that the L2 items come to be similar in the TL, with regards to the precise morphosyntactic, semantic and pragmatic features. The authors do not imply that the L2 speakers will not arrive at the intended meaning, only that they are slowed down by the irrelevant options that would not occur to an L1 speaker. The final potential source relates to what they described as the *comparisons* of planned target segments. They explained that in comparison to L1 speakers, L2 speakers process relatively slowly and perhaps with an unhelpful prosodic system. L2 speakers also experience greater difficulty accessing a lexicon that may include non-target-like semantico-pragmatic features. Therefore, Dekydtspotter and colleagues argued that these different “target” segments may reflect different “moments” in the parsing process. Dekydtspotter et al (2006) maintain that L2 processing is congruent with the same kind of structurally based processing used by L1 speakers, however, the general slowness of L2 processing can arguably explain the differences between L1 and L2 linguistic behaviour.

Closely related is *The Lexical Bottleneck Hypothesis*, proposed by Hopp (2014, 2018). The hypothesis proposes that L2 learners are more restricted, compared to L1 speakers, in terms of their ability to consistently access and retrieve lexical information due to lower-quality lexical representations. The differences between L1 and L2 processing may be due to characteristics of bilingual processing and/or factors extraneous to syntactic processing, such as the stages that precede or subserve lexical parsing, in particular, lexical processing. On the basis that lexical processing precedes syntactic processing in the functional architecture of the language processing system, delays, differences, or cross-linguistic influence in this earlier stage (i.e., lexical processing) may lead to non-target-like syntactic processing. This slower lexical processing comes from the combination of weaker links, and non-selective lexical access could yield input for syntactic processing that is less robust, more diffuse, or even delayed. If the delays, differences or cross-linguistic influence were taken into account or even removed, adult L2 learners may come to display target-like syntactic processing in the L2. Hopp (2018) maintains that L1 and L2 processing are qualitatively similar, however, L1-L2 differences are in reality indirect effects of difficulty with lexical access, which likely result in non-target syntactic processing.

Although Cunnings (2017) opposes the SSH, their proposed approach, the *memory interference hypothesis*, can also be identified as a difference-orientated account, claiming that L2 processing differs from L1 processing with respect to the neurocognitive (memory) systems recruited during real-time comprehension. Cunnings attributes the differences between L1-L2 processing to divergent abilities to retrieve from memory information recently encountered during processing. It is argued that the primary source of difficulty lies in which the information that is not currently the focus of attention must be encoded, stored and retrieved when needed, rather than implicating shallow parsing (as posited by the SSH, Clahsen & Felser, 2006). Cunnings maintains that the differences between L1-L2 processing are not due to L2 speakers' failure to compute full syntactic parses, but rather to their increased susceptibility to interference during memory retrieval and greater reliance on discourse-based cues, as opposed to syntactic cues, during memory retrieval. L2 speakers arguably suffer from greater cue-based interference, because cues that are overtly marked on lexical items, such as morphosyntactic agreement features, are easier for L2 speakers to employ during parsing (Hopp, 2018). One of the underlying causes for the greater susceptibility of interference in L2 learners is related to the quality and extent of lexical processing, determining the detail with which lexical items are stored in memory (Cunnings, 2017; Hopp, 2018).

The theoretical approaches briefly discussed in this section provide further evidence arguing that L1-L2 processing is qualitatively similar, but the observed differences are attributed to extraneous factors (Dekydtspotter et al., 2006), lexical processing (Hopp, 2014, 2018) and/or memory interference (Cunnings, 2017). Although these do not argue quantitative differences in terms of capacity effects, these theoretical accounts argue that L2 learners can develop processing patterns that are qualitatively similar to those of L1 speakers, however, any L1-L2 differences are influenced by individual differences.

5.5 Chapter summary

This chapter provided an overview of the theoretical approaches central to L2 gender processing, including the SSH (Clahsen & Felser, 2006, 2018), the DP model of lexicon and grammar (Ullman, 2001, 2004, 2005, 2015, 2020), and cognitive capacity approaches advocated by Hopp (2010) and McDonald (2006). Additionally, other approaches within a similar vein were discussed, including the Full Parse Hypothesis (Swanson & Dekydtspotter, 2022), the Lexical Bottleneck Hypothesis (Hopp, 2014, 2018) and the memory interference approach (Cunnings, 2017). These were discussed to better understand how bilingual speakers process sentences during real-time. It may possible that the current Welsh speaker processing data can be explained under one of these theoretical approaches, or particular aspects of one (or more) of the approaches. Understanding the Welsh speaker results in light of the theoretical approaches can help extend current processing literature to Welsh, a minority language that is in constant contact with English and within the broader field of bilingualism. The following chapter presents experiment two, investigating the processing of gender in Welsh-English adult bilinguals and the role of various individual difference variables.

Chapter 6
Experiment 2: The Processing of Welsh Grammatical Gender

This chapter presents the second experiment and extends the findings of experiment one, by investigating the processing of Welsh grammatical gender in Welsh-English bilingual adults. The first section describes the methodology used to empirically test the processing of Welsh grammatical gender (see section 6.1). The second section presents the results, detailing the participants' sensitivity to grammatical gender violations and the role of the individual difference variables under investigation (see section 6.2). The final section presents a brief discussion of these findings (see section 6.3).

6.1 Methodology: Overview

Section 6.1 outlines the methodology used to investigate the processing of Welsh grammatical gender in Welsh English bilingual adults. Details are provided on the research questions and hypotheses, participants, procedure, task details and the analyses employed.

6.1.1 Research Questions and hypotheses

This second experiment extends the findings from production and comprehension to the processing of Welsh gender, investigating Welsh speakers' representation of gender during real-time and considers the possible role of various additional individual difference variables. It investigates whether the patterns emerged from the first experiment, particularly the production data, is replicated in real time processing (Slabakova et al., 2020). Additionally, it explores whether the processing patterns emerging from this experiment are comparable to L1 or L2 processing patterns in speakers of other gendered languages. Therefore, the following research questions guide this second experiment:

1. To what extent do Welsh-English adult bilinguals show sensitivity to gender agreement violations when gender is independent of mutations, in conjunction with mutations and/or encoded through mutations, in local-contexts?
2. To what extent do cognitive and environmental individual differences play a role in Welsh-English adult bilinguals' processing of gender in Welsh?

Previous studies have consistently found monolingual L1 speakers to show sensitivity to noun-adjective (Barber & Carreiras, 2005; Keating, 2010; Sagarra & Herschensohn, 2010) and determiner-noun (Barber & Carreiras, 2005; Foucart & Frenck-Mestre, 2011; Hagoort & Brown, 1999) gender agreement violations, whereas studies investigating L2 adult learners have found mixed results (e.g., Keating, 2009; Klassen, Ferreira, & Schwieter, 2021; Sagarra & Herschensohn, 2010). However, the Welsh speaking participants in this study are simultaneous, early sequential, and late bilinguals of Welsh and English (late bilinguals having acquired the subsequent language at ages 7-11). Early bilinguals, given a sufficient quantity of input, are claimed to show characteristics similar to L1 processing (Foote, 2011; Guillelmon & Grosjean, 2001; Silvina Montrul et al., 2008; Ullman & Lovelett, 2018).

Therefore, it is hypothesised that the Welsh-English bilingual adults will demonstrate L1-like processing patterns, showing sensitivity to gender agreement violations. The Welsh speakers will emulate L1-like parsing strategies, where they fully compute complete syntactic representations during real time sentence processing (contra to the *SSH*, Clahsen & Felser, 2006). Following previous research (e.g., Barber & Carreiras, 2005; Foucart & Frenck-Mestre, 2011; Hagoort & Brown, 1999), the Welsh speakers will show sensitivity to gender agreement violations when gender is encoded via the determiner 'y', as this is the most common instantiation of gender encoded through mutations. Moreover, following patterns observed in experiment one, the Welsh speakers will show stronger sensitivity to gender agreement violations when gender occurs with no mutations (i.e., testing this instantiation using the numeral 4) than when gender is in conjunction with mutations (i.e., testing this instantiation using the numeral 2). It is also possible that they will show sensitivity to the agreement violations when mutations are independent of gender. Evidence will not indicate shallow or less detailed processing patterns in the Welsh-English bilingual speakers.

In line with previous research, it is hypothesised that certain individual differences will play a role in the processing of Welsh grammatical gender, namely, Welsh linguistic proficiency, language dominance and the number of years spent using Welsh in an educational and/or professional setting. Welsh proficiency is predicted to play a role in light of the findings from experiment 1, which showed that proficiency predicted outcomes on comprehension and production measures in Welsh bilinguals. Language dominance is hypothesised to play a role in the processing of Welsh gender because dominance in a language points to greater speed, accuracy, and automaticity in their dominant than in their non-dominant language (Birdsong, 2006). Therefore, speakers reflecting Welsh as their dominant language may show stronger sensitivity to Welsh gender agreement violations. Additionally, the number of years spent using Welsh in an educational and/or professional setting, is also predicted to play a role in the processing of Welsh gender. This is because the results form experiment one indicated that there could be more informative and different factors relevant to Welsh bilingual speakers, than maybe the traditional individual differences. Considering this factor is exploratory, however, it may prove itself to be a step in the right direction.

It is also hypothesised that working memory, procedural memory, declarative memory, and Welsh linguistic proficiency will not play a role in the processing of Welsh gender. This is because research has shown that working memory effects become irrelevant at ceiling in L1 speakers, meaning that if the Welsh bilinguals demonstrate L1-like processing patterns (thus show ceiling effects), working memory is no longer a significant factor (e.g., Sagarra & Herschensohn, 2010). However, L2 adult speakers with higher working memory span scores have shown to be more sensitivity to gender agreement violations (e.g., Sagarra & Herschensohn, 2010). Procedural memory is assumed not to play a role, given the fact that the Welsh bilinguals are predicted to emulate similar processing patterns to L1 speakers of other gendered languages, and would therefore be at ceiling. Finally, declarative memory is also assumed not to play a role, because with increasing L2 exposure, experience, and proficiency, L2 learners may become to rely on the same procedural memory system and mechanisms as L1ers to process grammar, and the Welsh-English bilinguals in this study emulate L1-like 'categorisation' more than L2 adult learners. The methods used to address these research questions and to test these predictions is outlined below.

6.1.2 Participants

In experiment one, forty self-reported Welsh-English bilingual adults participated – ten males, 29 females and an individual who identified as non-binary (age *M*=34, range 19-64). All forty participants were contacted via email to voluntarily participate in this follow up experiment (which took place approximately 14 months after experiment 1), of which twenty-one participated (age *M*=37, range 19-64, five males, 16 females).

The 21 participants are highly fluent Welsh speakers (as measured by the cloze test in experiment 1: *M*=95%, range 84-100%) and are highly educated, all having obtained (n=20) or working towards an undergraduate degree (n=1), of which twelve hold a master's degree and half of these hold a PhD (n=6). Notably, 14 of the 21 participants work in Welsh professionally, for instance, English-Welsh translators, Welsh lecturers or Welsh medium secondary school teachers. Sixteen of the twenty-one participants acquired Welsh from birth, while four participants acquired Welsh from the age of 3 and one participant from the age of 4. Only ten of the participants recorded that they acquired English from birth, while five of the participants acquired English from the age of 3, two from the age of 4, one from the age of 6, one from the age of seven and two from the age of 8. According to their global dominance scores as indicated by the BLP (at the time of testing for experiment 1), eleven out of the 21 participants are Welsh dominant (*M*=-58.15, range: -118.06 – -30.25), seven are English dominant (*M*=64.68, range: 23.16 – 146.88) and three are considered balanced bilinguals (*M*=-4.93, range: -18.80 – 19.63).

6.1.3 Battery of tasks and procedure

A battery of tasks was administered to the participants to investigate the processing of grammatical gender in Welsh and the possible role of individual differences, including working, declarative and procedural memory. The battery included six tasks, and these were completed in the following order:

- Self-paced reading task (SPR: part 1, online reading measure)
- Operation span task (OSPAN, working memory task)
- Continuous visual memory task (CVMT, declarative memory task)
- Self-paced reading task (SPR: part 2, online reading measure)
- Towers of Hanoi (TOH, procedural memory task)
- Offline gender decision task (Welsh gender assignment task)

The data were collected online via the software Gorilla.sc. Participants completed the experiment on their devices on their own. Data collection lasted approximately 45 minutes and participation was voluntary. Participants first read an information sheet and signed a consent form (in either Welsh or English). All instructions were presented bilingually, and they were provided with a bilingual debrief sheet upon completion.

Data collection took place during March, April, and May (2022). Each participant was given the right to withdraw at any given time, and all participant data were anonymised with each participant given an ID code in line with GDPR. The study was granted ethical approval by the Faculty of Humanities and Social Science Research Ethics Committee (approval number 250222/4936) and complied with the ethical guidelines for conducting research with adults, as outlined in the university's guidelines.

The following sub-sections describe each task in turn, outlining the stimuli and how the data were used to address the research questions posed in this experiment.

6.1.4 Self-paced reading task

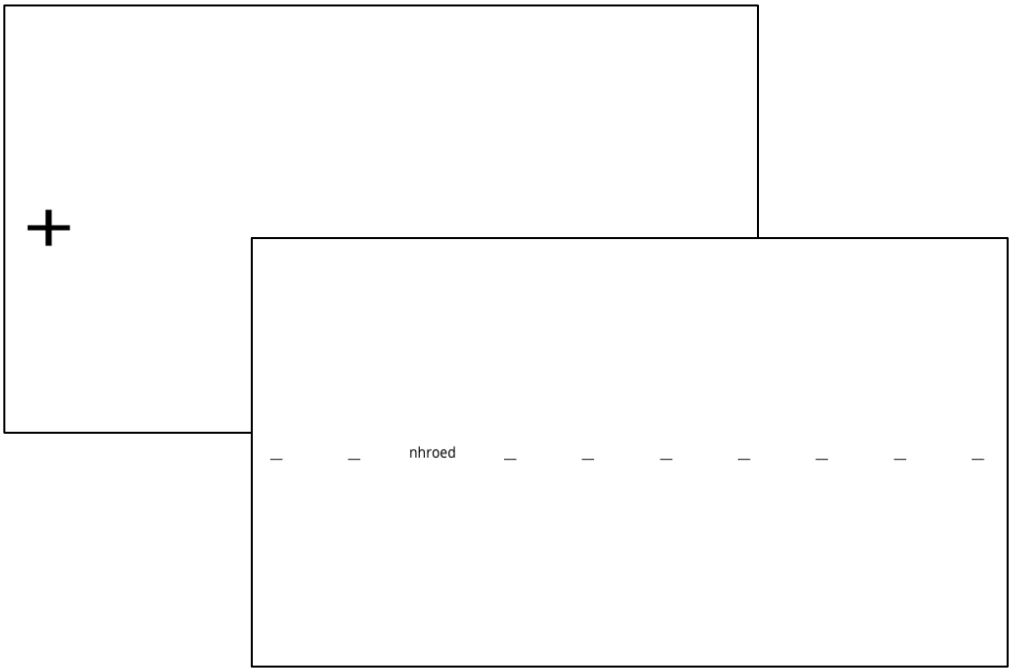
The primary task in this experiment was a self-paced reading task (SPR task) (Just, Carpenter, & Woolley, 1982). Self-paced reading tasks involve capturing information as the participant incrementally processes a sentence, allowing for a "fine-grained" analysis of online processing during real time (Keating & Jegerski, 2015, p.2). An underlying assumption of this technique is that the time taken to read a word is a direct reflection of the time taken to process the specific phenomenon, by comparing participants' RTs in two minimally different versions of a particular linguistic feature (Marsden, Thompson, & Plonsky, 2018; Slabakova et al., 2020). Collecting RTs allows researchers to determine with precision the moment where processing difficulty (longer RTs) or facilitation (shorter RTs) arises. The focus of the current SPR task is the notion of anomaly detection, namely, the processing of gender disagreement (Renaud, 2014; Sagarra & Herschensohn, 2010).

Simple alterations can be used to see whether speakers of gendered languages are sensitive to the change in agreement, for instance, violating the agreement between a noun and an adjective, so that they no longer agree in gender. Reading times (reaction time data) can help determine whether or not they have constructed an abstract representation of the gender system without intentionally drawing on awareness or explicit knowledge (Keating, 2009; Marsden et al., 2018). However, a lack of sensitivity to the gender errors may suggest evidence of a divergent representation (Keating, 2009). Data from SPR tasks claim to provide a window into implicit processing, ultimately telling the researcher whether or not the participant can compute linguistic representations during real-time processing (Slabakova et al., 2020). The way in which the task is created and administered is outlined below.

Following previous methodologies and recommendations (Jegerski, 2014; Jiang, 2004, 2007; Keating & Jegerski, 2015; Marsden et al., 2018; Slabakova et al., 2020), the word-by-word non-cumulative linear paradigm was employed, which is also referred to as the moving window paradigm. Here, each button press causes the 'window' of visible text (individual words) to move across the screen with the proceeding word disappearing upon appearance of the following word (Jiang, 2007; Marinis, 2010). Each sentence began with a 500ms fixation cross in order to land the participant's gaze at the beginning of the sentence on the left side of the screen (Jegerski, 2014). Each word in the sentence was represented by a dash line, with spaces in between each dash to present visible characteristics of a sentence (Marinis, 2010).

Figure 6.1

SPR Window Example



Another fundamental consideration in designing SPR tasks, is item design (Keating & Jegerski, 2015; Marsden et al., 2018). For inferences about the nature of processing under investigation to be meaningful, task items were carefully designed. To test the phenomenon of interest (grammatical gender agreement), there were two overarching groups of items: experimental conditions and non-experimental conditions. The experimental conditions included four different conditions. The non-experimental conditions included distractor conditions and filler conditions (50:50).

There was a total of 64 items made up across four experimental conditions (16 items per condition). The four conditions represented four contexts to disentangle gender from mutations. Experimental condition 1 tested gender in conjunction with mutations, by including the gendered forms of the numeral two (dwy [f] / dau [m]), triggering SM on post-numeral nouns with words beginning with a consonant participating in the mutation system (see section 2.4). Testing this aspect of the system provided information on the processing of gender when the two systems operate in conjunction with one another. Below are two examples (masculine & feminine).

Experimental condition 1

Example 6.1 Roedd dau bengwin[16] *yn sefyll yn stond yn y môr*

Be PST two Num (m) penguin N (m) PART standing VERB PART still ADJ in PREP the DET sea N (m)

'There were two penguins standing still in the sea'

Example 6.2 Roedd dwy wiwer yn casglu cnau ar gyfer y gaeaf

Be PST two Num (f) squirrels N (f) PART collecting VERB nuts N (f) for PREP the DET winter N (m)

'There were two squirrels collecting nuts for the winter'

Experimental condition 2 tested gender independent of mutations, by including the gendered forms of the numeral four (pedair [f] / pedwar [m]), where post-numeral nouns are produced in their bare form. Testing this aspect of the system provided information on the processing of gender when the two systems are independent from one another. Below are two examples (masculine & feminine).

Experimental condition 2

Example 6.3 Bydd pedwar cogydd yn brysur yng nghegin y tŷ bwyta

Will be four Num (m) chefs N (m) PART busy VERB in PREP kitchen N (f) the DET restaurant N (m)

'Four chefs will be busy in the restaurant kitchen'

Example 6.4 Mae pedair bydwraig ar y ward yn yr ysbyty leol

Be 3SG four Num (f) midwives N (f) on PREP the DET ward N (f) in PREP the DET hospital N (m) local ADJ

'There are four midwives on the ward in the local hospital'

Experimental condition 3 tested mutation independent of gender, where the SM can be seen on the noun regardless of the noun's gender. This is done by including pre-nominal adjectives triggering SM on following nouns (irrespective of gender). The adjectives used were '*hen*' (old), '*annwyl*' (dearest), '*prif*' (main/head/chief), '*hoff*' (favourite) and '*unig*' (only/lonely). This condition disentangled mutations from gender. This condition is similar to condition 1, in the sense that it is mutation regardless of grammatical gender, however, it is different

[16] In Welsh, there is no number agreement in the full DP where the head noun in the numeral phrase remains singular (involving numbers 1-9). However, there is the option of number + preposition 'o' + plural form, e.g., twenty men 'ugain dyn' (twenty men) or 'ugain o ddynion' (twenty of men). Also, there is number agreement with pronouns, e.g., 'Mae Fred a Sandra yn yfed coffi. Maen nhw'n ffrindiau' (Fred and Sandra is (are) drinking coffee. They are friends).

because they do not have to process a gendered form. Below are two examples (masculine & feminine).

Experimental condition 3

Example 6.5 Roedd hen gerdyn ar waelod y cist yn yr atig

Be PST old ADJ card N (m) on PREP bottom N (m) the DET chest N (f) in PREP the DET attic N (m)

'The old card was at the bottom of the chest in the attic'

Example 6.6 Mae'r brif afr yn arwain y gweddill i'r nant fach

Be 3SG+the DET chief ADJ goat N (f) PART leading VERB the DET rest to PREP+the DET stream N (m) small ADJ

'The chief goat is leading the rest to the small stream'

Experimental condition 4 tested gender encoded through mutation, or the absence of a mutation. This was done by using the definite article '*y*', which unlike many other gendered languages, does not distinguish gender. However, it causes a different mutation depending on gender in the subsequent noun. The definite article '*y*' triggers SM on post-determiner feminine nouns, whereas no mutation is triggered on post-determiner masculine nouns (i.e., produced in bare form). This condition tested the most common instances of gender and mutation in the Welsh language where the target nouns are singular in form, and mark gender via SM [F] or no mutation [M]. Below are two examples (masculine & feminine).

Experimental condition 4

Example 6.7 Bydd y cartŵn ar y sianel Disney yn ddoniol iawn

Will be the DET cartoon N (m) on PREP the DET channel N (f) Disney PN PART funny ADJ very ADV

'The cartoon on the Disney channel will be very funny'

Example 6.8 Roedd y gigfran yn hedfan yn swnllyd. uwch fy mhen

Be PST the DET raven N (f+SM) PART flying VERB PART noisy ADJ above PREP my PRON head N (m)

'The raven was flying noisily above my head'

The non-experimental conditions were divided into two categories, distractor conditions and filler conditions. The items in the distractor conditions were intentionally designed to contain similar linguistic forms and structures to the items in the experimental conditions, serving to counterbalance the characteristics of the experimental conditions to avoid making them stand out to the participant (Keating & Jegerski, 2015; Marsden et al., 2018). Within the distractor conditions, there were two different linguistic conditions. Half of the items in the distractor conditions (n=16) included the gendered forms of the numeral two (dau [m] / dwy [f]), however, unlike the items in the experimental condition, the nouns following ‘two’ started with letters that do not participate in the mutation system. These letters were *a, s, n* and *ff.* Two examples are shown below.

Distractor condition 1

Example 6.9 Roedd dau siocled plaen ar y bwrdd yn y gegin

Be PST two Num (m) chocolate N (m) plain ADJ on PREP the DET table N (m) in PREP the DET kitchen N (f)

‘*There were two plain chocolates on the table in the kitchen”*

Example 6.10 Bydd dwy ffatri fach yn cau eleni oherwydd diffyg gweithwyr

Will be two Num (f) factory N (m) small ADJ PART closing VERB this year ADV because CONJ lack of N (m) workers

‘Two factories will close this year due to a lack of workers’

The other half of the items in the distractor conditions (n=16) included the gendered forms of the numeral three (tri [m] / tair [f]) where post-numeral masculine nouns AM if they begin with *c*, *p* or *t*, while feminine nouns are produced in bare form (i.e., not mutated). Two examples are given below.

Distractor condition 2

Example 6.11 Mae tri cheffyl yn carlamu yn y cae bach heddiw

Be PST three Num (m) horse N (m) PART galloping VERB in PREP the DET field N (m) small ADJ today ADV

‘There are three horses galloping in the small field today’

Example 6.12 Roedd tair cwcw yn canu’n hapus ar gangen y goeden

Be PST three Num (m) cuckoo N (f) PART singing VERB happy ADJ on PREP branch N (m) the DET tree N (m)

‘Three cuckoos were singing happily on the branch on the tree’

In total, there were 32 items in the distractor conditions, 16 including the gendered numeral forms for *two* and 16 including the gendered numeral forms for *three.* In addition to this, there were 32 items in the filler conditions.

The items in the filler conditions were unrelated sentences that were not intended to elicit any specific type of processing effects (Keating & Jegerski, 2015; Marsden et al., 2018). These items contained plural nouns which were all preceded by the definite article ‘*y*’ (the). Plural nouns following *‘y’* are not associated with mutations and are produced in their bare form (i.e., not mutated). All nouns began with consonants which participate in the mutation system (*p, t, c, b, d, g, m, ll, rh)* to remain consistent with the items in the experimental conditions.

Additionally, to include variety in the stimuli, half of the items in the filler conditions included non-local gender marked contexts with third-person possessive forms. Recall that if the antecedent noun is masculine, '*ei*' triggers SM and if the antecedent noun is feminine, '*ei*' triggers AM. However, if the antecedent noun is plural, '*eu*' does not trigger a mutation on the following word. Therefore, the 16 items with non-local gender agreement included the plural third-person possessive form '*eu*' to agree with the plural nouns. The purpose of these items was to divert participants' attention from the items experimental conditions (Keating & Jegerski, 2015). In total, there were 32 items in the filler conditions, all of which contained plural nouns, with 16/32 items including non-local agreement. Examples of the two types of filler items are given below.

Non-critical filler condition

Example 6.13 Roedd y cantorion yn nerfus cyn perfformio ar y llwyfan
Be PST the DET singers N PART nervous VERB before CONJ performing PREP on PREP the DET stage N (m)
'*The singers were nervous before performing on the stage'*

Example 6.14 Daeth y breninesau o bell i drafod materion pwysig iawn
Come PST the DET queens N from PREP far ADV to discuss VERB matters N important ADJ very ADV
'The queens came from afar to discuss very important matters'

Non-critical filler condition with non-local agreement

Example 6.15 Roedd y crysau yn edrych yn smart gyda'u botymau sgleiniog
Be PST the DET shirts N PART looking VERB PART smart ADJ with PREP its DET buttons N shiny ADJ
'*The shirts looked smart with their shiny buttons'*

Example 6.16 Collodd y peli siap ar ôl iddynt gael eu cicio
Lost PST the DET balls N shape ADJ after CONJ to them POSS-3PL have VERB their POSS-ADJ N kicked VERB
'*The balls lost their shape after they were kicked'*

In total, there were 128 items, 50% experimental conditions (four conditions) and 50% non-experimental conditions (25% distractor, 25% filler), which can be found in appendix H. This 50:50 experimental to non-experimental conditions ratio was in keeping with the recommendation posited by Keating and Jegerski (2015), who suggest that at least 50% of items are non-experimental, with the preferred number to be 75%. However, the current SPR task includes four experimental conditions to disentangle gender from mutations, which in itself serve as distractors to each other. This factor combined with practical limitations of the number of sentences that participants can read before fatigue or boredom, balances out the necessary number of experimental to non-experimental conditions (Keating & Jegerski, 2015).

To maintain consistency across all of the items (Keating & Jegerski, 2015; Marinis, 2010), all sentences were 10 words in length and the target noun always appeared as the third word in the sentence. In the four experimental conditions and the non-experimental distractor conditions, the target nouns were two syllables in length, to avoid reading times differing drastically between participants (i.e., longer words take longer to read than shorter words) (Marinis, 2010). Additionally, the spill-over segments were similar across the four experimental conditions and the non-experimental distractor conditions (Keating & Jegerski, 2015), where the word immediately following the critical segment (i.e., the target noun) was always one syllable in length. In the non-experimental filler conditions, the target noun also appeared in the same position (3rd word) but varied in syllable length – 1 to 3 syllables depending on the pluralisation of the noun, with spill-over segments being either one or two syllables in length.

There were 128 nouns included in the task. Fifty-two of the nouns were taken from experiment one. Fifty-eight nouns were taken from the Welsh word lists (Morris & Meara, 2019), which are considered core vocabulary for beginner and intermediate learners of Welsh. An additional 18 nouns were used to create items, to balance gender and animacy. These nouns were checked for frequency using the KWIC tool on CorCenCC (Knight et al., 2020), which showed that in general, the 18 nouns were infrequent, with words per million ranging from 0.149 – 15.75. The 18 nouns were not found in the 'Geirfan' which includes 500 of the most frequent words in the Welsh language, designed for use by learners at A1/A2 levels of proficiency, or the 'Amliadur' which includes frequency lists for contemporary Welsh (Knight, Morris, Tovey-Walsh, Fitzpatrick, & Anthony, 2020). The 18 nouns and their frequencies can be found in appendix H.

The task consisted of one independent variable (grammaticality) composed of two levels (grammatical and ungrammatical). Therefore, experimental doublets were lexically matched at the two levels to maintain a high-level internal consistency within an item (Keating & Jegerski, 2015), where all 128 items had grammatical and ungrammatical versions. This means that the difference between the two versions in experimental condition 1 (numeral '2') and experimental condition 2 (numeral '4') was the violation of the gender agreement between the numeral and the target noun. For instance, the feminine gendered numeral form (*dwy)* was replaced by the masculine gendered numeral form (*dau),* and vice versa. This was also the case for the numeral four (pedwar>pedair / pedair>pedwar). Although the violation itself is the gendered numeral form, this violation is only apparent to the reader when they see the target noun (3rd word) in the sentence, as this is when they would be able to detect whether or not there is a gender agreement violation. In experimental condition 3, the violation was seen on the target noun by changing the SM noun to be produced in its bare form (e.g., g > c). The presence or the absence of the mutation on the noun itself was the violation, given that the pre-nominal adjective triggers SM on the noun irrespective of the noun's gender. In experimental condition 4, the violation was seen on the target noun by changing SM feminine nouns to bare form and masculine nouns to be SM. Similar to experimental condition 3, the presence or the absence of the mutation on the noun itself was the violation, given that the determiner 'y' triggers SM on following feminine nouns and no mutation on masculine nouns

For all of the experimental conditions, participants would be expected to slow down at the critical segment (i.e., on the target noun), producing slower reading times in identifying the gender agreement violation between the numeral / adjective / determiner and the target noun. However, participants might slow down in following spill over segments, e.g., immediate spill over, spill over +1 and/or spill over +2, showing slightly delayed processing effects (Avery & Marsden, 2019). This could possibly be due to the heavier processing costs associated with the complexity of the Welsh gender and mutations systems, or because the error itself is seen on the noun. Therefore, three spill-over segments were considered as these segments were very short (i.e., 1 syllable in length).

For the items in the non-experimental distractor conditions including the number *two*, the gender agreement between the numeral and the target noun was violated (e.g., replacing *dwy* for *dau).* For the distractor items including the numeral three, the gendered numeral forms were replaced with the incorrect form (*tri* replaced with *tair*), in addition to violating the mutation on the following noun (*tri* > AM, *tair* > no mutation). An example of the numeral three violation is shown below.

Distractor condition 2 'Three'

Grammatical

Example 6.17 Mae ***tri*** ***ch****effyl yn carlamu yn y cae bach heddiw*

Be PST three Num (m) horse N (m) PART galloping VERB in PREP the DET field N (m) small ADJ today ADV

'There are three horses galloping in the small field today'

Ungrammatical

Example 6.18 Mae ***tair*** ***c****effyl yn carlamu yn y cae bach heddiw*

Be PST three Num (f) horse N (m) PART galloping VERB in PREP the DET field N (m) small ADJ today ADV

'There are three horses galloping in the small field today'

The violation in the non-experimental filler conditions was the first verb in the sentence. In the grammatical versions of the items, the noun was plural, and the verbs were 3rd person singular in form. However, in the ungrammatical versions, verbs were violated for person, which were shown in the suffix of the verb to reflect 1st person singular (*'wn'* & *'ais'*), 2nd person singular *('et'* & *'aist'*), 2nd person plural ('*ech*' & '*och*') and 3rd person plural ('*en*'). The ungrammaticality came before the critical segment of interest in the non-experimental filler conditions. For instance:

Non-critical filler condition 'Plural'

Grammatical

Example 6.19 Roedd y cantorion yn nerfus cyn perfformio ar y llwyfan

Be PST the DET singers N PART nervous VERB before CONJ performing PREP on PREP the DET stage N (m)

'*The singers were nervous before performing on the stage'*

Ungrammatical

Example 6.20 Roeddet y cantorion yn nerfus cyn perfformio ar y llwyfan

Be PST the DET singers N PART nervous VERB before CONJ performing PREP on PREP the DET stage N (m)

'*The singers were nervous before performing on the stage'*

Another important aspect of a self-paced reading task is the post-stimulus distractor task, which provides a secondary offline measurement that is related to each item (Keating & Jegerski, 2015). Following previous methodologies and recommendations (Jegerski, 2014; Keating & Jegerski, 2015; Marinis, 2010;

Marsden et al., 2018; Slabakova et al., 2020), binary-choice comprehension questions (CQs) in Welsh were included after the experimental and non-experimental distractor conditions (75% of the items). They were not included after non-experimental filler conditions because of practical limitations (i.e., time and fatigue). The CQs were included after the non-experimental distractor conditions to prevent the participant from noticing the experimental conditions (Keating & Jegerski, 2015). The CQs appeared for a maximum of 15,000ms or until the participant responded.

The purpose of this post-stimulus task was (a) to provide a clear purpose for reading the stimuli, (b) to ensure that participants paid attention to the task, and (3) to score the participants for accuracy (Keating & Jegerski, 2015; Marsden et al., 2018). Responses to CQs were used as an exclusion criterion to the data. If the CQ response was incorrect, this trial was excluded (Keating & Jegerski, 2015). Additionally, CQ accuracy scores of 80% or over were needed for participants to be included in the data (Marsden et al., 2018). If participants scored below this threshold, they were excluded. However, all participants scored above 80% and higher, therefore, none were removed.

The wording of the CQs did not repeat the target form to avoid drawing attention to the violation present in the stimulus (Keating & Jegerski, 2015). The CQs were posed as statements, where participants were required to decide whether the statement was correct or incorrect in nature about the preceding target sentence. Participants responded by selecting a *tick* or a *cross* symbol. Statements were given in place of questions to avoid any confusion with answering *yes* or *no* in Welsh, because in Welsh, there are multiple ways to answer *yes* or *no*, depending on person, tense, and emphasis[17]. The CQs were identical for both grammatical levels of each item and were balanced for *correct / incorrect* across gender and animacy within experimental and non-experimental distractor conditions. The full list of CQs can also be found in appendix H.

For the presentation lists, a by-2 *Latin Square Design* was implemented (Keating & Jegerski, 2015; Slabakova et al., 2020). As the current SPR task contained experimental doublets (one grammatical & one ungrammatical), two

[17] For example, to answer questions in the 1st person present tense with *yes*, it is 'Ydw' and to answer *no*, it is 'Nac ydw'. However, to answer questions posed in the 3rd person future tense, *yes* is 'Bydd' and *no* is 'Na fydd', and to answer questions in the past tense, *yes* is 'Do' and *no* is 'Naddo' irrespective of person. This is not an exhaustive list. Please see https://www.bbc.co.uk/wales/learnwelsh/pdf/welshgrammar_ff_replies.pdf for a detailed explanation.

presentation lists (List I & List II) were created so participants were exposed to one version of each item. The following list represents the division of the items testing experimental doublets ('a' versions are grammatical & 'b' versions are ungrammatical).

List I: 1a, 2b, 3a, 4b, 5a, etc
List II: 1b, 2a, 3b, 4a, 5b, etc.

Two presentation lists each included 128 sentences, balanced across conditions and grammaticality. The sentences were randomly fixed for all participants using Random.org, with no two items from the same experimental and non-experimental conditions side by side in the presentation lists. Due to the high number of items in the current SPR task, the 128 items were divided into two tasks, containing 64 items in each part. This was done to avoid fatigue and boredom which could in turn affect reading times (RTs). Following the by-2 *Latin Square Design,* each participant saw either *List I* or *List II,* split across two tasks.

The SPR tasks were created in Gorilla.sc. The lists were counterbalanced across participants, using a branch node in the experimental design to ensure that for every two participants who saw *List I,* the next two participants saw *List II*. The tasks were not performed back-to-back but with two memory measures in between (these are outlined in the following sections). The lists across the two tasks were balanced for grammaticality, gender, animacy and the number of correct/incorrect CQs following the experimental and non-experimental conditions. Each task took roughly 8-10 minutes to complete (up to 20 minutes in total), with 4 practice trials at the start of task 1.

Reaction times in milliseconds between each button press were recorded to indicate how long the participant spent reading that particular word, reflecting the time taken to process the gender agreement and its violation in the different versions (Marsden et al., 2018; Slabakova et al., 2020). Data were trimmed in the following way:

1. If responses to CQs were incorrect, the item was excluded (Keating & Jegerski, 2015), and if participants scored less than 80% accuracy overall on CQs, they were excluded (Marsden et al., 2018).

2. If the participants assigned the incorrect gender to the noun in the gender decision task, this trial was removed from the data set (Grüter et al., 2012; Hopp, 2013, 2016a; Lemhöfer, Schriefers, & Indefrey, 2014).
3. If RTs were less than 200ms, this trial was removed from the data set.
4. If RTs were > 3 SDs from the mean (following the three previous exclusion criteria), this trial was removed from the data set.

The gender decision task methodology is outlined in section 6.1.8 and SPR data excluded as a result of the gender decision task accuracy is provided in section 6.2.1.1. The results from the SPR task will be used to address the primary research question posited in this experiment, which is 'To what extent do Welsh-English adult bilinguals show sensitivity to gender agreement violations when gender is independent of mutations, in conjunction with mutations and/or encoded through mutations, in local-contexts?'.

6.1.5 Operations Span task

The first memory task in the battery was the Operation Span (OSPAN) test, which is a measure of working memory, specifically a measure of the central executive (Turner & Engle, 1989). A role for working memory has been identified in language comprehension, specifically the processing of gender agreement errors in comprehension (Keating, 2009, 2010; Sagarra & Herschensohn, 2010), showing that participants with higher working memory are more sensitive to gender disagreement (Sagarra & Herschensohn, 2010).

During tasks which measure sensitivity to gender errors, (e.g., SPR tasks), subjects are required to identify, store and integrate sources of linguistic information whilst paying attention to upcoming parts of the sentence (Baddeley, 2003; Sagarra & Herschensohn, 2010). Tests of WMC that tap into these complex cognitive tasks have been found to be predictors of performance (Keating, 2010; Sagarra & Herschensohn, 2010). For instance, the Reading Span task (RST) (Daneman & Carpenter, 1980) combines the processing component and a storage component, requiring the subject to read sets of sentences and to remember an element from each sentence (which may be the final word in a sentence, a letter or a numeral). The participant is then asked to recall this final word, letter or number in their presentation order. As the secondary processing task, participants make judgements of the semantic acceptability or grammaticality of the sentence. This secondary task obstructs the rehearsal of the elements to be recalled in memory, which prevents the RST from becoming a measure of short-term memory alone. This task is typically administered in the subjects' L1. However, this can be problematic for bilinguals, as bilinguals may be dominant in one language over the other, and thus, this complicates the language in which it is administered. Additionally, as co-activation of the two languages happens, this can influence processing speed when having to selectively ignore one language over the other.

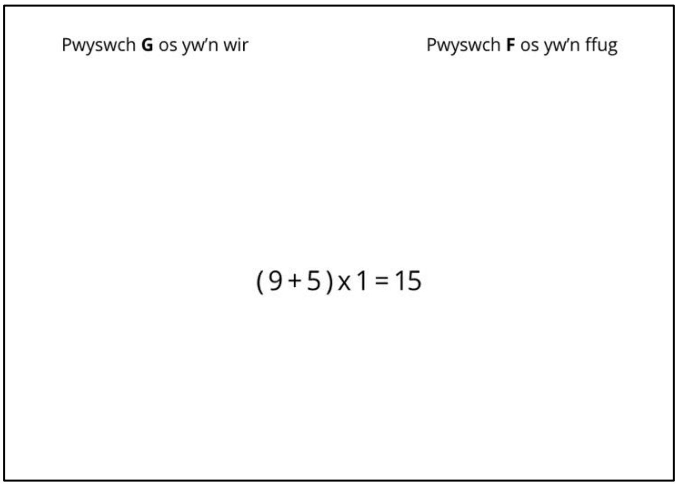
Unlike the RST, the Operation Span Task (OSPAN) (Turner & Engle, 1989) is language independent. The premise of the OSPAN is similar to the RST, however, the sentences are replaced with mathematical equations because the authors believed that subjects' reading abilities could be predicted without reading sentences (Turner & Engle, 1989). Participants are shown arithmetic operations and are required to solve these equations, followed by a word or a letter which they are required to remember. Similar to the RST, the OSPAN requires the subjects to recall the final letters (or words) in their presentation order. The OSPAN was selected to measure WM in the present experiment because it is language independent and is considered a reliable and valid measure of WMC (Conway et al., 2005).

The administered OSPAN was modelled after Turner and Engle (1989) and was cloned from an available version on Gorilla Open Materials (Brown et al., 2021). The cloned version was adapted to incorporate recommendations from more current research, to ensure a sound and robust design. These changes are outlined below.

Following previous methodologies and recommendations (Conway et al., 2005; Unsworth, Heitz, Schrock, & Engle, 2005), there were a total of 42 operations distributed in sets of two, three, four or five, with 3 sets for each set size. The order of set size varied randomly so that the participants could not predict the number of upcoming items. The participants were required to solve the mathematical operations by determining whether the equation answer was true or false, by clicking the keyboard letter 'G' for *True* ('G' for 'Gwir' in Welsh) or 'F' for *False* ('F' for 'Ffug' in Welsh) and were given 5000ms to do this. An example of this is shown in the figure below.

Figure 6.2

OSPAN Example



Following each mathematical operation, participants were presented with one of 12 phonologically distinct letters (consonants: B C D F G J L N P R S T) that exist in both Welsh and English. These letters were included to avoid reliance on one language over the other. Letters were chosen over words because previous research has shown that some of the shared variance between span tasks that use words and a measure of higher order cognition (e.g., reading comprehension) is due to word knowledge, word frequency, language proficiency and long-term memory strategies (Granena, 2013; Unsworth et al., 2005; Zalbidea, 2017). The letters remained on screen for 800ms (Granena, 2013; Unsworth et al., 2005) and the participants had to type the letters in their presentation order. This differed from the original version available on Gorilla, in which words were changed to letters.

The participants received the task title and instructions bilingually. There were five practice trials (one set of 2 operations and one set of 3 operations). Between each trial, there was a fixation cross for 1000ms to break up screen displays, with a fixation cross of 500ms between operation display and letter display. Participants recalled the letters at their own pace. Items were presented in the same order to all participants and took approximately 10 minutes to complete.

The scoring of WMC tasks, such as the OSPAN, is complicated (Conway et al., 2005; Juffs & Harrington, 2011). Scores are generally reported through absolute scoring or partial credit unit (PCU) scoring, with absolute scoring traditionally used for partially administered OSPAN measures (i.e., participants are stopped from continuing the tasks whenever they cannot recall a specific span range) and PCU scoring traditionally used for fully administered OSPAN measures (Najjari & Mohammadi, 2017). It has been argued that absolute span scores are generally inappropriate for individual differences research because it can greatly limit the sensitivity of the measure (Conway et al., 2005). In contrast, PCU scoring has shown an advantage in terms of the internal consistency of the task (Granena, 2013; Kane et al., 2004). Therefore, the PCU method with equal weighting was calculated (Conway et al., 2005; Granena, 2013; Najjari & Mohammadi, 2017).

PCU scoring calculates the average proportion of the to-be-remembered items (e.g., recalled letters in the task). Each set receives one mark. For example, the number of isolated letters in the task differed within each set so that there were three sets of two letters, three sets of three letters, three sets of four letters and three sets of five letters. Therefore, in a set of two, three, four, and five isolated letters, the correct recalling of each letter would receive 0.50, 0.33, 0.25 and 0.20 of the score 1, respectively. If a participant recalled all three letters in a set of three, they would receive a score of 1, while for the correct recalling of two letters, they would receive 0.67 of the score 1. These proportions were then averaged, i.e., the scores of all sets were totalled and divided by the number 12 (the total number of sets). This score was considered as the participant's WMC score (range possible from 0 to 12).

To understand the role of working memory capacity in the processing of grammatical gender in Welsh and to address the research question, 'To what extent do cognitive and environmental individual differences play a role in Welsh-English adult bilinguals' processing of gender in Welsh?', each individual's WMC scores as constructed by the OSPAN, was included in the generalised linear mixed effects model as a fixed factor to establish how it affects the outcome variable (reading times). The results are presented in sections **Error! Reference source not found.** and **Error! Reference source not found.** of this chapter.

6.1.6 Continuous Visual Memory Task

The second memory task in the battery was the Continuous Visual Memory test (CVMT: Trahan & Larrabee, 1988), which is considered a measure of declarative memory (Ullman, 2001, 2004). Declarative memory has been found to play a role in

L2 sentence processing and to an extent, the processing of grammatical gender (Morgan-Short et al., 2010). Research has shown that as L2 experience and proficiency increases, L2 learners become to rely more on the procedural memory system to process aspects of grammar, in particular rule-governed structure building (Stefaniak et al., 2021; Ullman, 2001). However, the mechanisms underlying the learning, representation and processing of grammar in the L1 are subserved by procedural memory. As the subjects in the current experiment are highly fluent Welsh bilinguals, it is not anticipated that declarative memory will play a role in the processing of gender. Yet, this needs to be checked, to determine whether gender processing in Welsh bilinguals relies on lexical/semantic processing (declarative memory) in place of grammatical processing (procedural memory).

Several tasks have been used to assess declarative memory, including the Continuous Visual Memory Task (Trahan & Larrabee, 1988), LLAMA-B (Meara, 2005) and the visual-auditory learning subtest of the Woodcock-Johnson III Tests of Cognitive Ability (Woodcock, Mather, & McGrew, 2001). Following previous research, the Continuous Visual Memory Task was selected to measure declarative memory (Carpenter, 2008; Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014; Ruiz, Chen, Rebuschat, & Meurers, 2019).

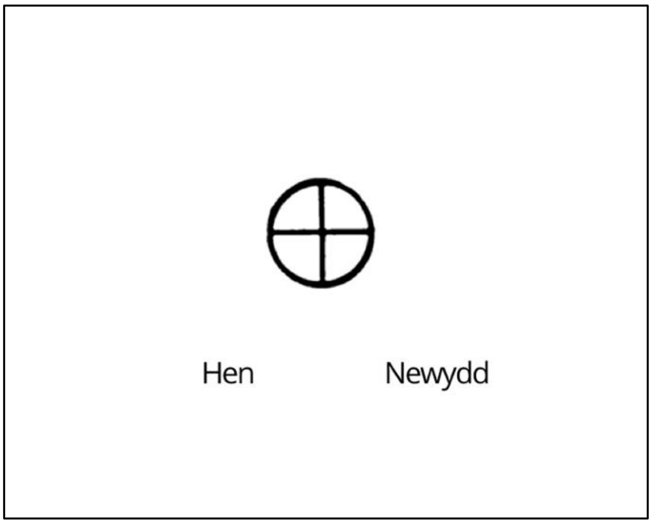
The CVMT served as a measure of nonverbal declarative memory and is a visual recognition task (Trahan & Larrabee, 1988). The test includes a recognition format to assess visual memory by using series of multiple recurring designs, many of which are perceptually similar to one another to help minimize the effects of verbal-spatial, motor-ability and verbal labelling of the test stimuli (Larrabee, 2009; Ullman et al., 1997). The CVMT in the current experiment is the same version as administered by Morgan-Short et al (2014) but programmed into Gorilla by Ryo Maie at Michigan State University. Permission was given by Kara Morgan-Short to clone the task, as it is not available on open materials. The CVMT methodology is outlined below.

The participants viewed a collection of complex abstract designs on the screen and were required to indicate whether the image they just saw was occurring for the first time (new) or an image that they had seen before (old). The complex abstract designs included polygons, patterns that appeared to be a dropped pile of sticks, and others that appeared to be a complex TV antenna (Larrabee, 2009). Seven of the designs were target items and were considered ‘old’ because they appeared seven times across the test (49 occurrences in total). There were an additional 63 distractor designs which only appeared once across the test. This totalled in 112 trials. The participants completed 11 practice trials prior to starting the task. All items were presented in a fixed order and each appeared on the screen for 2000ms (Morgan-Short et al., 2014).

To ensure the task was suitable for Welsh speakers, the instructions were translated from English to Welsh. The participants were instructed to respond to the images by selecting “Hen” (old) and “Newydd” (new) buttons on the screen, using the keyboard letter ‘S’ to select “Hen” and the letter ‘K’ for “Newydd”. The participants were shown the design for two seconds but were allowed to respond any time later (up to 10 seconds). After the participant responded, the presentation advanced to the next slide, which was a fixation cross displayed for 250ms in the centre of the screen between each trial. The CVMT took between 6 and 10 minutes to complete. Figure 6.3 shows an example of a trial in Gorilla. The participants received the task title and instructions bilingually.

Figure 6.3

CVMT Example



Following previous methodologies (Morgan-Short et al., 2014; Ruiz et al., 2019; Ullman et al., 1997), the total accuracy score was calculated to reflect the amount of learning. The total score was the participants’ accuracy responses, which was the number of designs correctly recognized as being new (“Newydd”) or old (“Hen”). This score was out of 112.

To understand the role of declarative memory in processing of grammatical gender in Welsh, each individual's raw score as constructed by the CVMT was modelled to establish how it affects the outcome variable. The results are presented in sections **Error! Reference source not found.** and **Error! Reference source not found.** of this chapter.

6.1.7 Tower of Hanoi task

The final memory task in the battery was the Tower of Hanoi (TOH), which is considered a measure of procedural memory (Takano et al., 2002; Winter, Broman, Rose, & Reber, 2001). With research showing that L1 speakers depend on the procedural memory system to process aspects of grammar, it is important to measure procedural memory to understand its role in the processing of grammatical gender in the population of interest. Measuring procedural memory may provide an indication on whether the highly fluent Welsh adult bilinguals rely on grammatical processing to process gender, by patterning similarly to L1 speakers of other gendered languages.

Various cognitive tasks have typically been used in L2 procedural memory research, including the dual-task Weather Prediction Task (DT WPT; Foerde, Knowlton, & Poldrack, 2006; Knowlton & Squire, 1995), the Serial Reaction Task (SRT; Nissen & Bullemer, 1987), the Alternating Serial Reaction Time Task (ASRT; (Howard & Howard, 1997) and the Tower of London (TOL) (Ettlinger, Bradlow, & Wong, 2014; Kaller, Unterrainer, & Stahl, 2012). The TOL is also frequently used to assess procedural memory learning ability in experimental and clinical research (Unterrainer et al., 2019). It is considered a reliable and valid measure of procedural memory, due to the use of implicit knowledge to solve the problems and subjects' demonstration of gradual improvement in accuracy and time (Buffington, Demos, & Morgan-Short, 2021). However, a variation of the TOL was administered in the current experiment, the Tower of Hanoi (TOH) (Simon, 1975). The TOL and TOH have typically been treated as interchangeable tasks because they both rely on a set of cognitive functions involved in planning and problem-solving which are controlled by prefrontal functions, namely, the frontal lobe of the brain (Bull, Espy, & Senn, 2004). Both the TOL and the TOH require subjects to transfer objects and have common rules that only one ball/disk can be moved at a time, and any ball/disk not being moved must remain on a peg (Bull et al., 2004; Welsh, Satterlee-Cartmell, & Stine, 1999). However, the TOH was selected to measure procedural memory over the TOL. This is because in the TOL, participants are instructed to achieve the configuration in a specified minimum number of moves as quickly as possible, whereas the participants are not given these instructions in the TOH. This difference promotes less online monitoring of the number of moves made in the TOH over the TOL (Bull et al., 2004). Therefore, it may be possible to see whether or not procedural learning ability was present across the task in the adult Welsh-English bilinguals.

The computerised version of the TOH available on Gorilla open materials was administered to the participants (Poort & Rodd, 2022). The participants were instructed to match the same configuration of coloured disks that rest on “towers” (pegs), from the first peg but on the final peg (Morgan-Short et al., 2014). Participants clicked and dragged disk-like shapes on to “towers”, from an initial configuration to a goal configuration (from the first to the final peg). In producing the goal configuration, participants were constrained by being able to move only the topmost disk on each “tower”, and when moved, the disk will fall to the lowest possible “tower” position. The participants were not able to place larger disks on top of smaller disks. For each trial, participants viewed a new initial end goal configuration and were told to match the goal configuration in the minimum number of moves as quickly as possible. The participants started with 3 disks and were given 600 seconds (10 minutes) to complete as many rounds as possible. This time restriction was put in place to avoid participants experiencing fatigue and boredom. The participants received the task title and instructions bilingually. Figure 6.4 shows a screenshot of the instructions and example, and Figure 6.5 shows the first set in the task – 3 disks.

Figure 6.4

TOH Instructions and Configuration Example

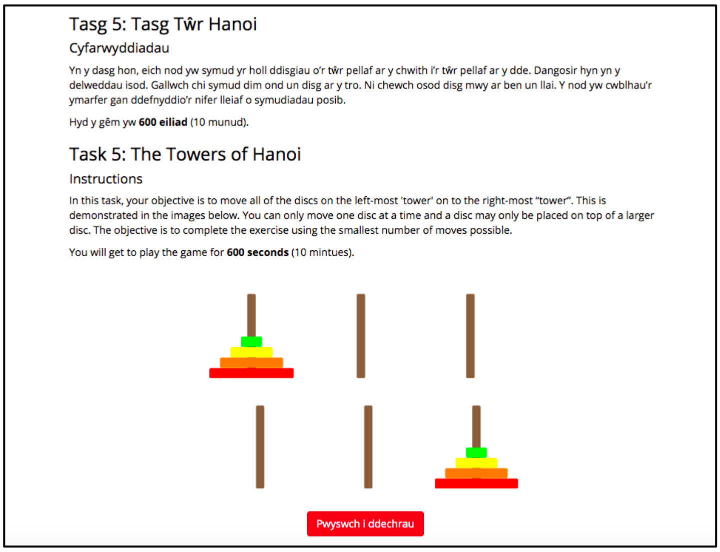
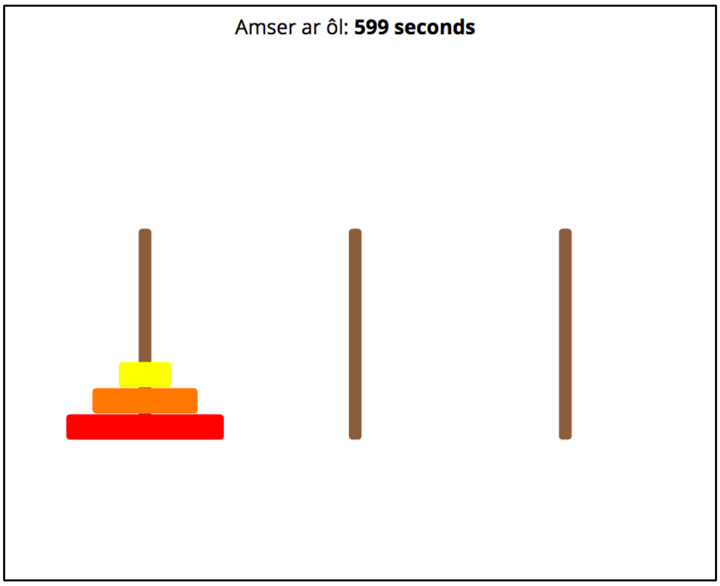


Figure 6.5

TOH First Set of 3 Disks



Several methods have been used to score solutions to tower problems (Bull et al., 2004). To measure improvement in studies using TOL, Unterrainer et al (2003) examined reaction time measures, namely the initial think time. This measures the time between the presentation of a trial and the first move made by the participant on that trial, regardless of whether the trial was solved in the specified number of moves. Unterrainer and colleagues (2003) also calculated the movement execution time, which is the time from first move to solution of the problem, while other studies have examined accuracy as the percentage of the correctly solved problems in the stated number of moves (Unterrainer et al., 2019). In Buffington and colleagues' study (2021), they measured the average percent change in initial think time, or planning time, with a higher percent change representing a greater decrease in initial think time and presumably more procedural learning (used in Morgan-Short et al., 2014). They also measured the average total time to match a goal configuration on the second administration of the task, normalized relative to the other participants, as a measure of overall improvement on the task, which may or may not be specific to procedural memory (used in Antoniou, Ettlinger, & Wong, 2016; Ettlinger et al., 2014).

The task administered to the participants was a cloned task by Poort and Rodd (2022) in Gorilla who used it as a distractor task in their study. Unfortunately, the data for *initial think time* was not collected and this was not identified prior to data collection, in time to change it to include this data. However, the following data were collected: the optimal number of moves, total number of moves made and the execution time (i.e., RT in milliseconds to complete the set of disks). Therefore, the following scoring system was calculated to measure improvement and procedural memory learning ability in the Welsh bilingual participants.

Step one took the number of moves taken by the participant for the successfully completed final set of disks (e.g., 66 moves to successfully move 5 disks). Step two took the optimal number of moves for completed final set of disks (e.g., 31 optimal moves to move 5 disks), and subtracted this optimal number from the number of moves taken by the participant (e.g., 66 – 31 = 35). In the final step, this calculated number (i.e., 35) was then divided by the number of disks in the successfully completed final set of disks (e.g., 35 ÷ 5 = 7). This final number / score (i.e., 7) is indicative of procedural memory learning ability in the participants. The lower the score, the more suggestive it is of an improvement in procedural memory learning ability throughout the task. A score of zero (or close to zero) shows that the participant moved the set of disks with the optimal number of moves, while a score further from zero (e.g., 10+), shows that the participant used an increased number of moves from the optimal number of moves, and that this did not improve throughout the task. This approach to scoring the TOH was innovative and experimental. It was done to capture the number of moves taken by the participant, with the number of moves directly reflecting a possible improvement in the task. This scoring system attempts to demonstrate procedural memory learning ability in the Welsh bilinguals. It is therefore considered as the Welsh bilingual participants' PM (procedural memory) score.

To understand the role of procedural memory in processing of grammatical gender in Welsh, each individual's score as constructed by the TOH was modelled to establish how it affects the outcome variable. The results are presented in sections **Error! Reference source not found.** and **Error! Reference source not found.** of this chapter.

6.1.8 Gender decision task

The final item in the battery of tasks was an offline gender decision task. Following previous methodologies (Grüter et al., 2012; Hopp, 2013, 2016a; Lemhöfer et al., 2014), this task measured participants' gender assignment of the target nouns, of the 64 target nouns used in the four SPR task critical conditions (16 nouns per condition, balanced for gender and animacy). This task was created because previous research has shown that variability in lexical gender representations moderates the syntactic processing of L2 gender agreement (Gruter et al., 2012; Hopp, 2013, 2016a).

Welsh noun gender assignment was assessed using the gendered numeral four (*pedwar* [m] / *pedair* [f]). This was done because participants produced the gendered forms of the numeral four most accurately in experiment one and the target nouns following the numeral are produced in their bare form, i.e., no involvement with mutations. A simple sentence structure was formed to assess gender assignment, with the target noun as the final word in the sentence, as seen in example 15 below. There were 64 trials, representing each of the 64 target nouns. There was no gender agreement present in the sentence (Grüter et al., 2012). In other words, the sentence structure followed Verb-Numeral-Noun structure and the numeral in the sentence was 'missing'. The participants were required in fill in the 'missing' numeral and in doing so, had to decide which form of the numeral four, masculine or feminine, agreed in gender with the target noun in the sentence, seen below in Figure 6.6. This way, the participants were required to assign gender to the target noun presented to them on screen. An example of the sentence-structure used to elicit gender assignment is presented below.

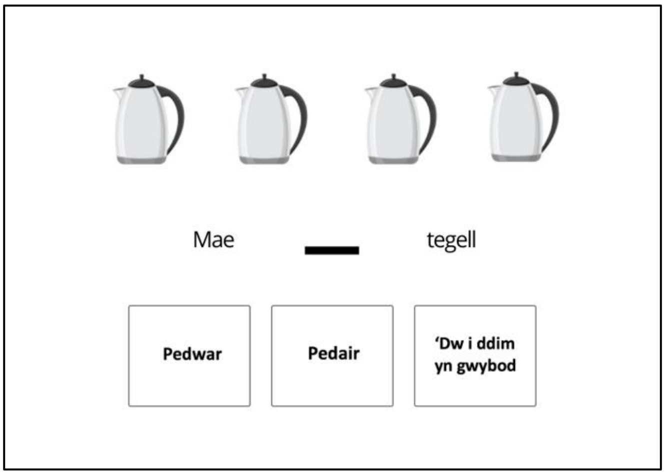
Mae	pedair	/	**pedwar**	tegell
Be-3SG	Numeral-F	/	Numeral-M	Noun-M
There are	four	/	four	kettle [M]

'There are four kettles'

The participants saw four singular images of the target item, e.g., four kettles, with the target sentence shown underneath. Below this, there were three button options – '*pedwar'* [m], '*pedair'* [f] and '*Dw I ddim yn gwybod'* (*I don't know*). The participants were required to select one of the three options, deciding which form, masculine or feminine, agreed in gender. An example of the trial display screen is shown in Figure 6.6.

Figure 6.6

Gender Decision Task Example



The participants received the task title and instructions bilingually. There were four practice trials. Between each trial, there was a fixation cross for 250ms to break up screen displays and to ensure that the participant is looking at the same place at the beginning of each trial. Items were presented in the same order to all participants. The task took 2-6 minutes to complete.

Participants' responses were coded for gender accuracy on the basis of the gendered numeral form chosen by the participant, with '*pedair'* assigning feminine gender to the noun and '*pedwar'* assigning masculine gender to the noun.

Participants received one point for correctly assigning the gender and zero for incorrectly assigning the noun's gender, with correct numeral selection reflecting correct gender assignment (Grüter et al., 2012). If the participant selected '*I don't know*', the response was coded as incorrect and received a score of zero.

The results of this task were used to exclude incorrectly assigned words from the self-paced reading task data, so that if the participants were inaccurate in the gender assignment in this task, then those trials were not included (Hopp, 2013, 2016a). Therefore, only trials that were correctly assigned gender in this task were included in the SPR data for analysis. This was done to allow for an accurate analysis of the data, in ensuring that the data analysed were based off participants knowing the gender of the noun(s). The number of trials removed is discussed in section 6.2.1.1.

6.1.9 Pilot

The battery of tasks was pilot tested with the same two Welsh speakers who pilot tested experiment 1 – two females, aged 24 and 62. Pilot testing was carried out to ensure that task instructions were clear, the tasks elicited the target responses, and to establish the length of time to complete the battery of tasks. This was done to better informed and to be transparent with the future participants. The battery was completed with understanding and ease, without any difficulties navigating the tasks online (via Gorilla.sc). Upon inspection, the tasks elicited the target responses. Both participants completed the battery within 45 minutes, and this was the guideline given to the participants. The following sections present the results from this follow up experiment.

6.1.10 Data Analysis

This section outlines the data cleaning process for the SPR, how the linear mixed effects models with the SPR data were built and how these were built with/without the additional variables under investigation.

6.1.10.1 Exclusion criteria

Only the data for the experimental conditions were analysed and not the non-experimental conditions. Therefore, following previous research, various exclusion criteria were applied to the SPR experimental conditions data. First, reading times were removed if the response to the CQ following the sentence was incorrect and this was applied first as it is standard practice in sentence processing research, particularly in L1 speaker populations (Keating & Jegerski, 2015; Marsden et al., 2018). Six participants answered all CQs correctly, while 15 participants answered between 1 and 8 questions incorrectly (out of a possible 96). In previous studies, participants who have achieved less than 80% overall on CQ accuracy have been removed from data sets (Marsden et al., 2018), however, the lowest score was 87.5%, therefore, no participants were removed. However, individual items were removed due to incorrect CQ responses, and this removed the data points for all 5 segments for that particular item. This information is provided below.

Table 6.1

Exclusion Criterion 1: Data Points Removed

Trials removed	Total data points	Percentage of data
44 trials	44 * 5 = 220	3.27%[18]

For the second exclusion criteria, data points were removed if the participant incorrectly assigned gender to the target noun in the gender decision task (Grüter et al., 2012; Hopp, 2013, 2016a; Lemhöfer et al., 2014). Therefore, if the gender assignment was inaccurate, the trial including that target noun in the SPR task was removed (data points for all 5 segments per item). The information is provided below.

Table 6.2

Exclusion Criterion 2: Data Points Removed

Trials removed	Total data points	Percentage of data
273 trials	273 * 5 = 1,365	20.3%

[18] 6,720 data points in total, from 21 participants reading 64 experimental items (21x64=1,344) and 5 segments per item (1,344x5=6,720).

The data were also trimmed for outliers (Marsden et al., 2018; Slabakova et al., 2020). Reading times below 200ms were removed (Marsden et al., 2018). This led to the exclusion of 53 data points which was 0.79% of the data. For the upper cut-off of reading times, there appears to be no general consensus, particularly for bilinguals who are highly proficient in both languages. In their systematic review, Marsden et al (2018) found upper cut-offs ranging from 2000ms to 20,000ms, usually depending on the different proficiency levels between the studies. Recently, Slabakova et al (2020) removed reading times greater than 5,000ms from their L2 speaker data set. However, Marsden et al (2018) reported that 20 out of the 48 studies used standard deviations (SDs) above the mean to identify outliers. For instance, Hopp (2016b) applied this method, where reading times >3 SDs from the mean were trimmed. Therefore, the upper cut-off for the current data were trimmed for > 3 SDs from the mean. This trimmed a further 137 data points which was 2.04% of the data.

Table 6.3 presents the mean, standard deviation, and range of the data (in milliseconds), following the first two exclusion criteria (i.e., CQs and gender accuracy) and removing the reading times below 200ms. The mean, standard deviation and range were calculated for each segment. The final column shows the number of data points removed from each segment for the upper cut-off range.

Table 6.3

SPR Data Cleaning / Trimming

Segment	**Mean**	**SD**	**Range**	**> 3 SDs**	**Trimmed > 3 SDs**
1 Pre-critical	526.12	412.97	200.5 – 4931.9	1651.88	22
2 Critical	583.67	489.21	201.7 – 4647.9	1956.84	30
3 Spill over	504.08	293.87	200.0 – 4626.0	1175.48	35
4 Spill over +1	520.18	374.72	200.0 – 4486.0	1498.88	25
5 Spill over +2	487.59	277.88	201.6 – 3352.6	1111.52	25
				TOTAL	137

In total, 1,775 (26.41%) data points were removed, leaving 4,945 (73.59%) data points in the data set for analysis. The section below details how the data were inputted into the model(s).

6.1.10.2 *SPR data and Models*

Linear mixed-effects models measure how well an outcome variable can be predicted by fixed and random effects. Fixed effects model how the independent variable(s) affect(s) the outcome variable, while random effects model variance that can be attributed to other factors inherent in the sampling of the study, such as participant or item variance (Cunnings, 2012). Therefore, The SPR are analysed with linear mixed effects modelling (LMMs) (Baayen, Davidson, & Bates, 2008), to investigate whether there are significant differences in the participants' reading times between the two levels of grammaticality in any of the four experimental conditions.

Before fitting mixed effects models, it is necessary to check whether the assumptions of LMMs have been met (Meteyard & Davies, 2020; Winter, 2020). One assumption is that the data should be normally distributed (Winter, 2013). An exploration into the data, plotting a histogram and a Q-Q plot of the raw reading times, revealed that the reading times were positively skewed. This can be seen in the figures below. The two plots show that the assumptions of normality of LMMs were not met.

Figure 6.7

Histogram: Raw RT Distribution

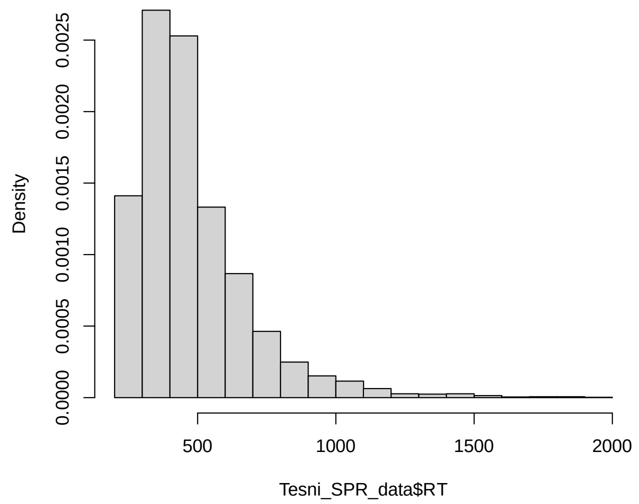
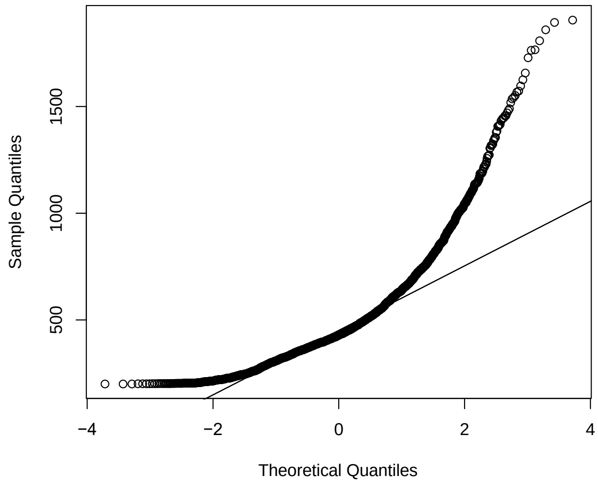


Figure 6.8

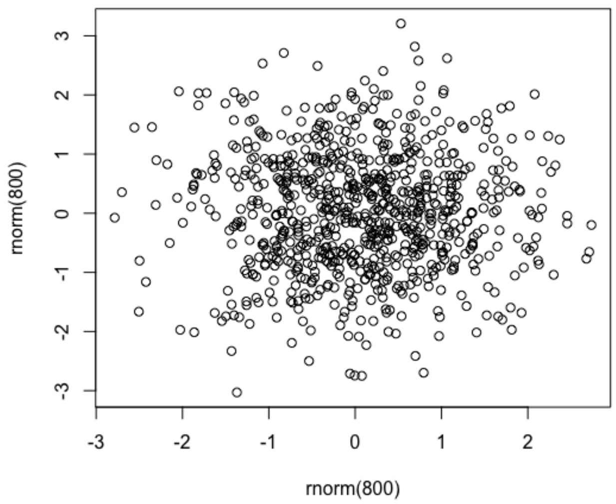
Q-Q Plot: Raw RT Distribution



Another assumption of LMMs is the absence of heteroscedasticity. This means that the residuals should be approximately equal across the range of the predicted values (Winter, 2013). For this to be met, the residuals of the model need to roughly have a similar amount of deviation from the predicted values. A residual plot with 800 fictious and random data points shows how a "good" residual plot looks like, where the data points are homoscedastic with no obvious pattern in the plot. Figure 6.9 shows this fictious and random data.

Figure 6.9

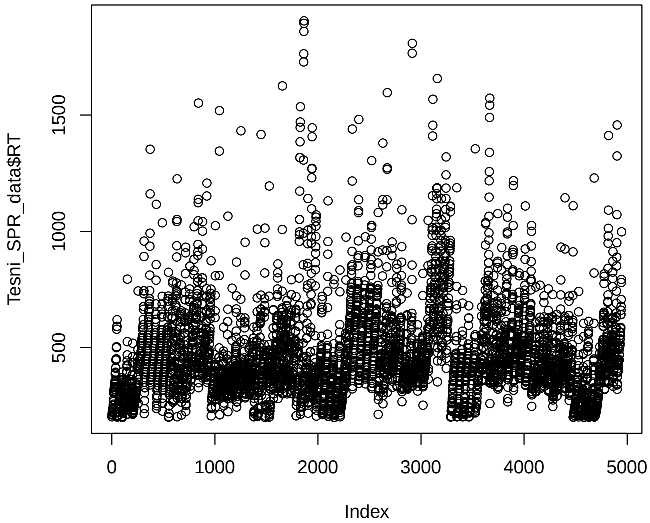
Residuals Plot: Fictitious Data Not Violating the Assumption of Absence of Heteroscedasticity



However, when the residuals are plotted for the raw reading times, Figure 6.9 shows that homoscedasticity was violated. This means that there was a problem with unequal variances. This is shown in the figure below.

Figure 6.10

Residuals Plot: Raw RT Distribution



The figures above show that the assumptions of normality and the absence of heteroscedasticity were not met. Because reaction time data is not normally distributed in nature, to account for these violations, researchers have log transformed the outcome variable (i.e., reaction time data), which can correct the problems of lack of normality and absence of homoscedasticity (Lo & Andrews, 2015). However, research has found that when data are log transformed, the differences disappear, because when applying a non-linear transformation (e.g., log-transformation), it can normalize the residuals and the differences between two (or more) samples can be obscured (Lo & Andrews, 2015). To solve this issue, generalized linear mixed effects models (GLMMs) offer an approach to allow statistical assessment on the original raw reaction time metric, but also to meet the mathematical constraints imposed by the statistical model (i.e., the mathematical criterion of normalized, homoscedastic residuals in linear regression) (Lo & Andrews, 2015). Therefore, following previous research (e.g., Fernandez et al., 2021; Lo & Andrews, 2015) GLMM(s) are utilized in place of LMMs (i.e., using LLMs and log-transforming the data) to analyse the reaction time data in this experiment.

When computing GLMMs, an appropriate distribution must be specified (Lo & Andrews, 2015). Although there is no consensus on the correct distribution, two of the two-parameter distributions currently implemented for GLMMs in the statistics package (LMER) as part of the *R* program for statistical computing (R Core Team, 2013), are the Gamma and Inverse Gaussian distributions. After having sought advice from Leigh Fernandez, a statistics expert, for performing the GLMM(s) in this experiment, Gamma distribution was selected as the appropriate distribution. This is because the Gamma distribution can reproduce surface characteristics of raw RTs - a unimodal skewed distribution with continuous responses greater than or equal to 0 (Lo & Andrews, 2015). For Gamma distribution, it is necessary to use a *link function*, which is a mathematical function characterizing the relationship between the predictors (fixed factors) and the outcome variable (i.e., RTs) (Lo & Andrews, 2015). The function binding the expected values produced by the predictors to the outcome variable is the *identity link* (which specifies a linear relationship between predictors and observed responses), which is appropriate to use because the data are situated well away from zero (in part because RTs less than 200ms were removed) (Lo & Andrews, 2015). This has been specified in previous studies as the Gamma distribution takes into account the positive skew of reading measures (e.g., Fernandez et al., 2021).

Following recommendations (Brauer & Curtin, 2018; Brown, 2021) and personal communication with Dr Leigh Fernandez and Dr Pablo Bernabeu (who extensively researched the different optimizers available to use with mixed effects models), control parameters were added to help the model converge (i.e., to find a good fit for the data within a reasonable number of iterations of attempting to estimate model parameters). These parameters included the argument *control = glmerControl* and the optimizer = *nloptwrap*. The number of iterations were also specified using the argument *maxeval = 1e8* and the argument *calc.derivs = FALSE* was specified to remove some of the derivative calculations that occur after the model has reached a solution.

The generalised linear mixed effects model reported in the results sections are the best-fitting models, by which models are fitted and compared using maximum likelihood ratio comparisons to find the one that best explains the data (e.g., Bates et al., 2015a; Bates et al., 2015b). The fitted models were evaluated via forward model comparison, using the anova() command for model comparison, to perform the likelihood ratio test criterion (Bates et al., 2015; Cunnings, 2012). First, the random effects structure was maximally specified (Barr, Levy, Scheepers, & Tily, 2013), but one that is theoretically motivated and supported by the data (e.g., Bates et al., 2015a; Bates et al., 2015b).

For the SPR data, three best-fitting models are found and reported. These three different models are for the different experimental conditions. Model one includes the SPR data for experimental conditions 1 and 2, as these both include the violation of the gendered numeral form (gender in conjunction with mutations via the gendered numeral 2 and gender independent of mutations via the gendered numeral 4, respectively). Model two includes the SPR data for experimental condition 3 as the violation is seen on the noun following the adjective (mutations independent of gender via adjectives). Model three includes the SPR data for experimental condition 3 as the violation is seen on the noun following the definite article 'y' (gender encoded through mutations via the determiner). Deviation coding was set for grammaticality, gender and condition (only in model 1) (Schad et al., 2020).

Regarding the random effects structure, model one had a by-participant random intercept as well as by-participant random slopes for grammaticality (grammatical, ungrammatical), condition (experimental conditions 1 and 2) and gender (masculine, feminine) and the interaction between the three, as well as associated correlation parameters. Models two and three had a by-participant random intercept as well as by-participant random slopes for grammaticality (grammatical, ungrammatical) and gender (masculine, feminine) and the interaction between grammaticality and gender, as well as associated correlation parameters. There was also a by-item random intercept. This did not lead to convergence errors (Barr et al., 2013). The random effects structure did not include a variance component for subject-related and item-related intercepts, for every within-subject and within-item fixed effect, and all of the possible correlations between these random effects (Barr et al., 2013), because of convergence issues.

6.1.10.3 Variables and Models

A secondary interest to this experiment is the roles of cognitive and environmental individual difference variables. Therefore, the effects and interactions will be checked for six continuous predictors once it has been established whether there are any main effects and interactions with the main linguistic manipulations (grammaticality / gender / condition). It will be possible to check whether there are additional main effects and interactions between the linguistic conditions and the six variables.

These six variables were five scores from the OSPAN (working memory), TOH (procedural memory), CVMT (declarative memory), BLP (dominance), Welsh cloze test (Welsh linguistic proficiency) and the final control predictor was the number of years a participant had spent using Welsh in an educational and/or professional environment. The data for the environmental individual difference variables (i.e., language dominance, Welsh linguistic proficiency, and the number of years a participant had spent using Welsh) were collected as part of experiment one.

The best-fitting models for the SPR data for the different experimental conditions will be taken as the base models, then, initially, the six variables will be included in the fixed effects structure only as main effects, as it may not be essential to include control predictors in random effects (Barr et al., 2013; Fernandez et al., 2021; Pereira et al., 2021). If there are significant effects for any of the variables in any of the models, additional models will be computed to see check the effects and interactions between the variables and the linguistic manipulations for the different experimental conditions.

All models were run in the R programme for statistical computing (R Core Team, 2022), using version 4.2.2 (2022.10.31) and version 1.1-30 of the lme4 package (Bates et al., 2015).

6.2 Results: Overview

Section 6.2 reports the results of the tasks administered in this second experiment, described in the previous section (Section 6.1). The battery of tasks was completed by 21 Welsh-English adult bilinguals. The results are divided into the following sections. Section **Error! Reference source not found.** presents the results for the gender task and the SPR task. Descriptive results are presented, followed by inferential statistics, using (generalised) linear mixed effects modelling to analyse the data. Section **Error! Reference source not found.** presents the descriptive results for the six cognitive and environmental individual difference variables, followed by inferential statistics. These findings are discussed in section 6.3.

6.2.1 Gender task and SPR task results

The two research questions which guide this experiment are (1) 'To what extent do Welsh-English adult bilinguals show sensitivity to gender agreement violations when gender is independent of mutations, in conjunction with mutations and/or encoded through mutations, in local-contexts?' and (2) 'To what extent do cognitive and environmental individual differences play a role in Welsh-English adult bilinguals' processing of gender in Welsh?'. In order to address these questions, the descriptive results for the gender decision task are first presented, as the results are used to remove items from the SPR data. This is followed by the SPR task results. These are followed by the cognitive and environmental individual differences results.

6.2.1.1 Gender decision task

The gender decision task measured participants' gender assignment of the 64 target nouns used in the four SPR task critical conditions. Participants' responses were coded for gender accuracy on the basis of the gendered numeral form chosen by the participant, with '*pedair'* assigning feminine gender to the noun and '*pedwar'* assigning masculine gender to the noun. Participants received one point for correctly assigning the gender and zero for incorrectly assigning the noun's gender. If they selected '*I don't know*', they received a score of zero.

First, the overall group accuracy is presented, followed by accuracy according to gender and the four conditions, as well as a brief analysis of items which may have proven to be somewhat problematic for the participants. The overall group accuracy reported in the table below.

Table 6.4

Gender Decision Task Descriptives: Overall

Gender descriptives	Raw	Percentage
Mean	50.24 / 64	78.50
SD	8.25	12.89
Range	41 – 64	64.1 – 100

Table 6.4 shows that the participants scored an overall group average of 78.5%, with a standard deviation of 12.89%. The histogram below shows the distribution of scores in this task.

Figure 6.11

Gender Decision Task Descriptives Distribution

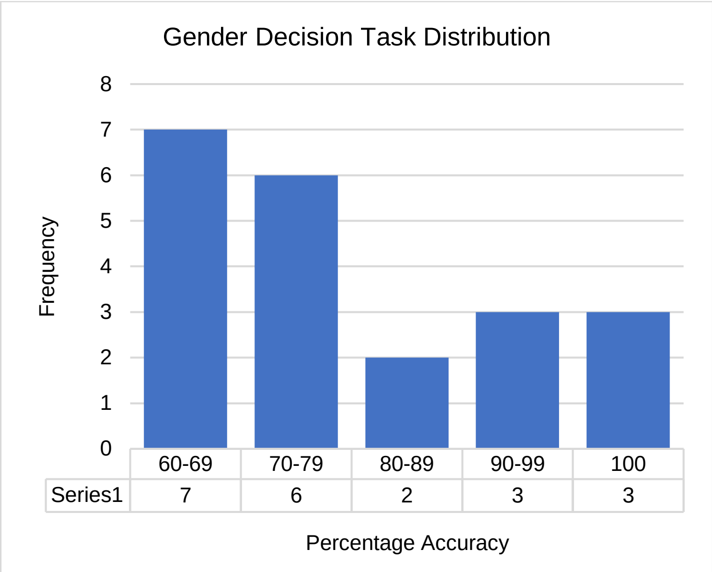


Figure ***6.11*** shows that seven participants scored between 60-69%, six scored between 70-79%, two scored between 80-89%, six scored above 90%, with three of those participants scoring 100% accuracy. The table and figure below show the accuracy according to masculine and feminine items.

Table 6.5

Gender Decision Task Descriptives: Gender

Gender descriptives	Masculine Raw	Masculine Percentage	Feminine Raw	Feminine Percentage
Mean	28/32	87.50	22/32	69.50
SD	3.78	11.82	5.86	18.30
Range	20 – 32	62.5 – 100	12 – 32	37.5 – 100

Figure 6.12

Gender Decision Task Descriptives: Gender

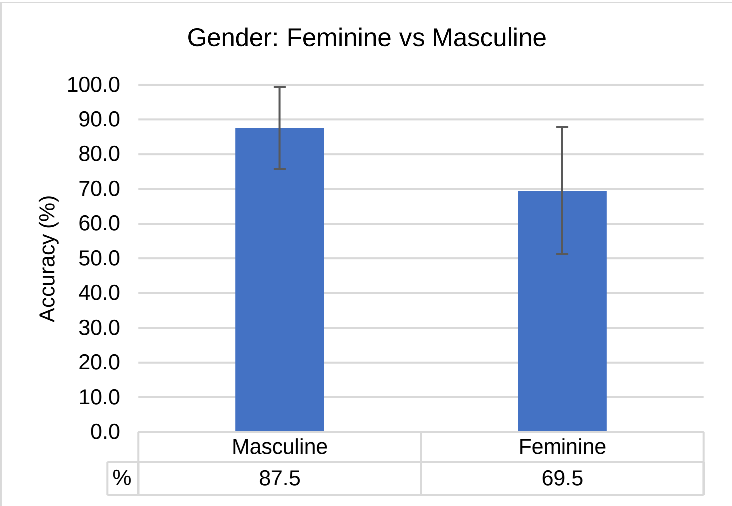


Table 6.5 and Figure 6.12 show that the participant's performed best on masculine nouns (*M*=87.5%, *SD*=11.8%) in comparison to feminine nouns (*M*=69.5%, *SD*=18.3%). The feminine nouns also have a larger standard deviation than the masculine nouns, which suggests a wider variation in accuracy scores. Table 6.5 shows that the range of accuracy scores for feminine nouns is from 12 (37.5%) to 32 (100%).

A paired samples student's t-test showed that this difference between the two genders was statistically significant (p = <0.001), and that the data were normally distributed (W = 0.966, p = 0.650). The difference in the scores may be explained by the fact that the majority of nouns in Welsh are masculine (approximately 69%) (Hammond, 2016). Therefore, the Welsh speakers may have higher gender assignment accuracy for masculine nouns as they are more common in the language. Additionally, it may be that the results reflect a default strategy, in that masculine forms are used as a default (Thomas, 2001). It is also possible that the higher accuracy for masculine forms in comparison to feminine forms stems from the fact that the masculine form of the numeral four is taught as part of the Welsh counting system.

When looking at the items individually, 17 out of 64 nouns were correctly assigned gender by all 21 participants, while another 41 nouns were assigned gender by over 50% of the participants. However, 6 nouns were only correctly assigned gender by less than 50% of the participants. These nouns were *gorsaf* [f] (English: 'station' as in *train station*, 10/21 participants), *cigfran* [f] (English: 'raven', 9/21 participants), *carreg* [f] (English: 'stone, 9/21 participants), *madfall* [f] (English: 'lizard' / 'newt', 8/21 participants), *cleren* [f] (English: 'fly', 8/21 participants) and *cawod* [f] (English: 'shower', 5/21 participants). All six nouns were feminine. Looking at the nouns on an item-by-item basis supports the data presented in Table 6.5 and Figure 6.12, showing that the participants were more accurate in assigning correct gender to masculine nouns than feminine nouns.

Interestingly, of these six nouns, *gorsaf* (station) and *cawod* (shower) are considered to be core vocabulary items that Welsh learners should learn according to the word lists created by Meara and Morris in 2019. Additionally, *carreg* (stone) is in the new frequency list for *Contemporary Welsh* and is included in the *Top 500 Noun Lemmas*, as identified, and created by *CorCenCC*. The three other nouns, however, are likely to be considered infrequent animal nouns in Welsh, each with a frequency of less than 1 word per million according to the KWIC tool available on *CorCenCC*.

As previously noted, this task was administered to control for the participants' individual knowledge (Hopp, 2013, 2016a). Any incorrect or missing ("I don't know") items from the task were removed from the SPR data. This was to ensure that any effects emerging from the analyses were due to participants' knowledge of the gender of the noun(s). This served as one of the exclusion criteria applied to the SPR data, which was discussed in the data analysis section. The following section reports the SPR task results.

6.2.1.2 Self-paced reading task

The participants completed a SPR task in Welsh, which was described in section 6.1.4. There were four experimental conditions and the data from these conditions are used to address the research questions posed in this second experiment. The results will examine the participants' RTs in two minimally different versions of gender agreement and gender disagreement across the four conditions. Reaction times were recoded for 5 segments: pre-critical, critical, immediate spill over, spill over +1 and spill over +2. Reading times at group level for the five segments and both levels of grammaticality across the four experimental conditions are presented in Table 6.6 are visually presented in Figure 6.13.

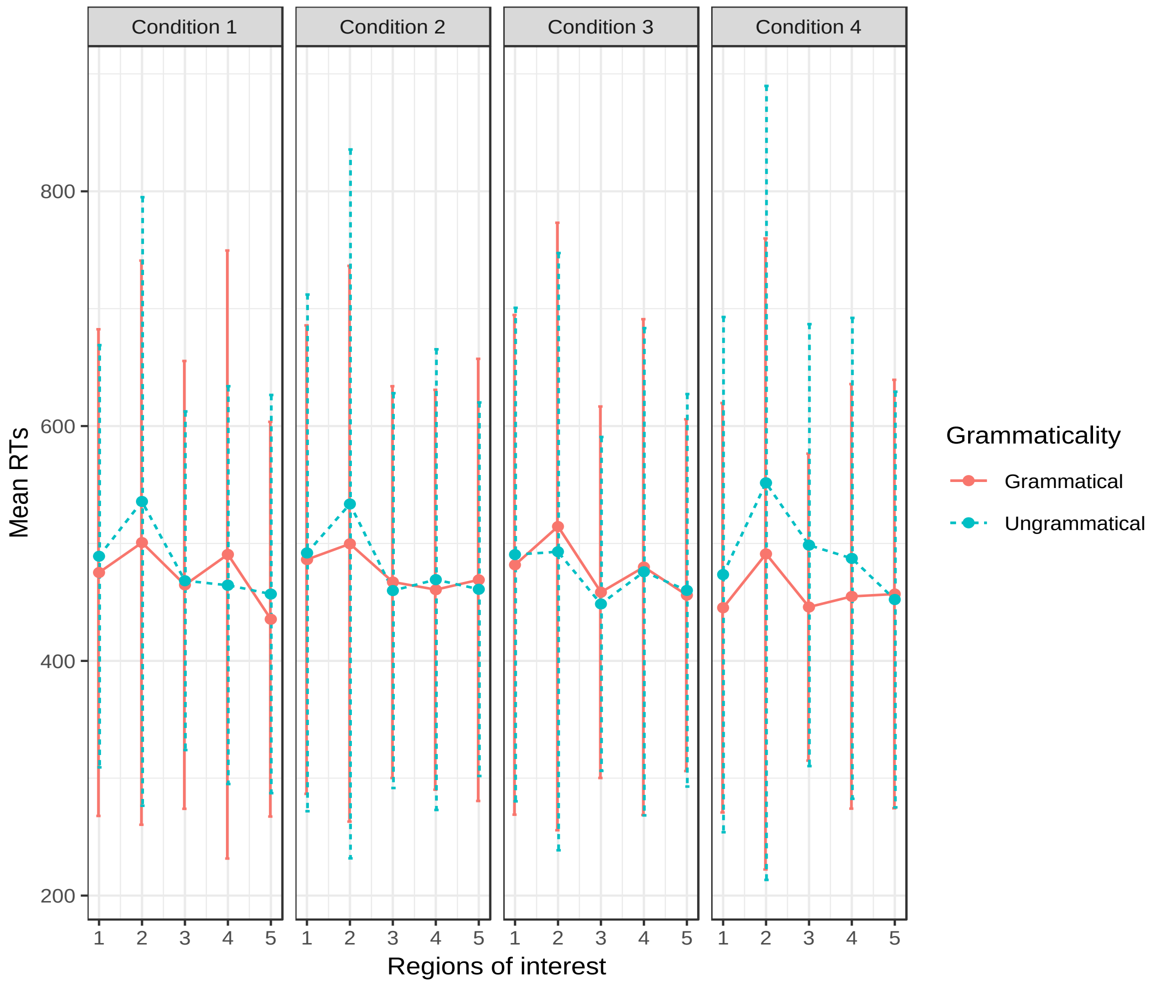
Table 6.6

Descriptive Statistics For Each Condition By Grammaticality and Segments

Descriptives		**Condition 1**		**Condition 2**		**Condition 3**		**Condition 4**	
		GR	UNG	GR	UNG	GR	UNG	GR	UNG
Pre-critical	*M*	475.17	489.11	486.24	491.94	481.78	490.52	445.27	473.46
	SD	207.36	179.81	199.51	220.11	212.85	210.22	174.38	219.42
Critical	*M*	500.75	535.73	499.72	533.63	514.46	492.95	491.04	551.58
	SD	240.29	259.18	236.74	301.90	258.68	254.38	268.82	338.18
Spill over	*M*	464.73	468.22	467.11	459.85	458.38	448.42	445.78	498.58
	SD	190.74	144.17	166.88	168.13	158.23	142.19	130.77	188.29
Spill over	*M*	490.58	464.43	460.56	469.15	479.85	475.83	454.90	487.28
+1	*SD*	258.98	169.43	170.26	196.14	211.20	207.51	180.85	204.75
Spill over	*M*	435.46	456.91	468.95	460.99	455.97	460.06	456.90	452.28
+2	*SD*	168.07	169.53	188.33	159.13	149.87	167.19	182.48	177.05

Figure 6.13

Descriptive Statistics for each Condition by Grammaticality and Segments



In condition 1, the participants' RTs for the grammatical items were faster in four out of the five segments. These were the pre- (13.94ms), critical (34.98ms), immediate spill over (3.49ms) and spill over+2 (21.45ms) segments. However, RTs were faster in the spill over+1 (26.15ms) segment for the ungrammatical items. The difference between the two levels of grammaticality in the critical segment of interest (34.98ms) suggests that on average, the participants took longer to read the target nouns following the incorrect form of the number two (i.e., *dau* [m] in place of *dwy* [f] and vice versa, presented in the pre-critical segment) than the correct form. This is difference is, however, small (34.98ms). The standard deviations show that there is a wide range in reading times across the segments, which are all almost half of the group mean RTs. When the RTs for the segments are collapsed to create a global score, the participants on average spend 2366.69ms reading the grammatical items, compared to 2414.40ms reading the ungrammatical items (difference of 47.71ms).

These results suggest a trend toward sensitivity to gender agreement violations when gender is conjunction with mutations via the numeral two. However, there is notable variability within the group, where sensitivity may be present in some individuals, but not others.

In condition 2, the participants' RTs for the grammatical items were faster in three out of the five segments. These were the pre- (5.70ms), critical (33.91ms) and spill over+1 (8.59ms) segments. However, the RTs were marginally faster in the immediate spill over (7.26ms) and spill over+2 (7.96) segments for the ungrammatical items. The difference between the two levels of grammaticality in the critical segment (33.91ms) suggests that on average, the participants took longer to read the target nouns following the incorrect form of the number four (i.e., *pedwar* [m] in place of *pedair* [f] and vice versa, presented in the pre-critical segment) than the correct form. This is difference is, however, small (33.91ms). The standard deviations show that there is a wide range in reading times across the segments, which are all almost half of the group mean RTs. When the RTs for the segments are collapsed to create a global score, the participants on average spend 2382.58ms reading the grammatical items, compared to 2415.56ms reading the ungrammatical items (difference of 32.98ms). These results suggest a trend toward sensitivity to gender agreement violations when gender is independent of mutation via the numeral four. As was found in condition one, condition two also shows a large amount of individual variation, where sensitivity to gender agreement errors when gender is independent of mutations is present in some speakers, but not others.

The picture is less clear in condition 3 than in conditions 1 and 2, as the participants' RTs for the grammatical items were faster in only two out of the five segments. The RTs for the grammatical items in the pre- (8.74ms) and spill over+2 (4.09ms) segments were marginally faster than the ungrammatical items. However, the RTs for the ungrammatical items in the critical (21.51ms), immediate spill over (9.96ms) and spill over+1 (4.02ms) segments were faster than the grammatical items. This is in contrast to the two previous conditions, which show faster RTs for the grammatical items than the ungrammatical items for the critical segment. The differences between two levels of grammaticality do not show evidence that the participants took longer to read the target nouns with the lack of mutation in place of the SM, following the pre-nominal adjectives. Similar to conditions 1 and 2, the standard deviations show that there is a wide range in reading times across the segments, which are all almost half of the group mean RTs. When the RTs for the segments are collapsed to create a global score, the participants on average spend 2390.44ms reading the grammatical items, compared to 2367.78ms reading the ungrammatical items, with a negative difference of 22.66ms. In contrast to the conditions 1 and 2, there is no evidence to suggest a trend toward sensitivity to the agreement violations between the pre-nominal adjectives and the soft mutation triggered on the target noun. There is, however, notable individual variation within the group results.

As was found in condition 1, the participants' RTs for the grammatical items were faster in four out of the five segments in condition 4. These were the pre- (28.19ms), critical (60.54ms), immediate spill over (52.8ms) and spill over+1 (32.38ms) segments. However, the participants RTs were faster in the spill over+2 (4.62ms) segment for the ungrammatical items. The difference between the two levels of grammaticality in the critical (60.54ms), immediate spill over (52.8ms) and spill over +1 (32.38ms) segments suggest that on average, the participants took longer to read the target noun and two of the following spill over segments, when the noun was incorrectly marked for gender via SM [f] or no mutation [m] following the definite article 'y' (the). The standard deviations show that there is a wide range in reading times across the segments, which are all almost half of the group mean RTs. When the RTs for the segments are collapsed to create a global score, the participants on average spend 2293.89ms reading the grammatical items, compared to 2463.18ms reading the ungrammatical items, with a difference of 169.29ms. This difference is similar to conditions 1 and 2, showing faster reading times for the grammatical than ungrammatical items. The difference in between these two levels of grammaticality in this fourth condition, is notably larger (169.29ms) than conditions 1 (47.71ms) and 2 (32.98ms). Condition 4 results suggest a trend toward sensitivity to gender agreement violations when gender is encoded through mutation via the determiner 'y'. There is evidence to suggest that when gender is incorrectly marked on the target noun itself following 'y', participants take longer to read the ungrammatical items than the grammatical items.

At group level, the descriptive results suggest that the Welsh speakers may show sensitivity to agreement violations when gender is conjunction with mutations via the numeral *two* (condition 1), when gender is independent of mutations via the numeral *four* (condition 2) and when gender is encoded through mutations via the determiner ‘y’ (the) (condition 4). However, there is no evidence to suggest that they show sensitivity to the agreement violations between the pre-nominal adjectives and the SM triggered on the target nouns. Moreover, the reading times for all four conditions show a large amount of individual variation (see Appendix I for individual z-scores).

To better understand the SPR data, linear mixed effects modelling can show whether there are significant differences in the participants’ reading times between the two levels of grammaticality in any of the four experimental conditions. To avoid losing sensitivity of the possible grammaticality effects across the different experimental conditions, *segment* is not included in any of the models as a fixed effect. Rather, grammaticality, gender, and condition (for model 1) are considered the main linguistic manipulations of interest to try and capture the possible effects in the data. The main effects and interactions are checked in each of the different models.

The first model includes the data from conditions 1 (gender in conjunction with mutations via the gendered numeral 2) and 2 (gender independent of mutations via the gendered numeral 4). To find the best-fitting model, three models were run. One model included the fixed effects of condition (1, 2), grammaticality (grammatical, ungrammatical) and gender (masculine, feminine). The second included the interaction between condition and grammaticality and a third included the interaction between condition, grammaticality, and gender. The models evaluated via forward model comparison revealed that the model fit improved with the interaction between all three ($\chi^2(3) = 9.72$, $p = 0.02$). Therefore, the most complex model justified by the data included the fixed effects of condition, grammaticality and gender, and the interaction between the three. The output from the best fitting model is below.

Table 6.7

Best-fitting model: SPR Conditions 1 and 2

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	526.25	14.58	36.11	**<0.001**
Grammaticality [grammatical]	2.06	9.96	0.21	0.836
Condition [1]	-15.86	8.21	-1.93	0.064
Gender [masculine]	6.33	10.08	0.68	0.530
Grammaticality [grammatical] x Condition [1]	3.33	15.07	0.22	0.825
Grammaticality [grammatical] x Gender [masculine]	0.80	12.36	0.07	0.948
Condition [1] x Gender [masculine]	55.66	19.86	2.80	**0.005**
(Grammaticality [grammatical] x Condition [1]) x Gender [masculine]	63.24	33.66	1.88	0.060

Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	26400.0	51.38	0.86	0.93
Grammaticality [grammatical]	1368.0	36.98	-	-
Condition [1]	732.1	27.10	-	-
Gender [masculine]	762.1	27.61	-	-
Grammaticality [grammatical] x Condition [1]	2069.0	45.49	-	-
Grammaticality [grammatical] x Gender [masculine]	676.4	26.01	-	-
Condition [1] x Gender [masculine]	3574.0	59.79	-	-
(Grammaticality [grammatical] x Condition [1]) x Gender [masculine]	12730.0	112.81	-	-

The model's intercept is significant (β = 526.25, SE = 14.58, z = 36.11, p = <0.001). There is no significant effect of grammaticality (β = 2.06, SE = 9.96, z = 0.21, p = 0.084), condition (β = -15.86, SE = 8.21, z = -1.93, p = 0.064) or gender (β = 6.33, SE = 10.08, z = 0.68, p = 0.530). The only effect approaching significance is condition (p = 0.054), with decreased RTs in condition 2 than in condition 1, yet this did not reach statistical significance. The model revealed a significant interaction between condition and gender (β = 55.66, SE = 19.86, z = 2.80, p = 0.005), but no other interactions reached significance (ps > 0.060). There was no significant interaction between grammaticality and condition (β = 3.33, SE = 15.07, z = 0.22, p = 0.83), nor was there a significant interaction between grammaticality and gender (β = 0.80, SE = 12.36, z = 0.07, p = 0.95).

To check the interactions between condition and gender, pairwise comparisons were run using *emmeans* (see Lenth et al., 2022) with the adjust argument specifying ‘Bonferroni’ correction. The results revealed that in condition 1, masculine items had increased reaction times than feminine items (β = 34.2, SE = 12.9, z = 2.65, p = 0.008), but there was no significant difference between the two genders in condition 2 (p = 0.16).

In summary, the best-fitting model showed that there is no evidence to suggest sensitivity to the gender agreement violations in condition 1, where gender operated in conjunction with mutations (via the numeral 2) or condition 2, where gender operated independently of mutations (via the numeral 4). There is also no evidence to suggest a significant difference in RTs between the two conditions, nor is there evidence to suggest a significant difference in RTs between the masculine and feminine items. However, pairwise comparisons revealed that in condition 1, masculine items had increased reaction times than feminine item, however, this effect was not found in condition 2.

The second model includes the data from condition 3 (mutations independent of gender via pre-nominal adjectives). To find the best-fitting model, two models were run. One model included the fixed effects of grammaticality (grammatical, ungrammatical) and gender (masculine, feminine). The second included the interaction between grammaticality and gender. The models evaluated via forward model comparison revealed that the model fit improved with the interaction between grammaticality and gender ($\chi^2(1) = 25.22$, $p < 0.001$). Therefore, the most complex model justified by the data included the fixed effects of condition, grammaticality and gender, and the interaction between the two. The output from the best fitting model is below.

Table 6.8

Best-fitting model: SPR Condition 3

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	523.87	64.98	8.06	**<0.001**
Grammaticality [grammatical]	24.64	37.95	0.65	0.516
Gender [masculine]	-0.93	40.72	-0.02	0.982
Grammaticality [grammatical] x Gender [masculine]	-22.31	24.91	-0.90	0.371
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	31660.0	177.95	9159.0	37.92
Grammaticality [grammatical]	15650.0	125.11	-	-
Gender [masculine]	12590.0	112.21	-	-
Grammaticality [grammatical] x Gender [masculine]	7206.0	84.89	-	-

The model's intercept is significant (β = 523.87, SE = 64.98, z = 8.06, p = <0.001). There is no significant effect of grammaticality (β = 24.64, SE = 37.95, z = 0.65, p = 0.52) or gender (β = -0.93, SE = 40.72, z = -0.02, p = 0.98). There was no significant interaction between grammaticality and gender (β = -22.31, SE = 24.91, z = -0.90, p = 0.37). Pairwise comparisons using *emmeans* (see Lenth et al., 2022) with the adjust argument specifying 'Bonferroni' correction, revealed that there was no statistically significant difference in RTs between grammatical and ungrammatical items for either masculine items (β = -2.33, SE = 15.4, z = -0.151, p = 0.88) or feminine items (β = 19.97, SE = 16.6, z = 1.20, p = 0.23).

In summary, the best-fitting model showed that there is no evidence to suggest sensitivity to the agreement violations in condition 3, where mutations operate independently of gender (via pre-nominal adjectives). There is also no evidence to suggest a significant difference in RTs between the masculine and feminine items. Additionally, there is no significant interaction between grammaticality and gender, and this was further supported by the pairwise comparison results.

The third model includes the data from condition 4 (gender encoded through mutation via the definite article 'y'). To find the best-fitting model, two models were run. One model included the fixed effects of grammaticality (grammatical, ungrammatical) and gender (masculine, feminine). The second included the interaction between grammaticality and gender. The models evaluated via forward model comparison revealed that the model fit did not improve with the interaction between grammaticality and gender (χ^2 (1) = o, *p* = 1.000). Therefore, the most complex model justified by the data included the fixed effects of condition, grammaticality, and gender, without interaction between the two. The output from the best fitting model is below.

Table 6.9

Best-fitting model: SPR Condition 4

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	526.73	17.71	29.73	**<0.001**
Grammaticality [grammatical]	-29.41	10.92	-2.69	**0.007**
Gender [masculine]	22.29	24.74	0.90	0.368
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	3757.0	61.29	1751.0	41.85
Grammaticality [grammatical]	2842.9	49.82	-	-
Gender [masculine]	2641.9	51.39	-	-
Grammaticality [grammatical] x Gender [masculine]	10660.0	103.25	-	-

The model's intercept is significant (*β* = 526.73, SE = 17.71, *z* = 29.73, *p* = <0.001). There is also a significant effect of grammaticality (*β* = -29.41, SE = 10.92, *z* = -2.69, *p* = 0.007), with decreased reaction times for grammatical items compared to ungrammatical items. However, there is no effect of gender (*β* = 22.29, SE = 24.74, *z* = -0.90, *p* = 0.368). Because grammaticality emerged as a significant effect, two additional models were run, one with the reaction times for the critical segment and another with the reaction times for the post-critical segments (i.e., immediate spill over, spill over +1, spill over +2). The results from the model including RTs for the critical segment revealed an effect of grammaticality (*β* = 77.81, SE = 27.78, *z* = -2.80, *p* = 0.005), with decreased reaction times for grammatical items compared to ungrammatical items. However, the results from the model including RTs for the post-critical segments did not reveal an effect of grammaticality (β = -18.78, SE = 11.35, z = -1.66, p = 0.098).

In summary, the best-fitting model showed that there is evidence to suggest sensitivity to the gender agreement violations in condition 4, where gender is encoded through mutations, in determiner-noun contexts (via the definite article ‘y’). There is evidence to suggest that this sensitivity arises in the critical segment in the sentence, which is where the reader would notice the violation on the target noun. However, there is no evidence to suggest a significant difference in RTs between the masculine and feminine items.

Taken together, there is no evidence to suggest sensitivity to the agreement violations in conditions 1 (gender in conjunction with mutations), 2 (gender independent of mutations) or 3 (mutations independent of gender). However, there is evidence to suggest sensitivity to the gender agreement violations in condition 4 (gender encoded through mutations) and this sensitivity arises in the critical segment of the sentence. These results are used to address the main research question in this experiment, which is (1) ‘To what extent do Welsh-English adult bilinguals show sensitivity to gender agreement violations when gender is independent of mutations, in conjunction with mutations and/or encoded through mutations, in local-contexts?’. The following section presents the results for the cognitive and environmental individual difference variables.

6.2.2 Individual difference variable results

A secondary interest in this experiment is the role of individual difference variables in the processing of Welsh gender. These include three variables related to *memory*, namely, working memory, procedural memory, and declarative memory, as well as three variables related to Welsh *use* and *experience*, namely, language dominance, Welsh linguistic proficiency and the numbers of years spent using Welsh in an educational / professional setting. The descriptive results for the six individual difference variables are reported in the following section.

6.2.2.1 Working Memory

The Operation Span task was administered as a measure of executive working memory. The group scores can be seen in Table 6.10.

Table 6.10

Operation Span Group Descriptive Results

OSPAN descriptives	Raw	Percentage
Mean	10.61	88.39
SD	1.65	9.83
Range	6.83 - 12	56.92 – 100

Table 6.10 shows a wide range in the scores, with some participants scoring close to 60% and others scoring 100%, with a group mean of 88%. In total, six participants accurately recalled all of the letters in the task. Generally, the participants were highly accurate in recalling the letters following the maths operations, however, not all participants demonstrated high-span scores as measured by the OSPAN task.

Previous studies have excluded participants based on their maths operation accuracy score. This has been done to ensure that participants were not trading off between solving the operations and remembering the letters, nor were they using any strategies for recalling the letters. Some studies have applied an 80% (or 85%) accuracy criterion (e.g., Conway et al., 2005; Unsworth et al., 2005; Zalbidea, 2017). However, Dokić et al (2018) explored the impact of the accuracy criterion on the processing component of the span task and recommended that researchers discard this accuracy criterion as a criterion for filtering the results for further statistical analyses. They suggested that discarding this criterion will directly improve the efficiency of administration of WM span tasks. If an accuracy criterion of 80% were applied to the current group of participants, six out of the 21 participants would be excluded. The maths accuracy results revealed that for these six participants, 40% of the errors were due to speed errors (i.e., not responding in enough time) while the other 60% were due to accuracy errors (i.e., responding to the maths operation incorrectly). However, due to the (already) relatively small number of participants in this second experiment and following the recommendation posited Dokić et al

(2018), the participants who scored less than 80% accuracy on the maths element of the OSPAN were not excluded from the data set. The six participants who scored less than 80% on the maths operations were a mix of ages. Four of the participants were female and were aged between 28 and 32, while the other two participants were older, and were aged 59 (female) and 63 (male). It is therefore not possible to say that the maths operations errors were related to age in any way. Therefore, the WM scores from the 21 participants were included as a continuous predictor in the models.

6.2.2.2 Procedural Memory

The Tower of Hanoi task was administered as a measure of procedural memory. The group scores can be seen below.

Table 6.11

Tower of Hanoi Group Descriptive Results

TOH descriptives	Score
Mean	5.68
SD	4.32
Range	0 – 16.25

Table 6.11 shows variability in the group scores, with scores ranging from zero to 16.25. The calculated score tells us that the lower the score, the more suggestive it is of an improvement in procedural memory learning ability throughout the task. A score of zero (or close to zero) shows that the participant moved the set of disks with the optimal number of moves, while a score further from zero (e.g., 10+), shows that the participant used an increased number of moves from the optimal number of moves. The group average indicates that the participants showed a trend towards overall improvement on the task. The procedural memory scores from the 21 participants were included as a continuous predictor in the models.

6.2.2.3 Declarative Memory

The Continuous Visual Memory test was administered as a measure of declarative memory. The group scores can be seen below.

Table 6.12

CVMT Group Descriptive Results

CVMT descriptives	Raw	Percentage
Mean	72.14	64.41
SD	5.39	4.81
Range	64 - 86	57.14 – 76.79

Table 6.12 shows that no participant scored 100% on this task, with scores ranging from 57% to 77% (M=64%). Unlike the other two memory measures, there is less variability in the group, however, the participants were less accurate in identifying the designs in the task. The group scored above chance in this task (i.e., 50%), however, the results do not indicate a high declarative learning ability score for these participants as measured by the CVMT. The declarative memory scores from the 21 participants were included as a continuous predictor in the model.

6.2.2.4 Language Dominance

The Bilingual Language Profile questionnaire was administered as a measure of language dominance (in experiment 1). The participants completed five sections, providing biographical information and information regarding their language history, use, proficiency, and attitudes (for both Welsh and English) (see section 4.1.4 for more details). The responses generated a global score, ranging from -218 to +218. A positive score indicated English dominance, while a negative score indicated Welsh dominance. A score close to 0 (+/- 20) suggested a 'balanced bilingual'. This score reflects the participant's dominant language. The group scores presented below are for the subset of 21 participants who completed experiment 2.

Table 6.13

BLP Descriptive Results

BLP descriptives	Group Score (n=21)	Welsh dominant (n=11)	English dominant (n=7)	Balanced (n=3)
Mean	-9.61	-58.15	64.68	-4.93
SD	65.50	32.68	39.98	21.33
Range	-118.06 – 146.88	-118.06 – -30.25	23.16 – 146.88	-18.80 – 19.63

Table 6.13 shows that 11/21 participants are Welsh dominant speakers, while seven are English dominant and three can be considered as ‘balanced’ bilinguals. This shows that more than half of the bilingual participants are dominant speakers of Welsh. The calculated global score takes into consideration a number of factors, including language history (e.g., at what age did the speakers acquire the two languages), language use (e.g., how much time do you use the two language with family), language proficiency (e.g., rating how well they speak the two languages) and language attitudes (e.g., how the speaker feels when using the two languages) (see appendix A for the BLP questionnaire). The BLP measures the relative strength of the two languages, in producing an inherently gradient (not categorical) raw score (Birdsong, 2016; Olson, 2023).

It is possible that language dominance plays a role in the processing of Welsh, as it has previously been found to be a significant predictor of an outcome variable (e.g., skills / RTs) (Bonvin et al., 2021), despite not emerging as significant in experiment 1 of this thesis. Each participant’s global dominance score is included as a continuous predictor in the models.

6.2.2.5 Welsh Linguistic Proficiency

A Welsh cloze test was administered as a measure of Welsh linguistic proficiency (in experiment 1). The participants read a 358-word passage in Welsh on BBC1’s Blue Planet and filled in 44 gaps. One point was given for each acceptable response and no points were given for unacceptable responses. Further details about the scoring system can be found in experiment 1 and the answer key can be found in appendix E. This cloze test score was used as an indication of the bilingual’s Welsh linguistic proficiency level. The group scores presented below are for the subset of 21 participants who completed experiment 2.

Table 6.14

Welsh Linguistic Proficiency Descriptive Results

Welsh proficiency descriptives	Raw	Percentage
Mean	42	95.45
SD	2.12	4.82
Range	37 - 44	84.1 – 100

Table 6.14 shows that the participants in this follow up experiment are highly proficient speakers of Welsh (*M*=95.5%, *SD*=4.8%). There is little variation in the group scores (84–100%), with the lowest score being 84% (37/44). In total, four participants scored 100%. These results indicate that the Welsh-English bilinguals, who participated in this follow up experiment, are highly proficient speakers of Welsh as measured by the Welsh cloze test. Each participant's proficiency score is included as a continuous predictor in the models.

6.2.2.6 Welsh in an Educational / Professional Setting

The final individual factor under consideration is the participant's use of Welsh in a formal setting. This information can be taken from the language history section of the BLP, where the participants answered the following question, '*How many years have you spent in a work or educational environment where Welsh is spoken?*', with years ranging from 0 to 20+. From this question, it is possible to determine the number of years spent in a Welsh speaking work or educational environment. The group results are shown below.

Table 6.15

Use of Welsh in a Formal Setting Descriptive Results

Use of Welsh descriptives	Years spent in Welsh speaking environment
Yes	n=14
Range	3 – 22 years
No	n=7

Table 6.15 shows that 14 out of the 21 participants use Welsh in a work or educational environment, with years ranging from 3 to 22, while 7 of the participants answered with zero years. For those who answered yes, the distribution of years can be seen in the histogram below[19].

Figure 6.14

Histogram: Years Spent Using Welsh in Work Distribution

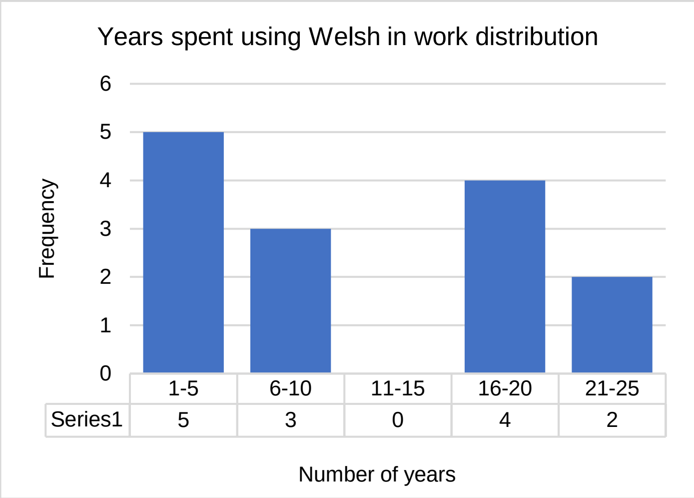


Figure 6.14 shows that of the 14 participants who use Welsh in a work or educational environment, six have spent more than 16 years, while the remaining eight participants have spent between one and ten years. Because the Welsh language is typically used more in a more prescriptive and formal way in education and work setting, than in day-to-day conversational environments, it is possible that an effect of the use of Welsh in a formal setting emerges. This would suggest that more frequent use of Welsh in a formal setting plays a role in the processing of Welsh gender. Each participant's number of years spent in a work or educational environment where Welsh is spoken, with years ranging from 0 – 22, is included as a continuous predictor in the models.

6.2.2.7 Inferential statistics

[19] Note that this also represents the age of some participants (i.e., 5 years of working in Welsh = 5 years since graduating from university), but not all.

In order to understand the possible roles of the six individual difference variables under investigation, the models presented for conditions 1 and 2 (Table 6.7), 3 (Table 6.8) and 4 (Table 6.9) are taken as the base models, and all six variables are then included as fixed effects to check for main effects. If any of the variables emerge as significant, separate models will be computed to understand their effect and possible interaction with linguistic conditions.

The first model computed includes the RTs for conditions 1 (gender in conjunction with mutations via the gendered numeral 2) and 2 (gender independent of mutations via the gendered numeral 4). Recall that the best-fitting model revealed no effect of grammaticality, condition, or gender. However, pairwise comparisons revealed that masculine items had increased reaction times than feminine item in condition 1, but not in condition 2.

Table 6.16

Best-fitting model: SPR Conditions 1 & 2 with IDs

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	504.16	16.88	29.94	**<0.001**
Grammaticality [grammatical]	-4.78	7.56	-0.63	0.528
Condition [1]	-16.10	7.02	-2.72	0.059
Gender [masculine]	3.08	10.25	0.30	0.764
Working Memory	75.21	9.60	7.83	**<0.001**
Declarative Memory	2.00	2.44	0.82	0.411
Procedural Memory	9.53	3.19	2.99	**0.003**
Dominance	-0.29	0.21	-1.36	0.173
Welsh Proficiency	-42.03	9.09	-4.62	**<0.001**
Welsh in Work	-6.36	1.97	-3.23	**0.001**
Grammaticality [grammatical] x Condition [1]	-1.97	15.51	-0.13	0.899
Grammaticality [grammatical] x Gender [masc]	-0.89	13.62	-0.07	0.948
Condition [1] x Gender [masc]	52.53	19.52	2.69	**0.007**
(Grammaticality [grammatical] x Condition [1]) x Gender [masc]	53.14	34.20	1.55	0.120
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	4037.00	63.53	1287.00	35.88
Grammaticality [grammatical]	533.50	23.10	-	-
Condition [1]	369.50	19.22	-	-
Gender [masculine]	821.60	28.66	-	-
Grammaticality [grammatical] x Condition [1]	2317.00	48.14	-	-
Grammaticality [grammatical] x Gender [masc]	1221.00	34.95	-	-
Condition [1] x Gender [masc]	3378.00	58.12	-	-
(Grammaticality [grammatical] x Condition [1]) x Gender [masc]	13420.0	115.84	-	-

As in the model presented in Table 6.7 (for conditions 1 and 2), the model's intercept is significant (β = 504.16, SE = 16.88, z = 29.94, p = <0.001). There is no significant effect of grammaticality (β = -4.78, SE = 7.56, z = -0.63, p = 0.528), condition (β = -16.10, SE = 7.02, z = -2.72, p = 0.059) or gender (β = 3.08, SE = 10.25, z = 0.30, p = 0.764). The model revealed a significant interaction between condition and gender (β = 52.53, SE = 19.52, z = 2.69, p = 0.007), but no other interactions reached significance (ps > 0.120).

The model also revealed effects of individual variables. It showed an effect of working memory (β = 75.21, SE = 9.60, z = 7.83, p <0.001), with higher working memory participants having slower reaction times. There was an effect of procedural memory (β = 9.53, SE = 3.19, z = 2.99, p = 0.003), with higher procedural memory participants having slower reaction times. It also showed an effect of Welsh linguistic proficiency (β = -42.03, SE = 9.09, z = -4.62, p <0.001), with more proficient speakers having faster reaction times. There was an effect of Welsh in work (β = -6.36, SE = 1.97, z = -3.23, p = 0.001), showing that participants with increased number of years of using Welsh in work having faster reaction times. However, there was no effect of declarative memory (β = 2.00, SE = 2.44, z = 0.82, p = 0.411) or language dominance (β = -0.29, SE = 0.21, z = -1.36, p = 0.173).

Four additional models are run, checking the interactions between the individual variable and the linguistic manipulations (grammaticality, condition & gender). The model with working memory only showed an interaction between working memory and grammaticality (β = 14.97, SE = 5.20, z = 2.88, p = 0.004), with faster RTs for grammatical items than ungrammatical items, for participants who have lower WM scores. There were no other significant interactions between (1) working memory and condition, (2) working memory and gender, (3) working memory, grammaticality, and condition, (4) working memory, grammaticality, and gender, or (5) working memory, grammaticality, condition, and gender (ps > 0.131). The model with procedural memory only did not reveal any statistically significant interactions between (1) procedural memory and grammaticality, (2) procedural memory and condition, (3) procedural memory and gender, (4) procedural memory, grammaticality, and condition, (5) procedural memory, grammaticality, and gender, or (6) procedural memory, grammaticality, condition, and gender (ps > 0.194). The model with Welsh linguistic proficiency only showed an interaction between proficiency and gender (β = -9.17, SE = 4.01, z = -2.29, p = 0.022), with faster RTs for feminine items than masculine items, for more proficient Welsh speakers. There were no other significant interactions between (1) proficiency and grammaticality, (2) proficiency and condition, (3) proficiency, grammaticality, and condition, (4) proficiency, grammaticality, and gender, or (5) proficiency, grammaticality, condition, and gender (*ps* > 0.071). The model with number of years spent in a Welsh speaking work or educational environment did not reveal any statistically significant interactions between (1) Welsh work/education and grammaticality, (2) Welsh work/education and condition, (3) Welsh work/education and gender, (4) Welsh work/education, grammaticality, and condition, (5) Welsh work/education, grammaticality, and gender, or (6) Welsh work/education, grammaticality, condition, and gender (*ps* > 0.076).

The next model includes the RTs for condition 3 (mutations independent of gender via pre-nominal adjectives). Recall that the best-fitting model revealed no effect of grammaticality or gender, nor was there a significant interaction between grammaticality and gender.

Table 6.17

Best-fitting model: SPR Condition 3 with IDs

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	526.73	65.52	8.04	**<0.001**
Grammaticality [grammatical]	16.05	37.85	0.42	0.672
Gender [masculine]	-9.50	40.41	-0.24	0.814
Working Memory	3.51	10.17	0.35	0.730
Declarative Memory	4.43	2.75	1.62	0.106
Procedural Memory	-0.53	3.59	-0.15	0.882
Dominance	0.07	0.24	0.28	0.776
Welsh Proficiency	-9.68	9.62	-1.01	0.315
Welsh in Work	-2.85	2.18	-1.31	0.192
Grammaticality [grammatical] x Gender [masc]	-16.23	24.83	-0.65	0.513
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	33530.0	183.11	1418.0	37.66
Grammaticality [grammatical]	15620.0	124.97	-	-
Gender [masculine]	12320.0	110.99	-	-
Grammaticality [grammatical] x Gender [masc]	7178.0	84.72	-	-

As shown in the base model (Table 6.8), the model's intercept is significant (β = 526.73, SE = 65.52, z = 8.04, p = <0.001). There is no significant effect of grammaticality (β = 16.05, SE = 37.85, z = 0.42, p = 0.672) or gender (β = -9.50, SE = 40.41, z = -0.24, p = 0.814) and there was no significant interaction between grammaticality and gender (β = -16.23, SE = 24.83, z = -0.65, p = 0.513). The model also revealed no significant effect of working memory (β = 3.51, SE = 10.17, z = 0.35, p = 0.730), declarative memory (β = 4.43, SE = 2.75, z = 1.62, p = 0.106), procedural memory (β = -0.53, SE = 3.59, z = -0.15, p = 0.882), dominance (β = -0.07, SE = 0.24, z = 0.28, p = 0.776), proficiency (β = -9.68, SE = 9.62, z = -1.01, p = 0.315) and Welsh in work (β = -2.85, SE = 2.18, z = -1.31, p = 0.192). Therefore, there is no evidence to suggest any different results to those revealed in the base model. Additionally, there is no evidence to suggest that any of the individual difference variables play a role in the processing of mutations when they operate independently of gender (via pre-nominal adjectives).

The final model(s) to be computed includes the RTs for condition 4 (gender encoded through mutations via the definite article 'y'). Recall that the best-fitting model revealed an effect of grammaticality, which surfaced in the critical segment but not the post-critical segments. However, there was no effect of gender.

Table 6.18

Best-fitting model: SPR Condition 4 with IDs

Fixed effects				
Parameters	Estimate	SE	*t*	Pr(>|z|)
Intercept	512.95	18.70	27.43	**<0.001**
Grammaticality [grammatical]	-42.39	9.57	-4.43	**<0.001**
Gender [masculine]	36.65	25.65	1.43	0.153
Working Memory	28.98	11.22	2.58	**0.010**
Declarative Memory	-2.16	2.89	-0.75	0.456
Procedural Memory	4.48	3.88	1.16	0.248
Dominance	0.10	0.26	0.38	0.704
Welsh Proficiency	-40.18	10.62	-3.78	**<0.001**
Welsh in Work	4.79	2.38	2.02	**0.044**
Random effects				
	By Subject		By Item	
Parameters	Variance	S.D.	Variance	S.D.
Intercept	4357.08	66.01	1846.73	42.97
Grammaticality [grammatical]	315.09	17.75	-	-
Gender [masculine]	2632.53	51.31	-	-
Grammatical [grammatical] x Gender [masc]	4517.85	67.22	-	-

As in the base model (Table 6.9), the model's intercept is significant (β = 512.95, SE = 18.70, z = 27.43, p = <0.001). There is also a significant effect of grammaticality (β = -42.39, SE = 9.57, z = -4.43, p <0.001), with decreased reaction times for grammatical items compared to ungrammatical items. However, there is no effect of gender (β = 36.65, SE = 25.65, z = 1.43, p = 0.153). The model also revealed effects of individual variables. It showed an effect of working memory (β = 28.98, SE = 11.22, z = 2.58, p = 0.010), with higher working memory participants having slower reaction times. It also showed an effect of Welsh linguistic proficiency (β = -40.18, SE = 10.62, z = -3.78, p <0.001), with more proficient speakers having faster reaction times. There was an effect of Welsh in work (β = 4.79, SE = 2.38, z = 2.02, p = 0.044), showing that participants with increased number of years of using Welsh in work having faster reaction times. However, there was no effect of declarative memory (β = -2.16, SE = 2.89, z = -0.75, p = 0.456), procedural memory (β = 4.48, SE = 3.88, z = 1.16, p = 0.248), or language dominance (β = 0.10, SE = 0.26, z = 0.38, p = 0.704).

Three additional models are run, checking the interactions between the individual variable and the linguistic manipulations (grammaticality, condition & gender). The model with working memory only, did not show any statistically significant interactions between working memory and grammaticality (p = 0.694) nor between working memory and gender (p = 0.451). The model with Welsh linguistic proficiency only, also did not show any statistically significant interactions between proficiency and grammaticality (p = 0.183) nor between proficiency and gender (p = 0.521). Similarly, the model with number of years spent in a Welsh speaking work or educational environment did not reveal any statistically significant interactions between Welsh work/education and grammaticality (p = 0.177) nor between Welsh work/education and gender (p = 0.588).

It is possible that the lack of significant interactions in these additional models is due to the fact that it includes all RTs for the fourth condition. Therefore, additional models are run, specifically on the critical segment of the sentence, as there is evidence suggesting that this is where the sensitivity to the violations emerged. The model with working memory only, did not show any statistically significant interactions between working memory and grammaticality (p = 0.207) nor between working memory and gender (p = 0.850). The model with Welsh linguistic proficiency only, also did not show any statistically significant interactions between proficiency and grammaticality (p = 0.423) nor between proficiency and gender (p = 0.537). Similarly, the model with number of years spent in a Welsh speaking work or educational environment did not reveal any statistically significant interactions between Welsh work/education and grammaticality (p = 0.362) nor between Welsh work/education and gender (p = 0.968).

Collectively, the additional models checking for interactions between working memory, Welsh linguistic proficiency and number of years spent in a Welsh speaking work or educational environment, and the two linguistic manipulations (grammaticality and gender), did not show any statistically significant interactions between any of the possible interactions. There is only evidence to suggest that each of these three individual variables are main effects, with (1) higher working memory participants having overall slower reaction times, (2) more proficient speakers having overall faster reaction times and (3) participants with increased number of years of using Welsh in work having overall faster reaction times.

The results from this section (6.2) are used to address the two research questions, (1) ‘To what extent do Welsh-English adult bilinguals show sensitivity to gender agreement violations when gender is independent of mutations, in conjunction with mutations and/or encoded through mutations, in local-contexts?’ and (2) ‘To what extent do cognitive and environmental individual differences play a role in Welsh-English adult bilinguals’ processing of gender in Welsh?’. The findings presented are discussed in light of these research questions and the relevant literature in the following section (6.3).

6.3 Discussion

The aim of this second experiment was to investigate the processing of grammatical gender in Welsh-English bilingual adults. It also set out to explore whether any of the patterns emerged from the first experiment (particularly the production data), are replicated in real time processing (Slabakova et al., 2020). A final goal was to investigate the role of individual difference variables, namely, working memory, declarative memory, procedural memory, language dominance, Welsh linguistic proficiency and the number of years spent using Welsh in a formal setting and/or work environment, in understanding whether each of these factors predict sensitivity to agreement violations in Welsh-English bilingual adults. Therefore, the following section summarises the results presented in section 6.2, followed by a discussion of these findings.

6.3.1 Results Summary and Welsh data Comparison

The Welsh-English bilingual adults completed a SPR task which included four experimental conditions. These linguistic conditions tested four different instances of gender with and without mutation(s). The results are summarised in this section and where possible, are briefly compared to experiment 1 findings and results from previous Welsh studies (for comprehension and production). However, this is the first study to examine the processing of grammatical gender in Welsh, therefore, it is not possible to make direct comparisons to any previous Welsh gender processing findings.

The results revealed that for condition 1, which tested gender in conjunction with mutations (via the gendered numeral 2) and condition 2, which tested gender independent of mutations (via the gendered numeral 4), there was no overall effect of grammaticality (p = 0.084), condition (p = 0.064), or gender (p = 0.530). There were no significant interactions between grammaticality and condition (p = 0.83), or grammaticality and gender (p = 0.95). However, there was a significant interaction between condition and gender (p = 0.005), showing slower reading times for masculine items compared to feminine items in condition 1 (p = 0.008), but with no difference in condition 2 (p = 0.16). To suggest sensitivity to the gender agreement errors in either of the two conditions, one would expect to see an overall effect of grammaticality and an interaction between grammaticality and conditions, however these results do not emerge from the model.

It was hypothesised that the Welsh speakers would show stronger sensitivity to gender agreement violations when gender is independent than when gender is in conjunction with mutations, however, there is no evidence to suggest sensitivity to the gender agreement violations in either of the two linguistic conditions. Therefore, neither of these predictions are met. There is also no evidence of a significant difference between the masculine and feminine items overall, yet, there was a significant difference between masculine and feminine items in condition 1 only.

Taken together, these results are interesting in light of those found in the production task in experiment 1. The production results showed that the participants produced the gendered numeral forms fairly accurately for both the masculine and feminine forms of *two* (‘dau’ [m] - 78.1%; ‘dwy’ [f] – 71.9%) and *four* (‘pedwar’ [m] – 82.2%; ‘pedair’ [f] – 70.3%). This shows that gendered marked numerals are a well-established gender feature in these Welsh speakers, suggesting that it is a fairly robust environment in Welsh for marking gender. Additionally, the production of nouns in bare form (i.e., no mutation) following ‘pedwar’ [m] (79.9%) and ‘pedair’ [f] (61.6%) are also fairly established. Therefore, it is somewhat surprising that there is no evidence of a grammaticality effect in condition 2, which tested gender independent of mutations, or even evidence of a gender effect in this condition, given that in experiment 1, the participants produced masculine nouns in bare form more accurately than feminine nouns. However, it could be argued that the lack of evidence to suggest a grammaticality effect in condition 1 is less surprising, given that in experiment 1, accuracy was lower for the production of nouns when they are soft-mutated following ‘dau’ [m] (66.3%) and ‘dwy’ [f] (55.6%). It may have also been expected to see an effect of gender here, but there was no evidence of such an effect.

The results revealed that for condition 3, which tested mutations independent of gender (via select pre-nominal adjectives), there was no effect of grammaticality ($p = 0.52$) or gender ($p = 0.98$), and there was no significant interaction between them ($p = 0.37$). To suggest sensitivity to the agreement violations, one would expect to see an effect of grammaticality, however this does not emerge from the model. It was also hypothesised that the bilinguals will show sensitivity to the agreement violations when mutations are independent of gender, however there is no evidence indicating this, therefore, this prediction is not met.

This instance of mutation operating independently of gender was not investigated in experiment 1, however, the results can be discussed in light of previous findings exploring the production of SM in non-gender contexts. For example, Thomas (2001) investigated the production of SM / no mutation on nouns following four prepositions, of which two trigger SM (*ar* “on” and *o* “from”) and two that do not trigger any mutation (*mewn* “in” and *(h)efo* “with”), in L1 Welsh speaking adults from North Wales. The results showed that the adults were more accurate in producing nouns in bare form following *(h)efo* “with” than *ar* “on” (p <

0.001) and *o* "from" ($p < 0.001$), as well as for nouns in bare form following *mewn* "in" than *ar* "on" ($p < 0.001$) and *o* "from" ($p < 0.001$). Errors in the expected production of SM nouns after *ar* "on" and *o* "from" were due to the adults producing them in bare form, rather than SM form. Despite the significant differences between nouns produced in bare form and SM nouns, the descriptive results showed that the SM nouns were still produced fairly accurately following *ar* "on" (73%) and *o* "from" (78%). Therefore, it may have been expected for the Welsh speakers in the current experiment to show evidence of sensitivity to the agreement violations on the post-adjectival noun, given that previous results in Thomas' study showed fairly accurate use of SM in non-gendered contexts. However, it may be that the use of mutations in gender-free contexts is more accurate for nouns in their bare form (i.e., no mutation) than SM nouns. One interesting take from this, is that the processing of agreement violations on nouns in non-gendered contexts could be compared in mutation gender-free contexts, between SM nouns and nouns produced in bare form. This could contribute to the overall understanding of the adults' processing patterns for gender mutation and non-gendered mutation contexts.

The results revealed that for condition 4, which tested gender encoded through mutations via the definite article 'y' (the), triggering SM on following feminine nouns but no mutation on masculine nouns, there was a significant effect of grammaticality ($p = 0.007$), which appeared to surface in the critical segment ($p = 0.005$) of the sentence but not the post-critical segments ($p = 0.098$), and there was no effect of gender ($p = 0.368$). To suggest sensitivity to the gender agreement violations, one would expect to see an effect of grammaticality, and this does emerge from the model. It was hypothesised that the Welsh speakers would show sensitivity to the gender agreement violations when gender is encoded through mutations in determiner-noun contexts, therefore, this prediction is met.

Similar to condition 3, gender encoded through mutations via the definite article 'y' was not tested in experiment 1 of this thesis, however, previous research has explored the production of Welsh gender in determiner-noun contexts. For example, Thomas (2001) investigated the production of grammatical gender in Det-Noun sequences in L1-Welsh speakers from North Wales. The results revealed that the adults were errorless in producing masculine nouns in their bare form, while accuracy for feminine nouns was 88%. The feminine noun errors were due to instances of producing the noun in bare form and two instances of double mutation.

Thomas suggested that the gender system is well established in adult speech, particularly in local contexts, but the adults were more accurate in masculine (no mutation) than feminine (with SM) noun production. Similarly, Sharp (2012) investigated the productive use of Welsh gender in Det-Noun contexts in L1 Welsh speaking adults from North Wales. The results revealed that the adults performed close to ceiling levels (80%+) for both feminine nouns (SM) and masculine nouns (bare form) and that there was no effect of noun gender. Sharp concluded that the Welsh adults have a productive command of the SM feature in relation to grammatical gender in local contexts. Therefore, in line with patterns emerging from previous studies, showing that the production of nouns following the definite article 'y' is highly accurate, the evidence of sensitivity to gender agreement violations when gender is encoded through mutations via 'y' is arguably unsurprising, but nonetheless interesting and important to see.

The results from this task are used to address research question one in this follow-up experiment, (1) 'To what extend do Welsh-English adult bilinguals show sensitivity to gender agreement violations when gender is independent of mutations, in conjunction with mutations and/or encoded through mutations, in local-contexts?'. Given that this is the first experiment to investigate the processing of grammatical gender in Welsh, using a SPR task, and there are no Welsh processing results to compare the Welsh bilinguals to, the following section briefly discusses the results in light of L1 / bilingual gender processing in other gendered languages. The results are then discussed in light of the theoretical approaches reviewed in chapter 5.

6.3.2 Previous L1 and Bilingual Processing Findings

The finding suggesting that the Welsh bilinguals showed sensitivity to the gender agreement violations when gender is encoded through mutations via 'y', is in line with previous L1 research. Several studies have investigated Det-Noun gender agreement violations in L1 Dutch speakers (e.g., Hagoort & Brown, 1999), L1 Spanish speakers (e.g., Barber & Carreiras, 2005) and L1 French speakers (e.g., Foucart & Frenck-Mestre, 2011), and found evidence of sensitivity to the violations. For example, Hagoort and Brown (1999) investigated Det-Noun gender disagreement in L1 Dutch speakers and found evidence of a P600 response, suggesting that the online processing of gender agreement information is a syntactic-form driven process, and not a conceptual/semantic process. Similarly, Barber and Carreiras (2005) examined Det-Noun gender agreement violations in L1 Spanish speakers and found a LAN effect, suggesting that they detected a mismatch between the morphosyntactic features when the nominal-phrase disagreed in gender. This was followed by reanalysis and repair processes which are reflected in the P600 component. Foucart and Frenck-Mestre (2011) also found similar results. They examined gender agreement violations in French L1 speakers, and the results showed a P600 response, suggesting that they detected the mismatch between the grammatical and ungrammatical forms, as well as reanalysis and repair processes.

Studies have also found highly proficient L2 speakers to show sensitivity to gender agreement violations. For instance, Tokowicz and MacWhinney (2005) tested gender agreement violations in L2 Spanish learners (L1 English) and the results showed P600s in response to Det-Noun gender agreement violations. The authors indicated that the L2ers were implicitly sensitive to these agreement violations. Similarly, Dowens, Guo, Guo, Barber and Carreiras (2011) investigated Spanish article-noun gender agreement violations in highly proficient L2 Spanish learners (L1 Chinese) and found that the L2ers evidenced a P600 effect. They argued that with increased proficiency, gender features that are not present in L1, can be acquired by L2 speakers. Arguably, the result from this experiment, suggesting that the Welsh bilinguals showed sensitivity to the gender agreement violations determiner-noun contexts, is in line with both L1 and L2 previous research. Notably, the Welsh speaking adults in this experiment are not monolingual Welsh speakers, nor are they consider L2 speakers of Welsh, they are highly proficient speakers of Welsh who are bilingual with English.

Empirical studies have also explored how early bilingual populations process gender agreement violations. One such study is by Alarcón (2020), who investigated gender agreement processing in highly proficient early (heritage language learners) and late Spanish-English bilinguals, as well as Spanish monolinguals. The results from a timed GJT showed that accuracy was similar across the 3 groups – the HLL scored 96.93%, the L2 learners scored 97.5% and the monolinguals scored 98.15% - with no significant difference between them. Reaction time data also showed similar RTs across the 3 groups, with an average of 1733ms for the HLL, 1625ms for the L2 learners and 1550ms for the monolinguals, again, with no significant difference between them. This indicates that performance was close to ceiling, with accuracy rates above 96%, and statistically indistinguishable RTs, averaging 1635.9ms. Alarcon suggested that early and late bilinguals displayed similar grammatical gender knowledge in their underlying grammars to L1 speakers.

In light of Alarcón's (2020) results, it is possible to suggest that the Welsh bilinguals detected the mismatch between the grammatical and ungrammatical forms, showing that they are able to build and manipulate the agreement information between the determiner and the noun. It is likely that they have grammatical gender knowledge in their underlying grammars and can therefore compute detailed syntactic representations during real time (Hopp, 2010). However, as noted previously, there is only evidence to suggest sensitivity to gender agreement errors in determiner-noun contexts (linguistic condition 4). Two other gender contexts were examined in this experiment, when gender was in conjunction with mutations (via '2') and when gender was independent of mutations (via '4'). The results from these two linguistic conditions did not point towards evidence of sensitivity in either instance. Therefore, it is possible to wonder whether any of the predictions made by theoretical accounts of L2 processing, including the SSH (Clahsen & Felser, 2006, 2018), the DP model of lexicon and grammar (Ullman, 2001, 2004, 2005, 2015, 2020) and computational/cognitive capacity accounts (e.g., Hopp, 2010; McDonald, 2006), can be used to explain the lack of evidence suggesting such sensitivity in these two instances.

6.3.3 Shallow Structure Hypothesis (SSH)

The SSH predicts that the syntactic representations computed by L2 speakers during comprehension are shallower and less detailed than those of L1 speakers, where grammatical representations lack complex hierarchical structure and more abstract elements of syntactic structure (Clahsen & Felser, 2006, 2018). Evidence to support such claims would be findings suggesting that the speaker / learner constructs underspecified syntactic representations and that they do not reflect early automatic structure-building processes. In turn, the absence of evidence indicating sensitivity to gender agreement violations, is sometimes used to suggest shallow processing strategies in these speakers / learners (e.g., Keating, 2009). Therefore, under the assumptions made by the SSH, it could be argued that the lack of evidence showing sensitivity to the gender agreement errors in linguistic conditions 1 (gender in conjunction with mutations) and 2 (gender independent of mutations) points towards evidence of 'shallow' processing. These results could be interpreted to indicate deficits in processing, suggesting that the Welsh adult speakers are experiencing problems building the abstract syntactic presentations in these two linguistic conditions. For instance, it could be suggested that there is difficulty in manipulating the necessary and relevant agreement information during real time, such as the agreement between the gendered numeral and the noun. This difficulty in turn points towards evidence of shallow and less detailed processing in these Welsh-English bilinguals. Additionally, it could be suggested that the shallow processing in the Welsh speakers may be due to a lack of knowledge of a particular grammatical rule or be in possession of incorrect knowledge regarding that particular rule. For example, it may be that they lack knowledge of Welsh gender and its different instantiations (e.g., in conjunction with, encoded through or independent of mutations), or that these abstract representations of the different instantiations are incorrect. Therefore, the linguistic representations cannot be computed during real-time.

Alternatively, in place of suggesting that the lack of evidence suggesting sensitivity to the gender agreement violations in these conditions serves as evidence of shallow processing, it is tentatively suggested that the violations in these two gender-contexts are not 'strong' enough for the Welsh speakers to recover from. This may be because the co-occurrences between the different agreement types are less frequent and/or are contradictory in the input. Previous research has found that speakers often rely on co-occurrence relations between gender marked modifiers and nouns (Grüter et al., 2012). Therefore, if the different gender agreement areas are rare in the input and/or are used in a more variable manner, these may be too simple to process, given the potential variability and/or inconsistency in the input. It is therefore possible to wonder because gender and mutations in Welsh are used in a variable manner both within and across speakers, these errors are subconsciously accepted. Therefore, the Welsh speakers are only processing certain errors in Welsh. It may be that there is only evidence suggesting that the Welsh speakers are processing errors following the determiner 'y', because this is the most common instantiation of gender in Welsh (det-noun sequences). Furthermore, it would be difficult to argue that the lack of evidence suggesting sensitivity to the gender agreement violations in these conditions is because the Welsh speakers lack knowledge of a particular grammatical rule, given the results in the production task in experiment 1. The results indicated that the production of gendered numeral nouns is relatively high, as well as the production of nouns following the gendered forms for ‘2’ and ‘4’. Therefore, it is not possible to argue that the Welsh speakers lack knowledge of the particular grammatical rule (i.e., numeral-noun contexts).

In light of the predictions made by the SSH and the findings from the current experiment, the SSH is not considered a suitable theoretical approach to explain the patterns emerging from the two linguistic conditions, which do not indicate sensitivity to gender agreement errors. The reasons behind this argument are outlined above. Alternatively, it may be that the predictions made by the SSH are unsuitable, given that they are intended for L2 adult learners and the bilinguals in the current experiment are not adult learners of Welsh. Additionally, the basis for much of the support for the SSH comes from studies comparing results between L1 speakers and L2 learners, where L2ers fail to demonstrate processing patterns similar to L1 speakers. Notably, this comparison is not possible to make, given that the highly proficient Welsh-English bilinguals may themselves be considered as the ‘base-line’ for Welsh adult speakers. Therefore, it is suggested that the predictions made by the SSH are not extended to the highly proficient Welsh-English adult bilinguals in this experiment.

6.3.4 Declarative Procedural (DP) Model

An alternative theoretical approach to try and explain why L2 speakers might not demonstrate processing patterns similar to L1 speakers, includes the DP model of lexicon and grammar (Ullman, 2001, 2004, 2005, 2015, 2020). The DP model argues that in the L1, aspects of grammar, in particular rule-governed structure building, are generally subserved by procedural memory, however, in the L2 (particularly at lower levels of exposure and corresponding proficiency), these learners rely on lexical/semantic processes in the declarative memory for the same functions (Hamrick et al., 2018; Morgan-Short et al., 2014; Stefaniak et al., 2021; Ullman, 2001, 2015). With increasing L2 exposure, experience and proficiency, this can lead to the proceduralization of grammar, where aspects of grammar may come to rely more on the same procedural memory system as those that underlie L1 grammar (Ullman, 2005; Ullman & Lovelett, 2018). Therefore, grammar should tend to rely more heavily on procedural memory for early-learned languages (L1 or L2) and more on declarative memory for later learned languages (Morgan-Short et al., 2022).

Based on the assumptions made by the DP model of lexicon and grammar, it would predict that procedural memory should be important for the processing of gender agreement violations, whereas the declarative memory system is not (Morgan-Short et al., 2010). However, if the Welsh bilinguals showed processing patterns similar to L1 speakers, neither declarative nor procedural memory should play a role in the processing of Welsh gender. Therefore, if there was evidence to suggest that declarative memory plays a role, it could be argued that the Welsh bilinguals demonstrate reliance on lexical/semantic processes in the declarative memory system. Alternatively, if there was evidence to suggest that procedural memory plays a role, it could be argued that the Welsh bilinguals show proceduralization of grammar, where aspects of grammar may come to rely more on the same procedural memory system as those that underlie L1 grammar (Ullman, 2005; Ullman & Lovelett, 2018). For either declarative or procedural memory to be understood as playing a role in the processing of Welsh gender, one would expect to see an interaction between declarative and/or procedural memory and grammaticality in each (or any) of the four linguistic conditions, showing that declarative and/or procedural memory predicts faster reading times for grammatical items compared to ungrammatical items.

The results revealed that for linguistic conditions 1 (gender in conjunction with mutations) and 2 (gender independent of mutations), declarative memory did not emerge as a significant effect (p = 0.411), whereas procedural memory did (p = 0.003), showing that participants with higher procedural memory had overall slower reaction times. An additional model was run to check the interactions between the linguistic manipulations (grammaticality, condition, and gender) and procedural memory, and it revealed no statistically significant results between any of the possible interactions (ps > 0.194). Therefore, even though in the main model, there is evidence to suggest that participants with lower procedural memory had overall faster reaction times (with lower procedural memory scores indicating an improvement in procedural memory learning ability), there is no evidence to suggest that procedural memory predicts faster reading times for grammatical items compared to ungrammatical items, in either linguistic condition. Similarly, there is no evidence to suggest that declarative memory plays a role in the processing of gender agreement violations when gender operates in conjunction with mutations and independent of mutations. For linguistic condition 3 (mutations independent of gender), neither declarative memory (p = 0.106) or procedural memory (p = 0.882) emerged as significant effects in the model. Therefore, there is no evidence to suggest that either play a role in the processing of agreement violations when mutations operate independently of gender. Similarly, for linguistic condition 4 (gender encoded through mutations via 'y'), neither declarative memory (p = 0.456) or procedural memory (p = 0.248) emerged as significant effects in the model. Therefore, there is no evidence to suggest that either play a role in the processing of gender agreement violations when gender is encoded through mutations.

Given that there is no evidence to suggest that declarative nor procedural memory play a role in the processing of agreement violations in any of the four linguistic conditions, it could be argued that this serves as evidence to suggest that the Welsh bilinguals are emulating L1-like processing patterns. This would suggest that the Welsh bilinguals do not rely on lexical/semantic processes to process the violations nor do they show the proceduralization of grammar (Ullman, 2005; Ullman & Lovelett, 2018). Therefore, it can be argued that both predictions are met, which assumed that neither procedural memory nor declarative memory would play a role.

These claims hold in light of previous findings. For example, Morgan-Short et al (2010) found that low proficient learners of an artificial language learning paradigm relied on lexical/semantic processes (i.e., declarative memory) for gender agreement processing, whereas highly proficient learners depended on procedural mechanism, similar to L1 processing mechanisms, for agreement processing. Similar results were found by Morgan-Short et al (2014), who administered the CVMT as a measure of declarative memory and the TOL as a measure of procedural memory, and found a relationship between declarative learning ability and syntactic development at early stages of acquisition, and a relationship between procedural learning ability and syntactic development at later stages of acquisition. They suggested that declarative memory learning abilities predicted L2 grammatical development at an early stage of L2 acquisition, while procedural memory learning abilities predicted L2 grammatical development at a later stage. However, one important consideration is the fact that there is no evidence to suggest sensitivity to the agreement violations in the first three linguistic conditions, therefore, this may be why there is no evidence of an effect of procedural or declarative memory in any of these conditions. There is only evidence to suggest sensitivity to the gender agreement violations in the fourth condition, where gender is encoded through mutations. Nonetheless, the absence of an effect of declarative or procedural memory, would support the arguments that the Welsh bilinguals are emulating L1-like processing patterns, suggesting that they do not rely on lexical/semantic processes to process the violations nor do they show the proceduralization of grammar (Ullman, 2005; Ullman & Lovelett, 2018).

Taken together, it is difficult to say whether this model is considered a suitable theoretical approach to explain the processing patterns emerging from this SPR experiment. On one hand, the assumptions made by the DP model can appropriately explain why there is no evidence that procedural or declarative memory play a role in the processing of agreement violations in the Welsh-English bilinguals, suggesting that they may be showing L1-like processing patterns. However, on the other hand, it could be argued that this is only relevant for the fourth linguistic condition, as there is only evidence to suggest sensitivity to agreement violations when gender is encoded through mutations, then when gender is in conjunction with mutations (condition 1), when gender is independent of mutations (condition 2) and when mutations are independent of gender (condition 3). Future research could consider testing new speakers of Welsh and comparing results to early speakers of Welsh to gain a clearer understanding of the potential roles of procedural and/or declarative memory in the processing of agreement violations in Welsh adult speakers.

6.3.5 Computational / Cognitive Capacity Accounts

Computational / cognitive capacity accounts have also been used to try and explain why L2 speakers might not demonstrate processing patterns similar to L1 speakers (e.g., Hopp, 2010; McDonald, 2006). In contrast to the SSH and the DP model, the cognitive accounts argue for quantitative differences, mainly attributed to cognitive resource limitations. Hopp (2010) and McDonald (2006) claim that processing in the L2 is more taxing for memory resources than L1 processing, with increased cognitive demands hindering access to working memory and other attentional and decoding abilities. The non-target-like performance by advanced L2ers reduces to the overburdening of a L1-like processing system (Hopp, 2014), therefore reflecting quantitatively non-target-like processing in the L2 due to capacity differences, rather than fundamental differences in grammar for the parser (as per the SSH). Therefore, given that it was predicted that the Welsh bilinguals would show processing patterns similar to L1 speakers, it was also hypothesised that working memory would not emerge as an important predictor, as the Welsh speakers would show ceiling effects. In turn, for there to be evidence suggesting that working memory plays a role in the processing of Welsh gender, and thus reflecting L2-like processing patterns, one would expect to see an interaction between working memory and grammaticality in each (or any) of the four linguistic conditions, showing that working memory predicts faster reading times for grammatical items compared to ungrammatical items.

The results revealed that for conditions 1 (gender in conjunction with mutations) and 2 (gender independent of mutations), working memory emerged as a significant effect ($p < 0.001$), showing that participants with higher working memory had overall slower reaction times. An additional model was run to check the interactions between the linguistic manipulations (grammaticality, condition, and gender) and working memory, and it revealed a significant interaction between grammaticality and working memory ($p = 0.004$), showing an effect of grammaticality (grammatical faster than ungrammatical) for the participants with lower working memory scores. No other interactions were significant ($ps > 0.131$). This suggests that working memory predicts faster reading times for grammatical items compared to ungrammatical items when gender is in conjunction with mutations and independent of mutations, but for participants with lower working memory scores. For linguistic condition 3 (mutations independent of gender), working memory did not emerge as a significant effect in the model ($p = 0.730$). Therefore, there is no evidence to suggest that working memory plays a role in the processing of agreement violations when mutations operate independently of gender. For linguistic condition 4 (gender encoded through mutations), working memory emerged as a significant effect ($p = 0.010$), showing that participants with higher working memory had overall slower reaction times. This is similar to conditions 1 and 2. However, an additional model was run to check the interactions between the linguistic manipulations (grammaticality and gender) and working memory, and it revealed no statistically significant results between any of the possible interactions (*ps* > 0.451). Therefore, there is no evidence to suggest that working memory plays a role in the processing of gender agreement violations when gender is encoded through mutations.

Taken together, it can be said that there is evidence to suggest that working memory plays a role in gender processing when it is in conjunction with mutations and independent of mutations, showing an effect of grammaticality at lower levels of working memory. In contrast, there is no evidence to suggest that working memory plays a role in the processing of agreement violations when mutations are independent of gender or when gender is encoded through mutations. Therefore, part of the prediction is met, with some evidence suggesting it does not play a role, while there is other evidence indicating that it does, but at lower levels of working memory.

This is interesting in light of previous findings, as studies have found that higher working memory scores predict grammaticality effects. For example, Sagarra and Herschensohn (2010) examined N-Adj agreement violations in L2 Spanish learners and found that L2 learners with higher working memory were more sensitive to gender agreement errors than those with lower working memory. Similarly, Keating (2010) also found that higher-span L2 learners showed increased sensitivity to N-Adj agreement violations in Spanish. The results indicated that higher-span learners spent more time reading ungrammatical adjectives relative to grammatical controls. In an ERP study, Gabriele et al (2021) also found working memory effects in advanced L2 Spanish learners in the processing of Spanish N-Adj gender violations. The results revealed that learners with better working memory showed larger positivities for N-Adj gender violations than those with lower working memory. In turn, they suggested that working memory plays a role when linguistic features need to be encoded and maintained during online processing, which involves a speaker's ability to retrieve the relevant linguistic features from memory and the online deployment of knowledge of agreement. In contrast, previous research has not found working memory effects in L1 gender agreement processing. Sagarra and Herschensohn (2010) found no WM effects in the L1 Spanish speakers processing of

N-Adj agreement violations. They suggested that this was a result of ceiling effects, proposing that working memory effects become irrelevant at ceiling levels in the processing of agreement errors. Based on this, it was predicted that working memory would not play a role in the processing of gender in Welsh bilinguals, given that the Welsh bilinguals would demonstrate L1-like processing patterns.

As noted previously, part of the prediction is met, with some evidence suggesting it does not play a role, while there is other evidence indicating that it does, but at lower levels of working memory. If there was evidence to suggest that higher working memory predicts differences in reading times, with faster reading times for grammatical compared to ungrammatical items, it may be possible to suggest that processing these instantiations of agreement places a greater burden on working memory (Hopp, 2006) and the Welsh bilinguals reflect patterns similar to L2 learners with high working memory. However, this is not the case. One possible explanation for the effect found in conditions 1 and 2, is that the higher working-memory capacity participants rely less on any kind of chunking strategy (Hopp, 2014). Therefore, instead of interpreting the different grammatical and ungrammatical agreement types in the sentences, the participants almost 'skim' over these parts, without incrementally processing the target structures (Hopp, 2014). It may be that the participants show no evidence of storing and integrating sources of linguistic information whilst paying attention to upcoming parts of the sentence (Baddeley, 2003; Sagarra & Herschensohn, 2010). Yet, this suggestion is tentative given that working memory only appears to play a role in gender processing when it is in conjunction with mutations and independent of mutations, showing an effect of grammaticality, but at lower levels of working memory.

Similar to the conclusions made for the DP model, it is difficult to say whether computational / cognitive capacity accounts are considered suitable theoretical approaches to explain the processing patterns emerging from this SPR experiment. On one hand, it can appropriately explain why there is no evidence that working memory plays a role in the processing of agreement violations when mutations are independent of gender and when gender is encoded through mutations, suggesting that they may be showing L1-like processing patterns. However, on the other hand, it struggles to explain the pattern emerging from conditions 1 and 2. Future research could also consider testing new speakers of Welsh and comparing results to early speakers of Welsh to gain a clearer understanding of the potential role of working memory in the processing of agreement violations in Welsh adult speakers.

6.3.6 Environmental Individual Difference Variables

Three environmental individual difference variables were also considered in this follow-up experiment, including language dominance, Welsh linguistic proficiency and the number of years spent using Welsh in an educational and/or professional setting. The results for the variables are discussed below.

It was hypothesised that language dominance would predict outcomes in the Welsh-English bilinguals' processing of gender. However, there was no significant effect of language dominance on reading times in any of the linguistic conditions (*ps* > 0.173), therefore suggesting that language dominance does not play a role in the processing of gender in Welsh. This is in line with results from experiment 1, which showed no evidence that dominance played a role in the comprehension or production of Welsh grammatical gender. However, it is in contrast with previous studies, which have found language dominance to be a predictive factor in linguistic outcomes in various different phenomena. For example, Amengual (2016b) found that language dominance was a strong predictor of performance in the production of the Catalan back mid-vowel contrast, with Catalan-Spanish bilinguals who were Catalan dominant having higher accuracy rates than the bilinguals who were Spanish dominant. Perpiñán (2017) also found language dominance to play a role in the production of the expression of Catalan clitics in simultaneous bilinguals. They found that Catalan-Spanish bilinguals who were Spanish dominant, exhibited higher errors rates than Catalan dominant speakers. Similar results were found by Bonvin, Brugger and Berthele (2021) who found a strong linear association between dominance and vocabulary proficiency.

One possible explanation for the lack of dominance effect may be because the results from the BLP were obtained as part of experiment 1, where data was collected in the winter of 2020, and these scores were included and linked with results from experiment 2, where data was collected in the spring of 2022. Given the fact the Welsh speakers are considered along a continuum of language dominance (Amengual, 2015), it may be that the participants' use of Welsh (and English) has shifted in the time between the two experiments. For instance, Solís-Barroso and Stefanich (2019) noted that a bilingual's language dominance score would vary depending on when the participant is assessed. Therefore, it is possible this has shifted, in turn adjusting the true reflection of the Welsh speakers' language dominance global score. Furthermore, research has found that gradient changes in the speakers' linguistic environment results in shifts between the two (or more) languages (Oppenheim et al., 2020). As a result, it is possible that the BLP scores collected in 2020 reflect different current and cumulative language experience and use in the Welsh bilinguals, particularly given the different environments in which the participants completed the two experiments, specifically, during the COVID-19 pandemic, where individuals spent a significant amount of time at home compared to coming out of the pandemic, where working in the workplace was being normalised again.

The role of Welsh linguistic proficiency was also explored in the processing of gender in Welsh. It was hypothesised that it would be found to play a role, particularly as evidence emerged from experiment 1, showing that Welsh proficiency predicted outcomes on comprehension and production measures in Welsh bilinguals. The results from this follow-up experiment revealed an effect of Welsh proficiency for linguistic conditions 1 (gender in conjunction with mutations), 2 (gender independent of mutations) ($p < 0.001$) and 4 (gender encoded through mutations) ($p < 0.001$), but not 3 (mutations independent of gender) ($p < 0.315$). For linguistic conditions 1 and 2, the results revealed that more proficient speakers had overall faster reading times. An additional model was run to check the interactions between the linguistic manipulations and proficiency, and it revealed a significant interaction between gender and proficiency ($p = 0.022$), showing faster reading times for feminine items than masculine items, for more proficient Welsh speakers. No other interactions were significant (*ps* > 0.071). For linguistic condition 4, the results also revealed that more proficient speakers had overall faster reading times. An additional model was run to check the interactions between the linguistic manipulations and proficiency, and it revealed no statistically significant results between any of the possible interactions (*ps* > 0.183).

Despite showing that increased proficiency scores predict faster reading times, there is no evidence to specifically suggest that proficiency predicts faster reading times for grammatical items compared to ungrammatical items in linguistic conditions 1, 2 or 4. Therefore, it is not possible to say that proficiency plays a role in the processing of gender agreement violations. This lack of evidence is in contrast to previous findings. For instance, Keating (2009) found that target-like processing of gender agreement violations in Spanish was subject to proficiency levels, with higher proficiency speakers showing increased sensitivity to the gender agreement errors. One possible explanation for the lack of evidence specifically suggesting that proficiency predicts faster reading times for grammatical items compared to ungrammatical items in linguistic conditions 1, 2 or 4, may be because the Welsh proficiency scores are extremely high (95.45%). There is very little variability in the groups scores (*SD*=4.8%), as out of the 21 participants, four scored 100%, twelve scored between 95-99%, two scored between 90-94%, two scored between 85-89% and one scored between 80-84%. This indicates that the Welsh speakers are highly proficient speakers of Welsh. It is possible that if there was more variability in the proficiency scores, then proficiency may be found to modulate the processing of gender. Future work could collect data from Welsh speakers of varying proficiency levels and examine its role in relation to the processing of Welsh gender.

The possible role of the number of years spent using Welsh in an educational and/or professional setting was also considered in this experiment. It was hypothesised that it would be found to play a role, because the results from experiment 1 indicated that there could be more informative / different factors relevant to Welsh bilingual speakers, than maybe the more ‘traditional’ individual differences. The results from this follow-up experiment revealed an effect of number of years spent using Welsh in an educational and/or professional setting for linguistic conditions 1, 2 ($p = 0.001$) and 4 ($p = 0.044$), but not 3 ($p = 0.192$). For linguistic conditions 1 and 2, the results revealed that participants with increased number of years using Welsh in work had overall faster reading times. An additional model was run to check the interactions between the linguistic manipulations and Welsh in work, and it revealed no statistically significant results between any of the possible interactions (*ps* > 0.076). For linguistic condition 4, the results also revealed that participants with increased number of years using Welsh in work had overall faster reading times. An additional model was run to check the interactions between the linguistic manipulations and Welsh in work, and it revealed no statistically significant results between any of the possible interactions (*ps* > 0.177).

Similar to the role of Welsh proficiency, despite the results showing that the number of years spent using Welsh in an educational and/or professional setting predicted overall faster reading times, there is no evidence to specifically suggest that this variable predicts faster reading times for grammatical items compared to ungrammatical items in linguistic conditions 1, 2 or 4. This is likely not to have emerged in conditions 1 and 2 as there is no evidence to suggest sensitivity to the gender agreement errors in these two conditions anyhow. However, it could have been expected to predict faster reading times for grammatical items compared to ungrammatical items in linguistic condition 4, given that there is evidence to suggest sensitivity to gender agreement violations when gender is encoded through mutations. Therefore, it is not possible to say that the number of years spent using Welsh in an educational and/or professional setting plays a role in the processing of gender agreement violations.

To my knowledge, this is the first time a factor such as this has been considered in relation to Welsh adults' use of grammatical gender. This factor was thought of in light of experiment 1 findings, which tentatively indicated that more environmental factors could be related to the comprehension and production of grammatical gender in Welsh speakers. Additionally, it was considered in light of previous research, which suggested that future work should consider how Welsh–English bilingual adults' lives influence their language proficiency in more detail, in order to establish which factors predict and influence higher performance on measures of morphology (Binks & Thomas, 2019). This factor is exploratory and could be investigated further, to understand the different variables that may play a role in the processing of gender in Welsh. This factor, arguably, is intrinsically linked with the amount a Welsh speaker reads, writes, and speaks in Welsh. Therefore, future work could consider Welsh adult speakers' contact with Welsh in more depth and detail.

The individual difference variable findings have been discussed in light of the second research question posed in this follow-up experiment, (2) 'To what extent do cognitive and environmental individual differences play a role in Welsh-English adult bilinguals' processing of gender in Welsh?'. The variables considered in this follow-up experiment are not the only factors that could be investigated and explored in relation to the processing of Welsh gender, however, considering these six variables is a step in the right direction. This experiment has attempted to explore a sub-set of internal and external factors which may influence gender processing in Welsh. Alternative internal (e.g., age of acquisition of the two languages, cognitive aptitude) and external variables (e.g., language attitudes towards the target language, motivation) could be considered in future work.

6.4 Chapter Summary

This second experiment set out to explore four things:

- To investigate how Welsh-English bilingual adults process agreement violations when gender is in conjunction with mutations, when gender is independent of mutations, when gender is encoded through mutations and when mutations are independent of gender.
- To explore whether any of the patterns emerged from the first experiment (particularly the production data), are replicated in real time processing.
- To consider the possible roles of, working memory, declarative memory, procedural memory, language dominance, Welsh linguistic proficiency and the number of years spent using Welsh in a formal setting and/or work environment, in understanding whether each or any of these factors predict sensitivity to agreement violations in Welsh-English bilingual adults.
- To see if patterns emerging from the processing data could be explained by the assumptions made by theoretical accounts of L2 / bilingual processing.

This section (6.3) discussed the results presented in section 6.2 of this chapter. A SPR task was conducted to attempt to understand how highly proficient Welsh-English bilingual adults process agreement errors during real time, and the possible roles of various individual variables. I have attempted to extend the predictions made by SSH, DP model and computational / cognitive capacity accounts to explain the patterns emerging from the data. It is possible to say that there is no obvious account that suitably describes these patterns, despite there being some aspects from the different approaches that are relevant. Notably, the theoretical approaches are designed to account for variability in L2 learners, describing how and why they might differ from L1 speakers, and importantly, the Welsh speakers in this experiment are not L2 learners of Welsh. Therefore, it could be argued that the theoretical approaches are not extendable to the Welsh bilinguals under investigation. Nonetheless, discussing the results in light of these approaches is enlightening and informative for the field, given that these Welsh-English adult bilinguals speakers use Welsh in minority language conditions, with English as the dominant societal language. It is possible that there is alternative theoretical approach that is more appropriate to explain the processing data, however, this exploration is beyond the scope of this thesis.

The following chapter brings together the findings from the two experiments to discuss the overarching themes of this thesis, which are:

- How do Welsh adult speakers, from a more geographically diverse population than those tested previously, comprehend, produce, and process grammatical gender in Welsh?
- Is the variable use of gender and mutations in Welsh adult speakers due to a lack of gender knowledge or rather, is it an indication of adults' lack of ability and consistency with producing mutations in general?
- Is there evidence to suggest that the comprehension, production, and processing of Welsh grammatical gender is more robust when gender operates without mutations than when it operates with mutations?
- To what extent do various individual difference variables play a role in the comprehension, production, and processing of Welsh grammatical gender?

This general discussion is presented Chapter 7.

Chapter 7

General Discussion and Concluding Remarks

This final chapter presents a general discussion by bringing together the findings from the two experiments. It also presents the conclusions drawn from the two experiments, discusses some of the limitations and offers some suggestions for future research.

7.1 General discussion

This thesis has investigated the comprehension, production, and processing of Welsh grammatical gender in Welsh-English bilingual adults. In experiment one, the comprehension findings addressing the first research question revealed that the Welsh speakers showed fairly accurate use of grammatical gender in comprehension when gender is independent of mutations and when gender is encoded through mutations, in non-local contexts. However, the individual variation in the group indicates that some of the participants show accurate use of gender in comprehension, while other participants show less accurate use of gender, when it is marked via possessive adjective forms (+mutations) and anaphoric pronouns (-mutations) in non-local contexts. Additionally, there is no evidence to suggest that accuracy is higher for gender when it is independent of mutations than when it is encoded through mutations, overall or for either gender specifically (masculine / feminine). Furthermore, the findings addressing the sub-research question revealed that there is evidence to suggest a relationship between Welsh linguistic proficiency and the comprehension of gender, but there is no evidence to suggest that proficiency affected either of the linguistic conditions differently. Moreover, there is no evidence to indicate that language dominance plays a role in the comprehension of gender in non-local contexts.

The production findings addressing the second research question revealed that the Welsh speakers have a good productive command of grammatical gender when gender is marked via gendered numerals, however, there is notable individual variation in the data. The results showed better performance for masculine numerals than feminine numerals, but there was no significant difference in performance between the different numeral forms. The gender-mutation accuracy results revealed no effect of gender but generally showed better performance on nouns produced in bare form compared to mutated nouns. Further analysis showed that no-mutation accuracy is higher for masculine nouns (following numeral 4) compared to feminine nouns (following numerals 3 & 4). Also, noun mutation accuracy is higher when it is SM (following numeral 2 [f]) compared to AM (following numeral 3 [m]). Collectively, this suggests that accuracy is highest when nouns are produced in bare form in comparison to when they are mutated, and when nouns are mutated, accuracy is higher for SM nouns (feminine only) in comparison to AM nouns (masculine). Furthermore, the findings addressing the sub-research question revealed that there is evidence to suggest there is a relationship between Welsh linguistic proficiency and the production of gender, for both gendered numeral form accuracy and for noun mutation / no-mutation accuracy, but there is no evidence to suggest that proficiency affected the linguistic instantiations differently. Moreover, there is no evidence to indicate that language dominance plays a role in the production of gender in local contexts (for either the production of numerals or nouns).

These results were discussed in light of theoretical approaches to L2 attainment, including the Interpretability Hypothesis (IH), and the Missing Surface Inflection Hypothesis (MSIH). It was concluded that predictions from both approaches were suitable to account for the bilinguals' production of gendered numeral forms in the production task, however, only predictions made by the MSIH could appropriately describe the patterns emerging from the noun mutation / no-mutation performance. Importantly, the MSIH does not fully explain all of the patterns emerging from the data, but it can explain the data in some parts, and this is promising given the complexity of the gender system in Welsh. Furthermore, only the predictions made by the IH were used to try and explain the pattens emerging from the comprehension task, given that the MSIH does not make explicit predictions for comprehension and was therefore not considered an appropriate theoretical approach to try and explain the data. It was concluded that the predictions made by the IH could not appropriately describe the patterns emerging the comprehension task. Taken together, it is not possible to say that one approach can account for all of the findings in the first experiment. The lack of applicability to the Welsh data may be because it is challenging to disentangle the potential sources of difficulty in the context of the gender and mutations in Welsh, or, alternatively, because the theoretical approaches are intended for L2 adult learners and the Welsh bilinguals who participated in this research are not L2 adult learners of Welsh.

The second experiment explored the processing of Welsh gender in a sub-group of the Welsh speakers from experiment one. The results addressing the main research question in the follow-up experiment revealed that there is no evidence to suggest sensitivity to agreement violations when gender is in conjunction with mutations, when gender is independent of mutations or when mutations are independent of gender. However, there is evidence to suggest sensitivity to gender agreement violations when gender is encoded through mutations, in determiner-noun phrases.

Regarding the cognitive and environmental individual differences, the findings addressing the second research question in the follow-up experiment revealed that when gender operated in conjunction with mutations and independent of mutations, procedural memory, Welsh proficiency, and the number of years spent in a Welsh speaking work or educational environment, predicted overall faster reading times. In contrast, working memory predicted overall slower reading times, revealing sensitivity to the gender agreement errors for Welsh bilinguals with lower working memory scores. Additionally, there was no evidence to suggest that declarative memory or language dominance played a role in either of the linguistic conditions. When mutations operated independently of gender, the results did not suggest that any of the variables predicted stronger sensitivity to the agreement violations. When gender was encoded through mutations, the results showed that Welsh proficiency, and the number of years spent in a Welsh speaking work or educational environment, predicted overall faster reading times, while working memory predicted overall slower reading times. There was no evidence to suggest that procedural memory, declarative memory, or language dominance played a role in this linguistic condition. Collectively, it appears that none of the cognitive and environmental factors investigated in this follow-up experiment played a positive role in the processing of gender in Welsh-English bilingual adults, in predicting stronger sensitivity to the agreement violations in any of the linguistic conditions.

These results were discussed in light of L2 / bilingual processing theories. It was concluded that the predictions made by the SSH are not considered suitable to explain the patterns emerging from the SPR task, and this is likely because it is intended for L2 adult learners and not highly proficient bilinguals who have learnt two langauges from a young age (0-11). It was also argued that the theoretical assumptions made by the DP model are able to (in part) appropriately describe the cognitive variable (declarative and procedural memory) results, while the assumptions made by the computational / cognitive capacity accounts are less suitable to describe the cognitive variable (working memory) results. Attention was also paid to environmental factors related to the speaker's use and experience of the Welsh language. Although no predictions are made specifically by the L2 processing theories for the involvement of such factors, these factors were discussed in light of previous research. Exploring these factors emphasised the need for research to consider how Welsh–English bilingual adults' lives influence their language proficiency in more detail, in order to establish which factors predict and influence higher performance on measures of morphology (Binks & Thomas, 2019). These factors are exploratory and could be investigated further, to understand the different variables that may play a role in the processing of gender in Welsh.

The findings from the two experiments have shown interesting patterns of results, notably, with high levels of variability in the data. There is no clear evidence suggesting that the Welsh bilinguals are displaying 'ceiling' effects in any of the comprehension, production, and processing tasks, nor is there strong evidence to suggest that the comprehension, production, or processing of gender is more robust when gender operates independently of mutations than when it is in conjunction with or encoded through mutations. Additionally, there is arguably only one individual difference variable that predicts higher performance on measures of comprehension, production, and processing, which is Welsh linguistic proficiency. However, there are several other considerations that are important when discussing these findings. Therefore, taken together, the findings from both experiments are now discussed in the broader field of bilingualism, to situate these findings within the wider bilingual context.

For some time, there existed an ideology for a bilingual speaker to be a perfectly balanced bilingual - two monolinguals in one, however, this notion is problematic (Grosjean, 2008). Given that the use of two or more languages is perhaps the most common characterization of language use in the world today, researchers are acknowledging that bilingualism is multidimensional and multifaced (Rothman et al., 2023a). Recent work is beginning to treat bilingualism as a continuum and therefore as a continuous variable in the range of analyses employed, with the hope to uncover the wide spectrum of experiences that give rise to bilingualism (Rothman et al., 2023a). Studies that are attempting to unpack the individual bilingual experiences showing that the relationships between factors, such as input (quality and quantity), use and proficiency, are complex, dynamic and non-linear, and this is likely due to the vast differences in opportunities and patterns of use in bilingual speakers (Rothman et al., 2023a). While this thesis may not have uncovered the key factors that contributed to Welsh bilingual intragroup variation, it has attempted to understand how a sub-set of individual differences and variables related to the Welsh bilinguals' experiences may play a role in the comprehension, production, and processing of Welsh gender. This is a step in the right direction, however, by no means does it provide a complete picture of the complexities and challenges faced by the Welsh-English bilingual speakers.

One of the most apparent explanations as to why the Welsh bilingual adults do not display 'ceiling' effects and in turn show variability in the results, is likely due to the complex and opaque nature of the Welsh gender (and mutation) system. In Welsh, gender assignment is generally arbitrary, and the noun form provides little information about noun gender. Gender is marked in the numerals 2, 3 and 4, which both have feminine and masculine forms. There are a small set of adjectives that have marked feminine forms that agree with the gender of the co-occurring noun (e.g., trwm [m] / trom [f] 'heavy') but these distinctions are being lost (Watkins, 1993). There are also some morphological generalisations that can be made to help infer the gender of the noun, but these are not absolute (Hammond, 2016; Thomas, 2001). Even though there are some generalisable trends, the irregularities and exceptions are so widespread that it would not be possible for a speaker to rely solely on such patterns whilst building the system. To add to this, nouns can differ in gender between regions and dialects across Wales, and even within Welsh dictionaries. These inconsistencies within the system makes gender in Welsh highly opaque and likely an arduous for new speakers of Welsh to acquire.

The gender system in Welsh is also intertwined with the mutation system which further adds to the complexity of the system. As noted throughout this thesis, gender is not marked on the determiner in Welsh, however 'y' ('the') triggers SM on following feminine nouns but does not trigger a mutation on masculine nouns (and this is only for nouns starting with letters participating in the mutation system). Here, SM marks feminine gender in local contexts, however SM marks masculine gender in non-local contexts via the possessive adjective ‘ei’. The nouns following ‘ei’ must agree in gender with the antecedent noun and this is marked by SM for masculine antecedent nouns or AM for feminine antecedent nouns. This, combined with the fact that the mutation system functions on its own without gender, results in a complex form-function mapping between the two highly interrelated phenomena. Therefore, (new) speakers of Welsh need to learn the mutation system and know when it functions with and without gender, as well as the various morphophonological forms and their respective conditioning environments (Cho & Slabakova, 2014; Lardiere, 2009). Although a Welsh speaker may not explicitly know the intricacy and complexity of the two systems and the challenges which it presents, the Welsh language is often referred to as a “difficult” or “hard” language to learn. Given that these descriptions are based on anecdotal experiences, to some extent, they do hold, given the fact that in order for Welsh speakers to have command of the grammatical gender system, they must also learn the mutation system (Thomas, 2001).

While it has not been focused on in this thesis, it has previously been suggested that there is a simplification of the mutation system underway (Thomas, 2001). Arguably, the mutation system is gradually being reduced from the original four-way alternation of the standard language (soft mutation, nasal mutation, aspirate mutation and basic form (i.e., no mutation)), to a binary alternation between the bare form and SM (Jones, 1998; Thomas, 2001). This is not a claim that can be extended to the findings in this thesis, given that mutations have largely been investigated in relation to the gender system. However, there was evidence showing the use of SM in place of AM in production, and the production of nouns in bare form in place of SM. It is possible that the Welsh gender system is showing change and this simplification may be due to the constant contact with English. Interestingly, there is current research showing change to a gender system in another gendered language. In a study by Rodina and Westergaard (2021), they investigated the ongoing change in the grammatical gender system of Norwegian. They noted that previous research had shown that the feminine form of the indefinite article is quickly disappearing from several dialects, which has led to claims that feminine gender is being lost from the language. Based on their findings, they concluded that there is evidence of change, showing that the indefinite article and prenominal possessives are affected, while other properties such as the definite suffix and the postnominal possessive are generally unaffected. Therefore, in light of previous research, when considering inter-participant variation in Welsh bilingual adult speakers, future work could pay closer attention to the possibility that this simplification is already underway and reflect on what this could mean for the gender system in Welsh.

This thesis has not focused on the potential cross-linguistic influence of English on Welsh, given that Welsh is the minority language in Wales, and it is in constant contact with the dominant societal language, English, however, it is worth considering at this point. In some areas of Wales, particularly in towns within the county of Gwynedd, Welsh is the dominant language in the community where over 70% of the population speak Welsh (Welsh Government, 2021). In contrast, there are large areas of Wales where less than 10% of the population speak Welsh (e.g., Newport Gwent) (Welsh Government, 2021). Even when Welsh is the dominant community language, English is still present in every aspect of life, for example, on the radio, television, and social media. Therefore, it may be that cross-linguistic influence from English plays a role in the development of Welsh bilinguals' grammar (Rothman et al., 2023b). This constant contact with English, a non-gendered language, has been argued to accelerate the convergence between Irish and English. Fhlannchadha and Hickey (2021) investigated the acquisition of grammatical gender in child speakers of Irish [+gender] and compared performance to proficient adult speakers of Irish. They found that the quality and quantity of input impacts the acquisition of gender in Irish, specifically, they noted that the high levels of variability in the adult speakers' language adds 'noise' to an already complex system. They argued that there is evidence of change in the Irish gender system, and this is because of the increasing contact and exposure to the majority language, English. In light of this, the authors suggested that future research should consider moving the 'goal-posts', in other words, what is considered the end point of successful acquisition, since there are rapid changes in Irish speaking adults' usage, particularly grammatical gender.

In light of previous research, given that English does not have grammatical gender but Welsh does, it may be that exploring the comprehension, production and processing of gender in Welsh bilingual speakers who have acquired Welsh from birth or a young age, and comparing the results to new speakers of Welsh who have started learning Welsh during adulthood, may inform the field of the precise parameters that bidirectional cross-linguistic influence has for bilingual development and ultimate attainment, particularly for minority language contexts (Rothman et al., 2023b). Furthermore, research could consider investigating the comprehension, production, and processing of gender in Spanish-Welsh bilingual speakers, who live in Patagonia. Following personal communication with Dr Rocio Perez-Tattam, they explained that production and comprehension data have been collected from L2 Welsh speakers whose L1 is Spanish and live in Patagonia, focusing on the use of gender in Welsh (in 2011). However, Dr Perez-Tattam explained that the results were inconclusive, with some evidence pointing towards no gender marking in this Welsh sepaking population. Therefore, future work could explore this further, as it is particularly interesting given that Spanish also possesses a binary grammatical gender system and in Patagonia, Spanish is the majority language. As a result, research could consider how contact with another gendered language impacts on the use of gender in speakers of Welsh.

Something that has not been discussed in this thesis, but is also worth noting here, is the possibility that some Welsh speakers may be recognised as *heritage* speakers of Welsh. For example, children may have Welsh at home, but Welsh is not the dominant language of the community (Rothman et al., 2023b). Even though this is likely the case for a large proportion of the speakers in Wales, Welsh is not claimed to be a heritage language in this thesis, given the complicated nature of the Welsh language and Welsh speakers in Wales. This thesis has attempted to avoid any grouping and/or classification of speakers, due to the complex and changing picture of Welsh in Wales. Recently, there has been a shift away from labelling speakers who have not learnt Welsh from a young age (i.e., teenager years+) as "second language learners" of Welsh, but instead, it is encouraged to refer to these speakers as "Siaradwyr Newydd" (new speakers). This shift has occurred in light of the call to be more inclusive of the speakers who are learning Welsh, particularly when some speakers have learnt Welsh in their adulthood and have spoken it for years. However, one important consideration that emerges from heritage speaker research, is the question of whether children are exposed to grammars that robustly instantiate the properties of gender assignment and/or gender agreement (Kupisch & Rothman, 2018; Rothman, 2007). For example, if children are exposed to the features day-to-day, and there is evidence to suggest that they acquire everything that they have been exposed to, but this input has high levels of variability, it is possible that acquisition may appear to be 'incomplete' or 'divergent' from the standard language agreement (Kupisch & Rothman, 2018; Rothman, 2007). It is known that variation exists between and within Welsh speakers and that under favourable conditions, such as mass exposure and optimal input levels, the acquisition process can be relatively unproblematic (e.g., Binks & Thomas, 2019; Gathercole et al., 2001; Gathercole & Thomas, 2005; Sharp, 2012). However, discussing the idea that Welsh may be recognised as a heritage language and what impact this has on the acquisition of Welsh gender is beyond the scope of this thesis. Future research could consider this in light of previous heritage speaker research.

While this thesis has focused on a sub-set of individual variables, it has uncovered more factors which deserve attention in future work. These include the degree of literacy in Welsh, as well as patterns of use at work and maybe even in social contexts. Attempting to develop a nuanced understanding of the individual differences and factors in this underrepresented bilingual context can inform the field of the variation within and across bilinguals (Fricke, Zirnstein, Navarro-Torres, & Kroll, 2019). Additionally, exploring key variables that could explain inter-participant variation will help to legitimize all language contexts and all individuals as being equally worthy of investigation (Rothman et al., 2023b). Historically, previous research has compared bilingual speakers' performance and patterns to L1 speakers of the target language under investigation. While this thesis has compared the patterns which have emerged from the Welsh bilingual data to L1, L2 and heritage speaker data, this thesis has not compared findings to adult L1 monolingual Welsh speakers, because these speakers do not exist. Nonetheless, given that the field recognises the monolingual comparative normativity in bilingualism research, the Welsh bilinguals in this thesis should be identified as bilingual speakers within their own right, in showing that the emerging patterns are their own, without being compared to L1 speakers of other gendered languages and treated as showing behavioural patterns that reflect "reduced", "incomplete" and/or "simplified" systems (Rothman et al., 2023b).

This thesis has investigated and discussed how grammatical gender in Welsh is comprehended, produced, and processed, in Welsh-English bilingual adults. It has provided a platform for highly proficient Welsh bilinguals from geographically diverse populations and of varying bilingual profiles. What is clear from the findings is that bilingualism is not black or white, and future work should continue to document the backgrounds and relevant experiences of the bilinguals, in order to gain a greater insight into the variables that contribute to bilingual intragroup variation (Rothman et al., 2023a; Rothman et al., 2023b). Unpacking these bilingual experiences and probing meaningfully into the different variables may highlight what, if any, the real-world world implications are (Rothman et al., 2023a).

7.2 Conclusions

This thesis set out to investigate the exploring the comprehension, production, and processing of Welsh grammatical gender in a more geographically diverse population of Welsh-English bilingual adult speakers, than those tested previously. It also set out to understand whether the variable use of gender in Welsh is due to adults' lack of ability and consistency with using gender and mutations together, rather than due to an issue with their gender knowledge. Finally, it also attempted to capture the key variables that could explain the patterns observed in the exploring the comprehension, production, and processing of Welsh grammatical gender.

To address these aims, Chapter 1 contextualised Welsh in Wales for the reader, while Chapter 2 introduced the linguistic domain of interest, grammatical gender, and its involvement with the mutation system. It also briefly addressed research on child language acquisition of gender cross-linguistically and specifically in Welsh. Chapter 3 reviewed the prior research on production and comprehension of grammatical gender in Welsh adults, before reviewing two theoretical approaches of ultimate attainment in L2 acquisition. After reviewing each theory, I presented the predictions for the Welsh adult bilinguals in the first experiment. Chapter 4 presented the first experiment of two which formed the heart of this thesis, investigating the comprehension and production of grammatical gender in Welsh-English adult bilinguals, as well as the role of various individual difference variables. Chapter 5 extended these findings from production and comprehension to processing. It also reviewed various theoretical approaches to the processing of grammatical gender in L2/bilingual speakers. Chapter 6 presented experiment two, exploring the processing of gender in a sub-group of the participants from experiment one and considered the role of cognitive and environmental individual difference variables. In this final chapter, I have brought together findings from the two experiments, chapters 4 and 6, and presented a general discussion, hoping to capture the true complexity of the nature bilingualism of Welsh alongside English in Wales. These results and discussion points should be interpreted with a word of caution and in the next section, I will outline some of the limitations of this research.

7.3 Limitations

There are limitations to this study that limit its generalizability. The limitations are discussed according to the two experiments. For experiment one, the first limitation acknowledged is the fact that gender agreement between the determiner ‘y’ and the noun was not explored. To mitigate this, a condition in the SPR task (experiment two) included this. However, future work could consider investigating the production and comprehension of gender when gender is encoded through mutations via ‘y’. One of the conditions in the elicited imitation task did test this, however, this task was removed due to some scoring issues. The second limitation is the lack of additional questions asked as part of the background questionnaire, the BLP, to develop a greater insight into the Welsh speakers’ contact with the Welsh language, in terms of literacy skills, including reading, speaking, and writing in Welsh. Future research should consider documenting detailed information to develop a better understanding of the various factors that may play a role in the comprehension, production, and processing of Welsh grammatical gender.

For experiment two, one of the main limitations is the fact that data were collected online, therefore it is likely that this affected the reading times in the SPR. The data were collected in Spring 2022, which was during the global COVID-19 pandemic, therefore, the experiment was administered online via the software Gorilla.sc. This meant that the participants completed the tasks on their own devices and on their chosen web browsers. Research has found reaction times in milliseconds to vary depending on which platforms (Win10 / MacOS) and browsers (Firefox, Edge, Chrome, Safari) are used when using Gorilla (Bridges, Pitiot, MacAskill, & Peirce, 2020). Therefore, the reading times reported in this thesis may vary less across individuals than expected because of these factors. Interestingly, differences in the use of devices has not been found to be a major problem when collecting online data compared to a lab setting (Mathôt & March, 2022), and recent studies have found no differences when comparing data obtained via the web and in the lab (Gastmann, Poarch, & Schimike, 2022). Nevertheless, it would be beneficial to replicate this SPR study in a lab setting, controlling for the device and web browser type to compare whether the results from this study can be replicated in the optimal data collection conditions, whether this be online or in a lab setting.

There is also a methodological consideration which should be noted, concerning the Tower of Hanoi task. Namely, the approach taken to score the TOH was innovative and experimental. It attempted to demonstrate procedural memory learning ability in the Welsh bilinguals. However, it is possible that the scoring system did not capture a direct reflection of procedural memory learning ability in the participants. These factors should be considered in future work exploring the role of procedural memory on outcome variables.

Another limitation to this dissertation concerns the number of participants in the two experiments. Experiment one collected data from 40 participants while experiment two collected data from 21 participants, who were a sub-group of the 40 participants. Arguably, these numbers are small and therefore make it difficult to pose any generalisable claims about the findings from the two experiments. However, it is worth noting that this relatively small but sizeable number was likely due to Welsh speaking participants motivation to be involved in online experiments during such challenging times (i.e., during the COVID-19 pandemic). I am extremely grateful for their time and contribution. However, larger numbers would have been better, particularly concerning the power in the statistical analyses. The participants were not recruited in a principled or representative way, and this is something future work could consider doing.

7.4 Directions for Future Research

Having discussed some of the limitations of this dissertation, some recommendations for further research are proposed. There are suggestions made throughout the thesis, however, the most noteworthy ones which can further add to the work carried out in this thesis, are outlined below.

This thesis has tested the gender and mutation systems separately from one another as well as together. However, future work should consider the different instantiations of gender and mutations together and independent of one another, in local and non-local contexts, in production, comprehension and online processing, in order to gain a more precise and robust understanding of the behavioural patterns observed in Welsh speakers' use of gender in Welsh. Future work might consider using more sophisticated psycholinguistic techniques to quantitatively measure processing in the Welsh-English bilingual adults, such as eye-tracking methods[20]. By doing so, it would provide a more fine-grained analysis of the data and help strengthen our knowledge of processing patterns in Welsh adult bilinguals.

Once a greater understanding is developed of these patterns in highly proficient Welsh adult bilinguals, the findings can inform Welsh child data in understanding how gender and mutations are acquired when they are tested separately from one another. To help develop this understanding, the experimental tasks in this thesis could be administered to Welsh speaking children and new speakers of Welsh who have started learning Welsh during adulthood. In doing so, it would provide a clearer picture of how gender is acquired, produced, comprehended, and processed in different bilingual Welsh speaking populations.

When the research community has a greater understanding of how gender in Welsh is acquired, produced, comprehended, and processed when independent and in conjunction with mutations, in local and non-local contexts, future work could consider using eye-tracking methodology to explore the predictive processing of grammatical gender in Welsh. In doing so, it would inform the field of whether Welsh adult bilingual speakers use gender cues to facilitate processing and establish potential anticipatory patterns in Welsh and compare these findings to bilingual speakers of other gendered languages.

Future work could also consider collecting more information from the Welsh speakers regarding their use of and experiences in Welsh. More specifically, speakers could be asked detailed questions regarding their literacy abilities, including the amount they speak, read, and write in Welsh, as well as the different environments in which Welsh is used and experienced by the speakers. This is suggested in light of the findings from experiment two, which indicated that the number of years spent in an educational/professional environment using Welsh influenced the processing patterns in the Welsh adult speakers. Exploring this further

[20] This is a suggestion for future research, as this was not possible in this thesis, due to restrictions during the global COVID-19 pandemic.

would inform the field of whether if and how contact with more formal or colloquial Welsh affects outcomes in measures of Welsh grammatical gender.

A final suggestion is that future work might consider comparing responses from production, comprehension, and processing tasks to the data available in the new Welsh Corpus – CorCenCC. It may be that the data in the corpus can show how gender is used in Welsh speakers outside of experimental settings, which can inform researchers of how Welsh gender and mutations operate in 'everyday' language. Additionally, this may lead researchers who are exploring the acquisition of gender in Welsh children to adjust the 'target-like' expectations one might anticipate the Welsh speaking children to aim towards.

7.5 Thesis summary

There are some noteworthy implications of this thesis. These are outlined below:

- It has researched an under-researched language (Welsh) in a minority bilingual context (Welsh alongside English).
- It reports on the first SPR task to investigate the processing of Welsh gender.
- The findings have been discussed in light of various L2 / bilingual theoretical approaches, with the hope that the predictions are suitable to describe the patterns emerging from the Welsh-English bilingual adults data.
- It situates the findings in light of literature exploring individual differences in bilinguals and considers bilingualism as multidimensional.
- It reports data from Welsh adult speakers who are from a more geographically diverse population than those tested previously, who also have diverse bilingual profiles.
- It has reported patterns of data in Welsh adult speakers, which researchers can consider what constitutes the 'end point' of successful acquisition, providing food for thought in contemplating whether there should be a sense of shifting the 'goal-posts'.

The work carried out in this thesis contributes to our understanding of how Welsh bilingual adults comprehend, produce and process gender in Welsh. It also contributes to the wider field of bilingualism, including whether it is possible to extend L2 / bilingual theoretical approaches to highly proficient Welsh-English bilingual adults who have learnt both languages before adulthood. Furthermore, it aids our understanding of how a sub-set of variables may play roles in the comprehension, production, and processing of Welsh gender. Given the minority-language bilingual context here in Wales, one might wonder whether these bilingual speakers are model subjects of study for those interested in the acquisition, use and/or processing of an underrepresented gendered language existing, and arguably thriving, alongside the dominant societal language, English.

Appendices

Appendix A. Bilingual Language Profile: BLP (English)

Section 1: Biographical Information

Background Questionnaire

Name:

Age (in years):

Date of birth:

Gender:

- Male
- Female
- Prefer not to say
- Other (please specify)

Do you have, or, ever had problems with your:

- Vision
- Hearing
- Learning
- Language (speaking)
- N/A
- Other (please specify)

Current place of residence (town/city):

Were you born in Wales?

- Yes
- No
- If 'No' please specify which country

Occupation:

Highest level of formal education:

- GCSE
- A-level
- Degree
- PGCE
- Professional training
- Masters
- PhD
- Other

Do you teach Welsh?

Please select...

Please list all the languages you know in order of **dominance**:

Please list all the languages you know in order of **acquisition** (your native language first):

Which language do you speak the **most**?

Section 2: Language History

Language History

1. At what age did you start learning **Welsh**? (E.g. since birth, 1-20+ years)

2. At what age did you start learning **English**?

3. At what age did you start to feel comfortable using **Welsh**? (E.g. as early as I can remember, 1-20+ years, not yet)

4. At what age did you start to feel comfortable using **English**?

5. How many years of classes (language, history, math, etc) have you had in **Welsh** (primary school through to university)? (Since birth, 1-20+ years)

6. How many years of classes (language, history, math, etc) have you had in **English** (primary school through to university)?

7. How many years have you spent in a country/region where **Welsh** is spoken? (Since birth, 1-20+ years)

8. How many years have you spent in a country/region where **English** is spoken?

9. How many years have you spent in a family where **Welsh** is spoken? (Since birth, 1-20+ years)

10. How many years have you spent in a family where **English** is spoken?

11. How many years have you spent in a work environment where **Welsh** is spoken? (Since birth, 1-20+ years)

12. How many years have you spent in a work environment where **English** is spoken?

Section 3: Language Use

Language use

Please answer with a percentage between 0 % and 100%. (Total use for all languages in a given question should equal 100%).

13. In an average week, what percentage of the time do you use **Welsh** with *friends*?

14. In an average week, what percentage of the time do you use **English** with *friends*?

15. In an average week, what percentage of the time do you use **Welsh** with *family*?

16. In an average week, what percentage of the time do you use **English** with *family*?

17. In an average week, what percentage of the time do you use **Welsh** at *university/work*?

18. In an average week, what percentage of the time do you use **English** at *university/work*?

19. When you talk to yourself, how often do you *talk to yourself* in **Welsh**?

20. When you talk to yourself, how often do you *talk to yourself* in **English**?

21. When you count, how often do you *count* in **Welsh**?

22. When you count, how often do you *count* in **English**?

23. In an average week, what percentage of the time do you spend watching *tv* in **Welsh**?

24. In an average week, what percentage of the time do you spend watching *tv* in **English**?

25. In an average week, what percentage of the time do you spend *reading* in **Welsh**?

26. In an average week, what percentage of the time do you spend *reading* in **English**?

27. In an average week, what percentage of the time do you spend *listening to the radio* in **Welsh**?

28. In an average week, what percentage of the time do you spend *listening to the radio* in **English**?

29. In an average week, what percentage of the time do you spend *listening to music* (e.g. on Spotify) in **Welsh**?

30. In an average week, what percentage of the time do you spend *listening to music* (e.g. on Spotify) in **English**?

31. In an average week, what percentage of the time do you spend *writing* in **Welsh**?

32. In an average week, what percentage of the time do you spend *writing* in **English**?

33. In an average week, what percentage of the time do you spend on *social media* (e.g. Instagram, Facebook, Twitter etc) in **Welsh**?

34. In an average week, what percentage of the time do you spend on *social media* (e.g. Instagram, Facebook, Twitter etc) in **English**?

Section 4: Language Proficiency

Language proficiency

In this section, we would like you to rate your language proficiency by giving marks from 0 to 6 (0=not very well, 6=very well).

35. How well do you *speak* **Welsh**?

0	1	2	3	4	5	6

36. How well do you *speak* **English**?

0	1	2	3	4	5	6

37. How well do you *understand* **Welsh**?

0	1	2	3	4	5	6

38. How well do you *understand* **English**?

0	1	2	3	4	5	6

39. How well do you *read* in **Welsh**?

0	1	2	3	4	5	6

40. How well do you *read* in **English**?

0	1	2	3	4	5	6

41. How well do you *write* in **Welsh**?

0	1	2	3	4	5	6

42. How well do you *write* in **English**?

0	1	2	3	4	5	6

Section 5: Language Attitudes

Language attitudes

In this section, we would like you to respond to statements about language attitudes by giving marks from 0-6 (0=disagree, 6=agree).

43. I feel like myself when I *speak* **Welsh**.

0	1	2	3	4	5	6

44. I feel like myself when I *speak* **English**.

0	1	2	3	4	5	6

45. I *identify* with a **Welsh-speaking** culture.

0	1	2	3	4	5	6

46. I *identify* with a **English-speaking** culture.

0	1	2	3	4	5	6

47. It is important to me to *use* (or eventually use) **Welsh** like a native speaker.

0	1	2	3	4	5	6

48. It is important to me to *use* (or eventually use) **English** like a native speaker.

0	1	2	3	4	5	6

49. I want others to think I am a native speaker of **Welsh**.

0	1	2	3	4	5	6

50. I want others to think I am a native speaker of **English**.

0	1	2	3	4	5	6

BLP (Welsh)

Adran 1: Bwybodaeth Bywgraffyddol

Holiadur Cefndir

Enw:

Oedran (mewn blynyddoedd):

Dyddiad geni:

Rhyw:

- Gwryw
- Menyw
- Well gennyf beidio â ddweud
- Arall

A ydych chi, neu, erioed wedi cael problemau gyda'ch:

- Golwg
- Clyw
- Dysgu
- Iaith (siarad)
- N/A (dim)
- Arall

Ble rydych chi'n byw (tref/dinas)?

Gawsoch chi'ch geni yng Nghymru?

- Ie
- Na
- Os 'Na', nodwch pa wlad

Swydd bresennol:

Lefel uchaf o addysg ffurfiol:

- TGAU
- Lefel-A
- Gradd
- TAR
- Hyfforddiant profesiynol
- Meistri
- Doethuriaeth
- Arall

Ydych chi'n addysgu'r Gymraeg

Dewiswch...

Rhestrwch yr holl ieithoedd rydych chi'n eu hadnabod yn nhrefn **goruchafiaeth iaith**:

Rhestrwch yr holl ieithoedd rydych chi'n eu hadnabod yn nhrefn **caffaeliad** (eich iaith frodorol yn gyntaf):

Pa iaith ydych chi'n siarad y **mwyaf**?

Adran 2: Hanes Iaith

Hanes iaith

1. Pryd dechreuoch chi ddysgu'r **Cymraeg**? (E.e. ers i mi gael fy ngeni, 1-20+ mlynedd)

2. Pryd dechreuoch chi ddysgu'r **Saesneg**?

3. Pryd dechreuoch chi deimlo'n gartrefol yn siarad **Gymraeg**? (E.e. cyn gynted ag y gallaf ei gofio, 1-20+ mlynedd, dim eto)

4. Pryd dechreuoch chi deimlo'n gartrefol yn siarad **Saesneg**?

5. Faint o flynyddoedd o ddosbarthiadau (iaith, hanes, mathemateg, ac ati) a gawsoch yn **Gymraeg** (ysgol gynradd hyd at y brifysgol)? (Ers i mi gael fy ngeni, 1-20+ mlynedd)

6. Faint o flynyddoedd o ddosbarthiadau (iaith, hanes, mathemateg, ac ati) a gawsoch yn **Saesneg** (ysgol gynradd hyd at y brifysgol)?

7. Faint o flynyddoedd ydych wedi'u treulio mewn gwlad/rhanbarth lle siaredir **Cymraeg**? (Ers i mi gael fy ngeni, 1-20+ mlynedd)

8. Faint o flynyddoedd ydych wedi'u treulio mewn gwlad/rhanbarth lle siaredir **Saesneg**?

9. Faint o flynyddoedd ydych wedi'u treulio mewn teulu lle siaredir **Cymraeg**? (Ers i mi gael fy ngeni, 1-20+ mlynedd)

10. Faint o flynyddoedd ydych wedi'u treulio mewn teulu lle siaredir **Saesneg**?

11. Faint o flynyddoedd ydych wedi'u treulio mewn amgylchedd gwaith lle siaredir **Cymraeg**? (1-20+ mlynedd)

12. Faint o flynyddoedd ydych wedi'u treulio mewn amgylchedd gwaith lle siaredir **Saesneg**?

Adran 3: Defnydd Iaith

Defnydd iaith

Atebwch gyda chanran rhwng 0% a 100%. (Dylai'r cyfanswm am y ddwy iaith mewn cwestiwn penodol fod yn gyfartal i 100%).

13. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n defnyddio **Cymraeg** gyda *ffrindiau*?

14. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n defnyddio **Saesneg** gyda *ffrindiau*?

15. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n defnyddio **Cymraeg** gyda *teulu*?

16. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n defnyddio **Saesneg** gyda *teulu*?

17. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n defnyddio **Cymraeg** yn y *brifysgol/gwaith*?

18. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n defnyddio **Saesneg** yn y *brifysgol/gwaith*?

19. Wrth i chi siarad â chi eich hun, pa mor aml fyddwch chi'n *siarad â chi eich hun* yn **Cymraeg**?

20. Wrth i chi siarad â chi eich hun, pa mor aml fyddwch chi'n *siarad â chi eich hun* yn **Saesneg**?

21. Wrth i chi *gyfri*, pa mor aml fyddwch yn *cyfri* yn **Cymraeg**?

22. Wrth i chi *gyfri*, pa mor aml fyddwch yn *cyfri* yn **Saesneg**?

23. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n gwylio'r *teledu* yn **Cymraeg**?

24. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n gwylio'r *teledu* yn **Saesneg**?

25. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *darllen* yn **Cymraeg**?

26. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *darllen* yn **Saesneg**?

27. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *gwrando ar y radio* yn **Cymraeg**?

28. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *gwrando ar y radio* yn **Saesneg**?

29. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *gwrando ar gerddoriaeth* (e.e. Spotify) yn **Cymraeg**?

30. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *gwrando ar gerddoriaeth* (e.e. Spotify) yn **Saesneg**?

31. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *ysgrifennu* yn **Cymraeg**?

32. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n *ysgrifennu* yn **Saesneg**?

33. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n ei dreulio ar *gyfryngau cymdeithasol* (e.e. Instagram, Facebook, Twitter a.y.b) yn **Cymraeg**?

34. Mewn wythnos gyffredin, pa ganran o'r amser ydych chi'n ei dreulio ar *gyfryngau cymdeithasol* (e.e. Instagram, Facebook, Twitter a.y.b) yn **Saesneg**?

Adran 4: Hyfedredd Iaith

Hyfedredd iaith

Yn yr adran hon, hoffem i chi roi sgôr i'ch hyfedredd iaith gan roi marciau o 0 i 6 (0=ddim yn dda o gwbl, 6=da iawn).

35. Pa mor dda ydych chi'n *siarad* **Cymraeg**?

0	1	2	3	4	5	6

36. Pa mor dda ydych chi'n *siarad* **Saesneg**?

0	1	2	3	4	5	6

37. Pa mor dda ydych chi'n *ddeall* **Cymraeg**?

0	1	2	3	4	5	6

38. Pa mor dda ydych chi'n *ddeall* **Saesneg**?

0	1	2	3	4	5	6

39. Pa mor dda ydych chi'n *darllen* **Cymraeg**?

0	1	2	3	4	5	6

40. Pa mor dda ydych chi'n *darllen* **Saesneg**?

0	1	2	3	4	5	6

41. Pa mor dda ydych chi'n *ysgrifennu* **Cymraeg**?

0	1	2	3	4	5	6

42. Pa mor dda ydych chi'n *ysgrifennu* **Saesneg**?

0	1	2	3	4	5	6

Adran 5: Agweddau Iaith

Agweddau iaith

Yn yr adran hon, hoffem i chi ymateb I osodiadau am agweddau iaith trwy roi marciau o 0 i 6 (0=anghytuno, 6=cytuno).

43. Rwy'n teimlo fel fi fy hun pan fyddaf yn *siarad* **Cymraeg**.

0	1	2	3	4	5	6

44. Rwy'n teimlo fel fi fy hun pan fyddaf yn *siarad* **Saesneg**.

0	1	2	3	4	5	6

45. Rwy'n *uniaethu* â diwylliant **Cymraeg** ei iaith.

0	1	2	3	4	5	6

46. Rwy'n *uniaethu* â diwylliant **Saesneg** ei iaith.

0	1	2	3	4	5	6

47. Mae'n bwysig i mi i *ddefnyddio* (neu ddefnyddio yn y pen draw) yr iaith **Cymraeg** fel siaradwr brodorol.

0	1	2	3	4	5	6

48. Mae'n bwysig i mi i *ddefnyddio* (neu ddefnyddio yn y pen draw) yr iaith **Saesneg** fel siaradwr brodorol.

0	1	2	3	4	5	6

49. Rwy' am i bobl eraill feddwl fy mod i'n siaradwr brodorol o'r **Gymraeg**.

0	1	2	3	4	5	6

50. Rwy' am i bobl eraill feddwl fy mod i'n siaradwr brodorol o'r **Saesneg**.

0	1	2	3	4	5	6

Additional BLP Questions

These additional questions were asked bilingually to the participants, but the questions are written in English only below and did not contribute to the dominance score.

Language *Use* section

- In an average week, what percentage of the time do you spend watching *tv* in Welsh / English?
- In an average week, what percentage of the time do you spend *reading* in Welsh / English?
- In an average week, what percentage of the time do you spend *listening to the radio* in Welsh / English?
- In an average week, what percentage of the time do you spend *listening to music* (e.g., on Spotify) in **Welsh** / English?
- In an average week, what percentage of the time do you spend on *social media* (e.g., Instagram, Facebook, Twitter etc) in **Welsh** / English?

Appendix B. Comprehension of Grammatical Gender Task

Target sentences in order of presentation

Item	Target noun	Gender	Animacy	Pronoun	Sentence 1	Sentence 2
1	Gwely *Bed*	M	Inanimate	Ei+SM	Dyma'r gyllell ddu a dyma'r gwely coch.	Ond mae'i goes wedi plygu.
2	Plismon *Policeman*	M	Human	Ei+SM	Dyma'r plismon dwl a dyma'r dywysoges dal.	Mae esgid oren ar ei droed.
3	Pêl *Ball*	F	Inanimate	Hi	Roedd y drwm tenau a'r bêl frown mewn bocs.	Ond cwympodd hi drwy'r gwaelod.
4	Tywysog *Prince*	M	Human	Ei+SM	Dyma'r fenyw ddwl a dyma'r tywysog da.	Mae het ar ei ben.
5	Desg *Desk*	F	Inanimate	Ei+AM	Dyma'r bwrdd brown a dyma'r ddesg goch.	Ond mae'i choes wedi torri.
6	Dafad *Sheep*	F	Animal	Hi	Aeth y ddafad dwp a'r ci tew am dro i'r cae.	Ond aeth hi’n sownd mewn perth!
7	Cath *Cat*	F	Animal	Ei+AM	Dyma'r gath dew a dyma'r draenog pigog.	Ond mae pili-pala ar ei thrwyn.
8	Gwasgod *Waistcoat*	F	Inanimate	Ei+AM	Dyma'r crys pinc a dyma'r wasgod ddu.	Ond mae'i phoced yn las.
9	Tap *Tap*	M	Inanimate	(F)E	Roedd y fasged goch a'r tap glas ar y ffenest.	Ond gwnaeth e dorri.
10	Blaidd *Wolf*	M	Animal	Ei+SM	Dyma'r afr frown a dyma'r blaidd glas.	Ond mae 'sanau ar ei glust.
11	Cwcw *Cuckoo*	F	Animal	Ei+AM	Dyma'r gwdihŵ porffor a dyma'r gwcw lwyd.	Ond mae hosan ar ei chlust.
12	Cyllell *Knife*	F	Inanimate	Hi	Dyma’r gyllell ddu a dyma’r plât coch.	Ond mae hi ar y bwrdd.
13	Crys *Shirt*	M	Inanimate	(F)E	Roedd y crys gwyrdd a'r wasgod binc yn y cwpwrdd.	Ond oedd e’n hen
14	Twrci *Turkey*	M	Animal		Dyma’r bioden ddu a	Mae het ar ei ben.

					dyma'r twrci brown.	
15	Pili-pala *Butterfly*	M	Animal	Ei+SM	Dyma'r pili-pala coch a dyma'r fuwch las.	Ond mae rhywbeth ar ei drwyn.
16	Prifathrawes *Headmistress*	F	Human	Ei+AM	Dyma'r brifathrawes grac a dyma'r brenin drwg.	Mae llyfr wedi cwympo ar ei throed.
17	Gafr *Goat*	F	Animal	Hi	Roedd y blaidd du a'r afr binc yn byw ar fferm.	Ond fel arfer, roedd hi mas yn y cae.
18	Twrci *Turkey*	M	Animal	(F)E	Roedd y fuwch ddu a'r twrci brown mewn cae.	Ond rhedodd e mas drwy'r gât.
19	Tegell *Kettle*	M	Inanimate	Ei+SM	Dyma'r bêl borffor a dyma'r tegell du.	Ond mae rhywbeth od iawn ar ei ben.
20	Menyw *Woman*	F	Human	Ei+AM	Dyma'r gŵr pwysig a dyma'r fenyw dda.	Mae het ar ei phen.
21	Pengwin *Penguin*	M	Animal	(F)E	Aeth y pengwin cysglyd a'r gath flinedig i'w gwelyau.	Aeth e i gysgu'n gyntaf.
22	Dyn *Man*	M	Human	Ei+SM	Dyma'r gyfnither dda a dyma'r dyn tal.	Mae'i got yn binc.
23	Llygoden *Mouse*	F	Animal	Hi	Penderfynodd y llygoden fach a'r pili-pala gwyn gael ras.	Y tro hwn, enillodd hi.
24	Drych *Mirror*	M	Inanimate	Ei+SM	Dyma'r drych tenau a dyma'r gadair las.	Ond mae llythyr ar ei ben.
25	Telyn *Harp*	F	Inanimate	Hi	Roedd y delyn denau a'r gitâr du yn yr ystafell.	Ond does dim llinynnau arni hi.
26	Brenhines *Queen*	F	Human	Ei+AM	Dyma'r cefnder crac a dyma'r frenhines bwysig.	Mae'i chot yn goch.
27	Draenog *Hedgehog*	M	Animal	(F)E	Dyma’r draenog pigog a'r bioden ddu.	Ond ddisgynodd e.

28	Torth *Loaf*	F	Inanimate	Ei+AM	Dyma'r dorth dew a dyma'r llyfr glas.	Ond mae cyllell ar ei phen.
29	Dafad *Sheep*	F	Animal	Ei+AM	Dyma'r ci du a dyma'r ddafad binc.	Mae cleren bach ar ei phen.
30	Blodyn *Flower*	M	Inanimate	(F)E	Roedd y blodyn piws a'r ddysgl las ar y bwrdd.	Ond torrodd e.
31	Gwely *Bed*	M	Inanimate	Ei+SM	Dyma’r gwely coch and dyma’r ddesg goch.	Ond mae’i goes wedi torri.
32	Basged *Basket*	F	Inanimate	Hi	Dyma’r bwrdd brown a’r fasged goch.	Ond torrodd hi.
33	Coeden *Tree*	F	Inanimate	Ei+AM	Dyma’r goeden werdd a dyma'r blodyn pinc.	Ond mae aderyn ar ei phen.
34	Potel *Bottle*	F	Inanimate	Hi	Dyma’r botel las a dyma'r plât gwyrdd. Ond mae hi ar y bwrdd.	Ond mae hi ar y bwrdd.
35	Tegell *Kettle*	M	Inanimate	(F)E	Roedd y tegell du a’r dorth dew ar y ffwrn.	Ond llosgodd e.
36	Drych *Mirror*	M	Inanimate	(F)E	Dyma'r delyn frown a dyma'r drych glas.	Ond cwympodd e.

Target nouns in original task (Binks & Thomas, 2019)

Original TN (Welsh)	**Original TN (English)**	**Gender**	**Animacy**	**Reason for change**
Gwely	Bed	Masculine	Inanimate	-
Plismon	Policeman	Masculine	Human	-
Pêl	Ball	Feminine	Inanimate	-
Tywysog	Prince	Masculine	Human	-
Desg	Desk	Feminine	Inanimate	-
Dafad	Sheep	Feminine	Animal	-
Cath	Cat	Feminine	Animal	-
Gwasgod	Wasistcoat	Feminine	Inanimate	-
Tap	Tap	Masculine	Inanimate	-
Blaidd	Wolf	Masculine	Animal	-
Tylluan	Owl	Feminine	Animal	*Tylluan* is NW form. *Gwdihŵ* is the form in SW but is masculine. Needed Fem, Animal noun ('cwcw' selected)
Cyllell	Knife	Feminine	Inanimate	-
Crys	Shirt	Masculine	Inanimate	-
Twrci	Turkey	Masculine	Animal	-
Glöyn byw	Butterfly	Masculine	Animal	*Glöyn byw* is NW form so changed to *pili-pala* which is SW form
Prifathrawes	Headmistress	Feminine	Human	-
Gafr	Goat	Feminine	Animal	-
Twrci	Turkey	Masculine	Animal	-
Tegell	Kettle	Masculine	Inanimate	-
Dynes	Woman	Feminine	Human	*Dynes* is is NW form. *Menyw* is SW form.
Pengwin	Penguin	Masculine	Animal	-
Dyn	Man	Masculine	Human	-
Tylluan	Owl	Feminine	Animal	Swapped *Tylluan* [f] for *Llygoden* [f] (mouse), as Tylluan is NW form and *Llygoden* is a more common noun
Drych	Mirror	Masculine	Inanimate	-
Telyn	Harp	Feminine	Inanimate	-
Brenhines	Queen	Feminine	Human	-
Draenog	Hedgehog	Masculine	Animal	-
Torth	Loaf	Feminine	Inanimate	-
Dafad	Sheep	Feminine	Animal	-
Blodyn	Flower	Masculine	Inanimate	-

Appendix C. Elicited Production Task

Target nouns in order of presentation

Target Noun (Welsh)	Target Noun (English)	Gender	Animacy
Mellten	Lightning	Feminine	Object
Mochyn	Pig	Masculine	Animal
Llyfr	Book	Masculine	Object
Tad-cu	Grandad	Masculine	Human
Broga	Frog	Masculine	Animal
Dafad	Sheep	Feminine	Animal
Mam	Mum	Feminine	Human
Troed	Foot	Feminine	Object
Bwrdd	Table	Masculine	Object
Cath	Cat	Feminine	Animal
Malwoden	Snail	Feminine	Animal
Menyw	Woman	Feminine	Human
Cacen	Cake	Feminine	Object
Drych	Mirror	Masculine	Object
Braich	Arm	Feminine	Object
Llygoden	Mouse	Feminine	Animal
Pluen	Feather	Feminine	Object
Pêl-droediwr	Footballer	Masculine	Human
Canwr	Singer	Masculine	Human
Planhigyn	Plant	Masculine	Object
Cylch	Circle	Masculine	Object
Buwch	Cow	Feminine	Animal
Corryn	Spides	Masculine	Animal
Cwningen	Rabbit	Feminine	Animal
Ci	Dog	Masculine	Animal
Cadair	Chair	Feminine	Object
Diod	Drink	Feminine	Object
Ceffyl	Horse	Masculine	Animal
Beic	Bike	Masculine	Object
Pysgodyn	Fish	Masculine	Animal
Tywysoges	Princess	Feminine	Human
Polyn	Pole	Masculine	Object
Tŷ	House	Masculine	Object
Peintiwr	Painter	Masculine	Human
Cneuen	Nut	Feminine	Object
Cloch	Bell	Feminine	Object
Coeden	Tree	Feminine	Object
Merch	Girl	Feminine	Human
Deilen	Leaf	Feminine	Object
Brenhines	Queen	Feminine	Human
Cwmwl	Cloud	Masculine	Object
Meddyg	Doctor	Masculine	Human
Brenin	King	Masculine	Human
Tegell	Kettle	Masculine	Object
Plât	Plate	Masculine	Object
Gwrach	Witch	Feminine	Human
Blodyn	Flower	Masculine	Object
Gardd	Garden	Feminine	Object

Appendix D. Productive Vocabulary Task

Target nouns in order of presentation

Item	Target Noun (Welsh)	Target Noun (English)	Lexically acceptable answers
1	Gwely	Bed	
2	Ci	Dog	
3	Tad-cu	Grandad	Grandfather
4	Cyllell	Knife	
5	Pêl	Ball	
6	Dafad	Sheep	
7	Menyw	Woman	Female
8	Bwrdd	Table	
9	Crys	Shirt	
10	Draenog	Hedgehog	
11	Pêl-droediwr	Footballer	Football player
12	Desg	Desk	
13	Gwasgod	Waistcoat	Jacket
14	Cath	Cat	
15	Tywysoges	Princess	
16	Tap	Tap	
17	Plât	Plate	
18	Blaidd	Wolf	
19	Canwr	Singer	
20	Diod	Drink	
21	Cloch	Bell	
22	Gafr	Goat	
23	Prifathrawes	Headmistress	Female headteacher
24	Tegell	Kettle	
25	Drych	Mirror	
26	Peintiwr	Painter	
27	Cacen	Cake	
28	Troed	Foot	
29	Brenhines	Queen	
30	Blodyn	Flower	
31	Twrci	Turkey	
32	Meddyg	Doctor	
33	Pluen	Feather	
34	Deilen	Leaf	
35	Buwch	Cow	
36	Mam	Mum	Mother
37	Llyfr	Book	
38	Polyn	Pole	
39	Pengwin	Penguin	
40	Plismon	Policeman	Police officer
41	Cneuen	Nut	
42	Gardd	Garden	
43	Cwcw	Cuckoo	
44	Gwrach	Witch	
45	Cwmwl	Cloud	
46	Planhigyn	Plant	
47	Pili-pala	Butterfly	
48	Tywysog	Prince	

49	Cadair	Chair	
50	Telyn	Harp	
51	Llygoden	Mouse	
52	Merch	Girl	Daughter
53	Tŷ	House	
54	Beic	Bike	
55	Pysgodyn	Fish	
56	Coeden	Tree	
57	Torth	Loaf	
58	Malwoden	Snail	
59	Cylch	Circle	
60	Corryn	Spider	
61	Ceffyl	Horse	
62	Brenin	King	
63	Basged	Basket	
64	Potel	Bottle	
65	Cwningen	Rabbit	
66	Mochyn	Pig	
67	Dyn	Man	
68	Mellten	Lightning	
69	Broga	Frog	Toad
70	Braich	Arm	

Appendix E. Cloze Tests

Welsh cloze test

1st passage includes all words

2nd passage includes the gaps

Passage 1:

Wrth ffilmio'r ail gyfres o *Blue Planet* ar BBC1, dywedodd cynhyrchwyr y rhaglen mai prin oedd yr adegau pan nad oedden nhw'n dod ar draws plastig yn y môr.

"Wrth ddeifio yn y môr wrth ffilmio, roedd darnau o blastig fel lein bysgota, papurau losin a photeli plastig yn olygfa gyffredin iawn," meddai Sarah Conner, un o'r cynhyrchwyr."

"Pan oedden ni'n teithio ar y cwch, os oedden ni'n sylwi ar ddarn o blastig, roedden ni'n stopio ac yn gwneud ein gorau i'w godi, fel y byddai unrhyw un sy'n poeni am ddyfodol ein planed. Ond weithiau roedd ynysoedd o blastig wedi crynhoi fel ei bod yn amhosibl cael gwared ohono i gyd."

Mae plastig yn achosi problem i nifer o greaduriaid y môr a phrofwyd hynny yn un o'r penodau pan welwyd crwban gwalchbig yn mynd yn sownd mewn sach blastig. Yn ffodus, y tro hwn, llwyddodd un o'r dynion camera i ryddhau'r crwban yn ddianaf. Ond nid pob creadur sydd mor lwcus. Yn ddiweddar, ymddangosodd fideo ar *YouTube* yn dangos crwban mewn poen difrifol gan fod gwelltyn yfed plastig wedi mynd yn sownd yn un o'i ffroenau. Ar ôl cryn drafferth, a'r creadur yn gwingo wrth i'r ecolegwyr geisio'i achub, llwyddwyd i dynnu'r gwelltyn.

Yn ystod y cyfnod ffilmio, cofnododd criw *Blue Planet* ymhle roedden nhw'n dod o hyd i sbwriel a faint ohono oedd yno. Roedden nhw'n synnu mai ar lannau ynysoedd pellennig, heb lawer o bobl yn byw arnyn nhw roedd y llygredd plastig mwyaf.

Weithiau roedd yn rhaid i'r criw ymyrryd pan oedden nhw'n gweld y plastig yn niweidio bywyd gwyllt. "Gwelson ni adar wedi marw oherwydd bod eu coesau wedi mynd yn sownd mewn bagiau plastig ac wedi methu hedfan. Bu'n rhaid i ni achub morfil cefngrwm oddi ar arfordir Canada gan ei fod yn llusgo tunelli o raffau a chewyll pysgota ac yn stryffaglu i nofio. Arhoson ni gyda'r morfil am naw awr i'w ryddhau. Y peth anoddaf oedd gorfod gwrando ar sŵn y morfil a oedd yn amlwg mewn poen dychrynllyd ac yn ofnus iawn."

Gellir gweld y bennod nesaf o *Blue Planet* nos Sul am 9 o'r gloch ar BBC1.

Passage 2:

Wrth ffilmio'r ail gyfres o *Blue Planet* ar BBC1, dywedodd cynhyrchwyr y rhaglen mai prin oedd yr adegau pan nad oedden nhw'n dod ar draws plastig yn y môr.

"-1- ddeifio yn y môr wrth ffilmio, -2- darnau o blastig fel lein bysgota, -3- losin a photeli plastig yn olygfa gyffredin -4-," meddai Sarah Conner, un o'r -5-."

"Pan oedden ni'n teithio ar -6- cwch, os oedden ni'n sylwi ar ddarn -7- blastig, roedden ni'n stopio ac yn -8- ein gorau i'w godi, fel y -9- unrhyw un sy'n poeni am ddyfodol -10- planed. Ond weithiau roedd ynysoedd o -11- wedi crynhoi fel ei bod yn -12- cael gwared ohono i gyd."

Mae plastig -13- achosi problem i nifer o -14- y môr a phrofwyd hynny yn -15- o'r penodau pan welwyd crwban gwalchbig -16- mynd yn sownd mewn sach blastig. -17- ffodus, y tro hwn, llwyddodd un -18- dynion camera i ryddhau'r crwban yn -19-. Ond nid pob creadur sydd mor -20-. Yn ddiweddar, ymddangosodd fideo ar *YouTube* -21- dangos crwban mewn poen difrifol gan -22- gwelltyn yfed plastig wedi mynd yn -23- yn un o'i ffroenau. Ar ôl -24- drafferth, a'r creadur yn gwingo wrth -25- ecolegwyr geisio'i achub, llwyddwyd i dynnu'r -26-.

Yn ystod y cyfnod ffilmio, cofnododd -27- *Blue Planet* ymhle roedden nhw'n dod -28- hyd i sbwriel a faint ohono -29- yno. Roedden nhw'n synnu mai -30- lannau ynysoedd pellennig, heb lawer o bobl -31- byw arnyn nhw roedd y llygredd -32- mwyaf.

Weithiau roedd yn rhaid -33- criw ymyrryd pan oedden nhw'n gweld y -34- yn niweidio bywyd gwyllt. "Gwelson ni -35- wedi marw oherwydd bod eu coesau -36- mynd yn sownd mewn bagiau plastig -37- wedi methu hedfan. Bu'n rhaid i -38- achub morfil cefngrwm oddi ar arfordir Canada -39- ei fod yn llusgo tunelli o -40- a chewyll pysgota ac yn -41- i nofio. Arhoson ni gyda'r morfil -42- naw awr i'w ryddhau. Y peth -43- oedd gorfod gwrando ar sŵn y morfil -44- oedd yn amlwg mewn poen dychrynllyd ac yn ofnus iawn."

Gellir gweld y bennod nesaf o *Blue Planet* nos Sul am 9 o'r gloch ar BBC1.

Bank of exact and lexically acceptable answers in order of presentation

Number	**Exact answer**	**Lexically acceptable answers**
1	Wrth	
2	roedd	mae
3	papurau	
4	iawn	
5	cynhyrchwyr	tîm, deifiwr, merch gamera
6	y	
7	o	
8	gwneud	
9	byddai	
10	ein	
11	blastig	sbwriel, fes,
12	amhosibl	anodd,
13	yn	
14	greaduriaid	Anifeiliaid, bywydau, pysgod
15	un	
16	yn	
17	Yn	
18	o’r	
19	ddianaf	diddifrod, di-glwyf, annifroddedig,
20	lwcus	ffodus, llwyddiannus
21	yn	
22	fod	
23	sownd	styc
24	cryn	tipyn, llawer
25	i’r	
26	gwelltyn	planhigyn
27	criw	tîm
28	o	
29	oedd	
30	ar	
31	yn	
32	plastig	sbwriel
33	i’r	
34	plastig	sbwriel
35	adar	
36	wedi	
37	ac	
38	ni	
39	gan	
40	raffau	wifrau, llinynnau, cortynnau, cordau
41	stryffaglu	
42	am	
43	anoddaf	
44	a	

English cloze test

1st passage includes all words

2nd passage includes the gaps

Passage 1:

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. ***His*** ever-restless brain has used the ***knowledge*** and the wisdom of his ancestors ***to*** improve his way of life. Since ***man*** is able to walk and run ***on*** his feet, his hands have always ***been*** free to carry and to use ***tools***. Man's hands have served him well ***during*** his life on earth. His development, ***which*** can be divided into three major ***periods***, is marked by several different ways **of** life.

Up to 10,000 years ago, ***all*** human beings lived by hunting and ***fishing***. They also picked berries and fruits, ***and*** dug for various edible roots. Most ***often,*** the men were the hunters, and ***the*** women acted as food gatherers. Since ***the*** women were busy with the children, ***the*** men handled the tools. In a ***man's*** hand, a dead branch became a ***tool*** to knock down fruit or to ***dig*** for tasty roots. Sometimes, an animal ***bone*** served as a club, and a ***sharp*** piece of stone, fitting comfortably into ***the*** hand, could be used to break ***nuts*** or to throw at an animal. ***One*** stone was chipped against another until ***it*** had a sharp edge. The primitive ***man*** who first thought of putting a ***sharp*** stone at the end of a ***stick*** made a brilliant discovery: he ***had*** joined two things to make a ***very*** useful tool, the spear. Flint, found ***in*** many rocks, became a common cutting ***tool*** in the Paleolithic period of man's ***development***. Since no wood or bone tools ***have*** survived, we know of this man ***by*** his stone implements, with which he ***could*** kill animals, cut up the meat, ***and*** scrape the skins, as well as ***draw*** pictures on the walls of the ***caves*** where he lived during the winter.

In the warmer seasons, man wandered on ***the*** steppes of Europe without a fixed ***home***, always foraging for food. Perhaps the ***women*** carried nuts and berries in shells ***or*** skins or even in light, woven ***baskets***. Wherever they camped, the primitive people ***made*** fires by striking flint for sparks ***and*** using dried seeds, moss, and rotten ***wood*** for tinder. With fires that he kindled himself, man could keep wild animals away and could cook those that he killed, as well as provide warmth and light for himself.

Passage 2:

Man is the only living creature that can make and use tools. He is the most teachable of living beings, earning the name of Homo sapiens. ***-1-*** ever restless brain has used the ***-2-*** and the wisdom of his ancestors ***-3-*** improve his way of life. Since ***-4-*** is able to walk and run ***-5-*** his feet, his hands have always ***-6-*** free to carry and to use ***-7-***. Man's hands have served him well ***-8-*** his life on earth. His development, ***-9-*** can be divided into three major ***-10-***, is marked by several different ways **-11-** life.

Up to 10,000 years ago, ***-12-*** human beings lived by hunting and ***-13-***. They also picked berries and fruits, ***-14-*** dug for various edible roots. Most ***-15-,*** the men were the hunters, and ***-16-*** women acted as food gatherers. Since ***-17-*** women were busy with the children, ***-18-*** men handled the tools. In a ***-19-*** hand, a dead branch became a -20- to knock down fruit or to ***-21-*** for tasty roots. Sometimes, an animal ***-22-*** served as a club, and a ***-23-*** piece of stone, fitting comfortably into ***-24-*** hand, could be used to break ***-25-*** or to throw at an animal. ***-26-*** stone was chipped against another until ***-27-*** had a sharp edge. The primitive ***-28-*** who first thought of putting a ***-29-*** stone at the end of a ***-30-*** made a brilliant discovery: he ***-31-*** joined two things to make a ***-32-*** useful tool, the spear. Flint, found ***-33-*** many rocks, became a common cutting ***-34-*** in the Paleolithic period of man's ***-35-***. Since no wood or bone tools ***-36-*** survived, we know of this man ***-37-*** his stone implements, with which he ***-38-*** kill animals, cut up the meat, ***-39-*** scrape the skins, as well as ***-40-*** pictures on the walls of the ***-41-*** where he lived during the winter.

-42- the warmer seasons, man wandered on ***-43-*** steppes of Europe without a fixed ***-44-***, always foraging for food. Perhaps the ***-45-*** carried nuts and berries in shells ***-46-*** skins or even in light, woven ***-47-***. Wherever they camped, the primitive people ***-48-*** fires by striking flint for sparks ***-49-*** using dried seeds, moss, and rotten ***-50-*** for tinder. With fires that he kindled himself, man could keep wild animals away and could cook those that he killed, as well as provide warmth and light for himself.

Bank of exact and lexically acceptable answers in order of presentation

Gap Number	**Exact Answer**	**Lexically acceptable answers**
1	his	man's, our, the
2	knowledge	accomplishments, culture, cunning, examples, experience(s), hands, ideas, information, ingenuity, instinct, intelligence, mistakes, nature, power, skill(s), talent, teaching, technique, thought, will, wit, words, work
3	to	
4	man	he
5	on	upon, using, with
6	been	felt, hung, remained
7	tools	adequately, carefully, conventionally, creatively, diligently, efficiently, freely, implements, objects, productively, readily, them, things, weapons
8	during	All, for, improving, in, through, throughout, with
9	which	Also, basically, conveniently, easily, historically, however, often, since, that, thus
10	periods	Areas, categories, divisions, eras, facets, groups, parts, phases, sections, stages, steps, topics, trends
11	of	For, in through, towards
12	all	Early, hungry, many, most, only, primitive, the, these
13	fishing	Farming, foraging, gathering, killing, scavenging, scrounging, sleeping, trapping
14	and	Often, ravenously, some, the
15	often	Always, emphatically, important, nights, normally, of, times, trips
16	the	All, house, many, most, older, their, younger
17	the	All, many, married, most, often, older, primate, these
18	the	All, constructive, many, most, older, primate, tough, younger
19	man's	Able, big, closed, coordinated, creative deft, empty, free, human('s), hunters, learned, needed, needy, person's, right, single, skilled, skilful, small, strong, trained
20	tool	Club, device, instrument, pole, rod, spear, stick, weapon
21	dig	Burrow, excavate, probe, search, test
22	bone	Arm, easily, foot, had, hide, horn, leg, skull, tail, tusk
23	sharp	Big, chipped, fashioned, flat, hard, heavy, large, rough, round, shaped, sizeable, small, smooth, solid, strong, soft, thin
24	the	A, his, man's, one('s)
25	nuts	Apart, bark, bones, coconuts, down, firewood, food, heads, ice, items, meat, objects, open, rocks, shells, sticks, stone, things, tinder, trees, wood
26	one	A, each, flat, glass, hard, obsidian, shale, softer, some, the, then, this
27	it	Each, one, they
28	man	Being, creature, human's, hunter, men, owner, people, person
29	sharp	Glass, hard, jagged, large, lime, pointed, sharpened, small
30	stick	Bone, branch, club, log, pole, rod, shaft
31	had	Accidentally, cleverly, clumsily, conveniently, creatively, dexterously, double, easily, first, ingeniously,

		securely, simply, soon, suddenly, tastefully, tightly, then, would
32	very	Bad, extremely, good, hunter's, incredibly, intelligent, long, modern, most, necessarily, new, portentously, quite, tremendously, useful
33	in	All, among, amongst, by, inside, on, that, using, within
34	tool	Device, edge, implement, instrument, item, material, method, object, piece, practice, stone, utensil
35	development	Age, ancestry, discoveries, era, evolution, existence, exploration, history, life, time
36	have	Actually, apparently, ever
37	by	And, for, from, had, made, through, used, using
38	could	Did, would
39	and	Carefully, help, or, skilfully, then, would
40	draw	Carve, create, drawing, engrave, hang, paint, painting, place, sketch, some, the
41	cave(s)	Animals, place(s), room
42	in	And, during, with
43	the	Across, aimless, all, barren, in, dry, flat, high, long, many, plain, stone, through, to, toward, unknown, various
44	home	Appetite, camp, course, destination, destiny, diet, direction, domain, foundation, habitat, income, knowledge, location, lunch, map, meal, path, pattern, place, plan, route, supplement, supply, time, weapons
45	women	Children, families, group, human, hunter, man, men, people, primitives, voyager, wanderers, woman
46	or	And, animal, animal's covered, in, like, of, on, their, using, with
47	baskets	Bags, blankets, chests, cloth(es), fabric, garments, hides, material, nets, pouches, sacks
48	made	Began, built, lighted, lit, produced, started, used
49	and	Also, by occasionally, or, then, together, while
50	wood	Bark, branches, dung, forage, grass, leaves, lumber, roots, skin, timber, tree(s)

Appendix F. Comprehension of gender task – individual results

ID	**Dominance Score** (raw)	**Welsh proficiency score** (raw/44)	**Possessive adjective - mutations**		**Anaphoric pronouns – no mutations**	
			Ei + SM [m] Raw / 10	Ei + AM [f] Raw / 10	‘fe’ [m] Raw / 8	‘hi’ [f] Raw / 8
1	34.32	36	10	10	8	8
2	45.78	30	7	4	3	5
3	-37.96	44	10	10	8	8
4	39.24	40	1	6	7	7
5	-15.62	43	10	10	8	8
6	-35.51	43	8	9	6	5
7	-44.41	43	9	9	6	4
8	75.2	42	5	8	4	5
9	3	36	4	7	7	5
10	-55.3	42	7	8	7	6
11	48.86	43	10	9	8	8
12	29.98	37	7	8	5	6
13	-31.24	44	5	8	7	7
14	-100.08	40	7	10	7	5
15	57.49	38	5	6	4	4
16	-32.23	41	7	9	5	6
17	-37.96	41	9	7	6	8
18	18.26	39	6	4	5	5
19	23.16	42	8	9	8	8
20	39.97	40	5	9	8	5
21	-30.25	43	10	10	8	8
22	-118.06	44	9	10	7	8
23	-44.05	43	10	10	8	8
24	-13.52	35	7	4	4	4
25	60.67	31	4	3	3	3
26	-8.99	37	7	6	5	4
27	80.19	28	8	5	8	8
28	19.63	42	10	10	8	8
29	146.88	43	9	7	3	5
30	-36.68	43	9	10	6	8
31	-92.63	32	10	10	8	8
32	63.67	37	8	5	2	3
33	-18.8	44	10	10	8	8
34	-42.04	40	6	8	4	6
35	-51.03	43	9	6	5	2
36	88.27	34	9	10	6	8
37	-111.34	43	10	10	8	8
38	63.31	40	5	8	6	7
39	24.97	40	10	8	8	8
40	46.41	38	8	4	2	4

Appendix G. Elicited Production task - individual results

Gender accuracy by numeral – scoring system 1

ID	Dominance (raw)	Welsh proficiency (raw/44)	2		3		4	
			Dau [m]	Dwy [f]	Tri [m]	Tair [f]	Pedwar [m]	Pedair [f]
1	34.32	36	7	5	8	5	4	6
2	45.78	30	8	0	8	0	6	3
3	-37.96	44	8	6	8	7	8	6
4	39.24	40	7	3	7	3	5	4
5	-15.62	43	8	8	8	7	7	8
6	-35.51	43	8	8	7	6	8	7
7	-44.41	43	7	6	5	3	5	7
8	75.2	42	5	6	5	3	8	7
9	3	36	4	8	4	7	8	3
10	-55.3	42	6	8	7	6	7	8
11	48.86	43	8	8	8	7	8	8
12	29.98	37	8	8	8	2	8	7
13	-31.24	44	4	8	5	6	5	5
14	-100.08	40	8	3	5	4	7	4
15	57.49	38	7	4	2	7	8	5
16	-32.23	41	3	7	4	4	5	6
17	-37.96	41	3	7	6	6	7	4
18	18.26	39	8	3	0	8	5	2
19	23.16	42	7	7	5	4	6	7
20	39.97	40	5	7	4	6	7	5
21	-30.25	43	7	8	8	6	6	7
22	-118.06	44	8	8	8	6	8	6
23	-44.05	43	7	7	8	7	8	8
24	-13.52	35	0	8	1	8	3	5
25	60.67	31	8	2	7	2	8	1
26	-8.99	37	8	1	8	0	8	0
27	80.19	28	8	4	7	4	8	4
28	19.63	42	5	8	7	7	5	7
29	146.88	43	6	3	6	5	6	5
30	-36.68	43	5	7	8	5	6	5
31	-92.63	32	8	8	8	8	8	8
32	63.67	37	2	5	5	5	6	4
33	-18.8	44	8	8	8	8	8	8
34	-42.04	40	3	5	2	7	6	7
35	-51.03	43	6	7	2	7	4	6
36	88.27	34	8	1	8	2	5	7
37	-111.34	43	7	7	8	8	8	7
38	63.31	40	5	6	6	4	7	3
39	24.97	40	8	8	7	7	7	7
40	46.41	38	8	2	8	5	8	1

Gender-mutation accuracy by numeral – scoring system 2

ID	**Dominance** (raw)	**Welsh proficiency** (raw/44)	**2 + mutation**		**3 +/- mutation**		**4 - mutation**	
			Dau [m] SM	Dwy [f] SM	Tri [m] AM	Tair [f] No mut	Pedwar [m] No mut	Pedair [f] No mut
1	34.32	36	5	4	6	3	3	5
2	45.78	30	0	0	2	1	6	2
3	-37.96	44	8	6	7	7	8	6
4	39.24	40	1	2	2	4	5	3
5	-15.62	43	8	8	5	7	7	8
6	-35.51	43	8	8	2	6	8	7
7	-44.41	43	6	6	2	2	5	7
8	75.2	42	2	3	1	3	8	7
9	3	36	1	3	1	7	8	3
10	-55.3	42	6	8	2	7	7	8
11	48.86	43	8	8	3	7	8	8
12	29.98	37	6	8	2	2	8	7
13	-31.24	44	3	7	1	6	5	4
14	-100.08	40	6	2	2	4	7	4
15	57.49	38	1	3	1	6	8	5
16	-32.23	41	1	6	1	3	5	1
17	-37.96	41	2	7	2	6	7	4
18	18.26	39	4	2	0	6	4	0
19	23.16	42	1	6	1	3	6	7
20	39.97	40	1	6	1	6	7	3
21	-30.25	43	6	8	2	5	6	7
22	-118.06	44	8	8	2	6	8	6
23	-44.05	43	6	7	8	7	8	8
24	-13.52	35	0	8	0	3	3	2
25	60.67	31	0	1	2	2	7	1
26	-8.99	37	6	1	3	0	6	0
27	80.19	28	6	4	6	4	8	5
28	19.63	42	5	8	5	7	5	7
29	146.88	43	4	3	1	5	6	5
30	-36.68	43	5	7	2	5	6	5
31	-92.63	32	8	8	8	8	8	8
32	63.67	37	2	5	1	3	2	1
33	-18.8	44	8	8	4	8	8	8
34	-42.04	40	3	4	1	3	6	7
35	-51.03	43	3	7	2	3	3	5
36	88.27	34	8	1	6	2	5	6
37	-111.34	43	7	7	2	8	8	7
38	63.31	40	2	6	5	2	7	2
39	24.97	40	8	7	2	7	7	7
40	46.41	38	5	1	2	4	8	1

Appendix H. Self-paced reading task items

Experimental Condition 1: Gender in conjunction with mutations via the numeral ‘2’

Noun	Item: grammatical	Item: ungrammatical	CQ	Gender / animacy
Pengwin Penguin	Roedd **dau** bengwin yn sefyll yn stond yn y môr	Roedd **dwy** bengwin yn sefyll yn stond yn y môr	Roedden nhw yn y dŵr ✓	M Animal
Bochdew Hamster	Mae’r **ddau** fochdew yn rolio o gwmpas yn y cawell	Mae’r **ddwy** fochdew yn rolio o gwmpas yn y cawell	Roedden nhw yn y cwpwrdd **X**	M Animal
Meddyg Doctor	Roedd **dau** feddyg yn barod i weld eu cleifion heddiw	Roedd **dwy** feddyg yn barod i weld eu cleifion heddiw	Meddygon ydyn nhw ✓	M Human
Bachgen boy	Mae **dau** fachgen mawr yn chwarae pêl-droed yn y parc	Mae **dwy** fachgen mawr yn chwarae pêl-droed yn y parc	Merched ydyn nhw **X**	M Human
Tegell Kettle	Mae **dau** degell ar ben y cownter yn y gegin	Mae **dwy** degell ar ben y cownter yn y gegin	Maen nhw yn y gegin ✓	M Inanimate
Cwmwl Cloud	Gwelais **ddau** gwmwl yn pasio yn yr awyr las heddiw	Gwelais **ddwy** gwmwl yn pasio yn yr awyr las heddiw	Nos ydy hi **X**	M Inanimate
Capel chapel	Bydd **dau** gapel yn cau yn y pentref yn anffodus	Bydd **dwy** gapel yn cau yn y pentref yn anffodus	Maen nhw yn y pentref ✓	M Inanimate
Darlun picture	Mae **dau** ddarlun pert yn hongian yn yr amgueddfa fawr	Mae **dwy** ddarlun pert yn hongian yn yr amgueddfa fawr	Maen nhw yn y siop **X**	M Inanimate
Gwiwer Squirrel	Roedd **dwy** wiwer yn casglu cnau ar gyfer y gaeaf	Roedd **dau** wiwer yn casglu cnau ar gyfer y gaeaf	Roedden hyn ar gyfer y tywydd oer ✓	F Animal
Cleren Fly	Mae’r **ddwy** gleren yn hedfan yn uchel wrth y ffenestr	Mae’r **ddau** gleren yn hedfan yn uchel wrth y ffenestr	Roedden nhw’n nofio **X**	F Animal
Modryb Aunty	Bydd **dwy** fodryb yn mynd i weld eu nith newydd	Bydd **dau** fodryb yn mynd i weld eu nith newydd	Byddan nhwn gweld eu nith ✓	F Human
Menyw Woman	Mae’r **ddwy** fenyw yn edrych yn debyg iawn i’w gilydd	Mae’r **ddau** fenyw yn edrych yn debyg iawn i’w gilydd	Brodyr ydyn nhw **X**	F Human
Telyn Harp	Mae **dwy** delyn ar werth yn y papur newydd lleol	Mae **dau** delyn ar werth yn y papur newydd lleol	Mae piano ar werth **X**	F Inanimate
Coeden Tree	Mae’r **ddwy** goeden yn y dref yn ganmlwydd oed heddiw	Mae’r **ddau** goeden yn y dref yn ganmlwydd oed heddiw	Maen nhw yn y dref ✓	F Inanimate
Deilen Leaf	Chwythwyd **dwy** ddeilen ar y lawnt gan y gwynt cryf	Chwythwyd **dau** ddeilen ar y lawnt gan y gwynt cryf	Roedd hi’n bwrw glaw **X**	F Inanimate
Calon Heart	Bydd **dwy** galon yn curo’n drwm trwy wneud naid bynji	Bydd **dau** galon yn curo’n drwm trwy wneud naid bynji	Byddan nhw’n curo’n drwm ✓	F Inanimate

Experimental Condition 2: Gender independent of mutations via the numeral '4'

Noun	Item: grammatical	Item: ungrammatical	CQ	Gender / animacy
Draenog Hedgehog	Roedd **pedwar** draenog yn cysgu'n sownd dan y berth heno	Roedd **pedair** draenog yn cysgu'n sownd dan y berth heno	Roedden nhw'n cysgu ✓	M Animal
Broga Frog	Eisteddodd **pedwar** broga tew wrth ymyl y pwll pysgod pert	Eisteddodd **pedair** broga tew wrth ymyl y pwll pysgod pert	Pysgod ydyn nhw **X**	M Animal
Garddwr Gardener	Bydd **pedwar** garddwr yn cwrdd yn y sioe amaethyddol wedyn	Bydd **pedair** garddwr yn cwrdd yn y sioe amaethyddol wedyn	Byddan nhw yn y sioe ✓	M Human
Cogydd Cook	Bydd **pedwar** cogydd yn brysur yng nghegin y tŷ bwyta	Bydd **pedair** cogydd yn brysur yng nghegin y tŷ bwyta	Byddan nhw ar lan y môr **X**	M Human
Blodyn Flower	Mae **pedwar** blodyn yn tyfu'n dda yng ngardd fy mam	Mae **pedair** blodyn yn tyfu'n dda yng ngardd fy mam	Coed ydyn nhw **X**	M Inanimate
Polyn Pole	Mae **pedwar** polyn ar bob cornel cae chwarae'r ysgol fach	Mae **pedair** polyn ar bob cornel cae chwarae'r ysgol fach	Roedden nhw ar gae chwarae'r ysgol ✓	M Inanimate
Cwpwrdd Cupboard	Roedd **pedwar** cwpwrdd pren yng nghegin yr hen blas enfawr	Roedd **pedair** cwpwrdd pren yng nghegin yr hen blas enfawr	Roedden nhw yn y lolfa **X**	M Inanimate
Castell Castle	Trwsiwyd **pedwar** castell G*ŵyr* yn wych gan yr adeiladwyr newydd	Trwsiwyd **pedair** castell G*ŵyr* yn wych gan yr adeiladwyr newydd	Roedden nhw'n wych ✓	M Inanimate
Dafad Sheep	Roedd **pedair** dafad yn pori'n hapus braf yn y cae	Roedd **pedwar** dafad yn pori'n hapus braf yn y cae	Roedden nhw mewn coedwig **X**	F Animal
Llewes Female lion	Roedd **pedair** llewes ddewr yn chwyrnu'n uchel yn eu cwsg	Roedd **pedwar** llewes ddewr yn chwyrnu'n uchel yn eu cwsg	Roedden nhw'n cysgu ✓	F Animal
Mam-gu Grandmothe r	Roedd **pedair** mam-gu yn cerdded yn y parc gyda'u hwyrion	Roedd pedwar mam-gu yn cerdded yn y parc gyda'u hwyrion	Roedden nhw yn yr amgueddfa **X**	F Human
Bydwraig Midwife	Mae **pedair** bydwraig ar y ward yn yr ysbyty leol	Mae **pedwar** bydwraig ar y ward yn yr ysbyty leol	Maen nhw'n gweithio mewn ysbyty✓	F Human
Basged Basket	Prynais **bedair** basged o farchnad y pentref y bore 'ma	Prynais **bedwar** basged o farchnad y pentref y bore 'ma	Daethon nhw o'r pentref ✓	F Inanimate
Potel Bottle	Cwympodd **pedair** potel o'r oergell pan agorais y drws trwm	Cwympodd **pedwar** potel o'r oergell pan agorais y drws trwm	Roedden nhw yn y cwpwrdd **X**	F Inanimate
Pluen Feather	Bydd **pedair** pluen bert yn ymddangos yn yr het ffasiynol	Bydd **pedwar** pluen bert yn ymddangos yn yr het ffasiynol	Roedd yr het yn ffasiynol ✓	F Inanimate
Cawod Shower	Bydd **pedair** cawod ar gael yn ystod yr *ŵ*yl gerddoriaeth	Bydd **pedwar** cawod ar gael yn ystod yr *ŵ*yl gerddoriaeth	Bydd bath ar gael **X**	F Inanimate

Experimental Condition 3: Mutations independent of gender via select pre-nominal adjectives

Noun	**Item: grammatical**	**Item: ungrammatical**	**CQ**	**Gender / animacy**
Morfil Whale	Golchwyd **hen f**orfil i arfordir Sir Fôn neithiwr yn anffodus	Golchwyd **hen m**orfil i arfordir Sir Fôn neithiwr yn anffodus	Roedd yn Sir Fynwy X	M Animal
Twrci Turkey	Roedd **hen d**wrci yn crwydro'n araf ar fuarth y fferm	Roedd **hen t**wrci yn crwydro'n araf ar fuarth y fferm	Roedd yn araf ✓	M Animal
Brenin King	Roedd **prif f**renin y byd yn garedig ac yn hael	Roedd **prif b**renin y byd yn garedig ac yn hael	Roedd yn gas **X**	M Human
Tad-cu Grandad	Mae **f'annwyl d**ad-cu yn hapus iawn wrth yrru'i gar newydd	Mae **f'annwyl t**ad-cu yn hapus iawn wrth yrru'i gar newydd	Mae'n gyrru car ✓	M Human
Llyfr Book	Fy **hoff l**yfr oedd yn ffrind cyson a ffyddlon iawn	Fy **hoff ll**yfr oedd yn ffrind cyson a ffyddlon iawn	Roedd yn ffrind da iawn ✓	M Inanimate
Cwrw	Fy hoff gwrw yn dod o'r bragdy enwog yn Abergwaun	Fy hoff cwrw yn dod o'r bragdy enwog yn Abergwaun	Mae'r bradgy yn Abertystwyth **X**	M Inanimate
Cerdyn Card	Roedd **hen g**erdyn ar waelod y cist yn yr atig	Roedd **hen c**erdyn ar waelod y cist yn yr atig	Roedd yn y cist ✓	M Inanimate
Mynydd Mountain	Fy **hoff f**ynydd yw Pen-y-Fan achos y golygfeydd hardd iawn	Fy **hoff m**ynydd yw Pen-y-Fan achos y golygfeidd hardd iawn	Roedd golygfa ofnadwy **X**	M Inanimate
Caseg Female horse	Mae'r **hen g**aseg yn mwynhau gorffwys yn y cae glas	Mae'r **hen c**aseg yn mwynhau gorffwys yn y cae glas	Mae'n hoffi rhedeg **X**	F Animal
Gafr Goat	Mae'r **brif** afr yn arwain y gweddill i'r nant fach	Mae'r **brif g**afr yn arwain y gwellydd i'r nant fach	Maen nhw'n mynd i'r dŵr ✓	F Animal
Dawns-wraig Dancer	Roedd **unig dd**awnswraig ar y llwyfan yn ystod y perfformiad	Roedd **unig d**awnswraig ar y llwyfan yn y ystod y perfformiad	Roedd sawl un yno **X**	F Human
Dynes Woman	Bydd **hen dd**ynes yn gwerthu tocynnau y raffl yn fuan	Bydd **hen d**ynes yn gwerthu tocynnau y raffl yn fuan	Bydd raffl ar gael ✓	F Human
Cneuen Nut	Roedd **hen g**neuen yn pydru ar waelod y bag siopa	Roedd **hen c**neuen yn pydru ar waelod y bag siopa	Roedd yn ffres **X**	F Inanimate
Gorsaf Station	Y **brif** orsaf ar y lein ydy'r un yng Nghaerdydd	Y **brif g**orsaf ar y lein ydy'r un yng Nghaerdydd	Prif ddinas Cymru oedd y lle ✓	F Inanimate
Carreg Stone	Bydd **hen g**arreg o oes y deinosoriaid ar werth yfory	Bydd **hen c**arreg o oes y deinosoriaid ar werth yfory	Roedd ar werth ddoe **X**	F Inanimate
Clustog Pillow	Roedd **hen g**lustog ar y gadair ac roedd yn gysurus	Roedd **hen c**lustog ar y gadair ac roedd yn gysurus	Roedd yn gyfforddus ✓	F Inanimate

Experimental Condition 4: Gender encoded mutations via the determiner 'y' (SM [f] nouns and no mutation [m] nouns)

Noun	**Item: grammatical**	**Item: ungrammatical**	**CQ**	**Gender / animacy**
Llwynog Fox	Roedd y **ll**wynog yn y teulu yn chwarae'n hapus braf	Roedd y **l**wynog yn y teulu yn chwarae'n hapus braf	Roedd yn bwyta **X**	M Animal
Crwban Tortoise	Roedd y **c**rwban yn crwydro'n ddiogel yn Fferm Ffoli ddoe	Roedd y **g**rwban yn crwydro'n ddiogel yn Fferm Ffoli ddoe	Roedd yn saff ✓	M Animal
Cigydd Butcher	Roedd y **c**igydd yn cystadlu yn y sioe Frenhinol, Llanelwedd	Roedd y **g**igydd yn cystadlu yn y sioe Frenhinol, Llanelwedd	Roedd yn y gwesty **X**	M Human
Dysgwr Learner	Roedd y **d**ysgwr da yn adrodd ar y llwyfan echddoe	Roedd y **dd**ysgwr da yn adrodd ar y llwyfan echddoe	Roedd yn perfformio ✓	M Human
Gwesty Hotel	Roedd y **g**westy ar gau yn y ddinas oherwydd Covid	Roedd y westy ar gau yn y ddinas oherwydd Covid	Roedd yn dŷ **X**	M Inanimate
Bwthyn Cottage	Bydd y **b**wthyn ger y traeth ar gael i'r ymwelwyr	Bydd y **f**wthyn ger y traeth ar gael i'r ymwelwyr	Roedd ar lan y môr ✓	M Inanimate
Cartŵn Cartoon	Bydd y **c**artŵn ar y sianel Disney yn ddoniol iawn	Bydd y **g**artŵn ar y sianel Disney yn ddoniol iawn	Bydd ar S4C **X**	M Inanimate
Dosbarth Class	Roedd y **d**osbarth yn yr ysgol gynradd ar drip heddiw	Roedd y **dd**osbarth yn yr ysgol gynradd ar drip heddiw	Roedd taith o'r ysgol ✓	M Inanimate
Madfall Lizard	Roedd y **f**adfall frown yn mwynhau torheulo draw ym Majorca	Roedd y **m**adfall frown yn mwynhau torheulo draw ym Majorca	Mae'n hapus yn yr haul ✓	F Animal
Cigfran Raven	Roedd y **g**igfran yn hedfan yn swnllyd uwch fy mhen	Roedd y **c**igfran yn hedfan yn swnllyd uwch fy mhen	Roedd yn dawel **X**	F Animal
Morwyn Maid / Maiden	Roedd y **f**orwyn yn cael ei chyflogi yn y palas	Roedd y **m**orwyn yn cael ei chyflogi yn y palas	Roedd merch yn gweithio ✓	F Human
Dysg-wraig Female learner	Bydd y **dd**ysgwraig yn trafod ei gwaith yn y gynhadledd	Bydd y **d**ysgwraig yn trafod ei gwaith yn y cynhadledd	Bydd yn gweiddi **X**	F Human
Dinas City	Bydd y **dd**inas yn cynnal ras seiclo noddedig yn fuan	Bydd y **d**inas yn cynnal ras seiclo noddedig yn fuan	Bydd pobl yn reidio beic ✓	F Inanimate
Teisen Cake	Roedd y **d**eisen boeth yn ddeniadol ac yn flasus iawn	Roedd y **t**eisen boeth yn ddeniadol ac yn flasus iawn	Roedd yn oer **X**	F Inanimate
Cadair Chair	Bydd y **g**adair bren yn dod i'r arddangosfa erbyn yfory	Bydd y **c**adair bren yn dod i'r arddangosfa erbyn yfory	Bydd ar gael yfory ✓	F Inanimate
Blanced Blanket	Roedd y **f**lanced ar y gwely ac roeddwn i'n oer!	Roedd y **b**lanced ar y gwely ac roeddwn i'n oer!	Roeddwn i'n dwym **X**	F Inanimate

Non-experimental Condition: Distractor condition 1, via the gendered numeral '2' and nouns starting with non-mutatable letters

Noun	**Item: grammatical**	**Item: ungrammatical**	**CQ**	**Gender / animacy**
Asyn Male donkey	Mae'r **ddau** asyn yn cerdded yn hamddenol i'w stablau clud	Mae'r **ddwy** asyn yn cerdded yn hamddenol i'w stablau clud	Maen nhw'n mynd i'r stablau ✓	M Animal
Ebol Foal	Roedd **dau** ebol bychan yn prancio'n fywiog ar y fferm	Roedd **dwy** ebol bychan yn prancio'n fywiog ar y fferm	Roedden nhw yn y goedwig **X**	M Animal
Ffermwr Male farmer	Siaradodd **dau** ffermwr hapus tan oriau man y bore bach	Siaradodd **dwy** ffermwr hapus tan oriau man y bore bach	Roedden nhw'n sgwrsio tan y bore ✓	M Human
Siopwr Male shopper	Cwynodd **dau** siopwr dig heddiw am bris anheg eu nwyddau	Cwynodd **dwy** siopwr dig heddiw am bris anheg eu nwyddau	Roedd cwyn am y maes parcio **X**	M Human
Siswrn Scissor	Mae **dau** siswrn yn y bocs ar y silff uchaf	Mae **dwy** siswrn yn y bocs ar y silff uchaf	Roedd papur yno **X**	M Inanimate
Siocled Chocolate	Roedd **dau** siocled plaen ar ôl ar fwrdd y gegin	Reodd **dwy** sioceld plaen ar ôl ar fwrdd y gegin	Roedden nhw yn y gegin ✓	M Inanimate
Ffosil fossil	Darganfyddwyd **dau** ffosil tlws ger y môr ar draeth Trefdraeth	Darganfyddwyd **dwy** ffosil tlws ger y môr ar draeth Trefdraeth	Roedden nhw yn y ddinas **X**	M Inanimate
Ffreutur Refectory	Mae **dau** ffreutur ar gael ar gampws newydd Prifysgol Abertawe	Mae **dwy** ffreutur ar gael ar gampws newydd Prifysgol Abertawe	Bydd llefydd bwyta ar gael ✓	M Inanimate
Neidr Snake	Roedd **dwy** neidr gas iawn yn hisian yn eu nyth	Roedd **dau** neidr gas iawn yn hisian y neu nyth	Roedden nhw'n hyfryd iawn **X**	F Animal
Asen Female donkey	Mae'r **ddwy** asen yn cerdded ar hyd y llwybr cul	Mae'r **ddau** asen yn cerdded ar hyd y llwybr cul	Roedden nhw'n cerdded ✓	F Animal
Siopwraig Female shopper	Bydd **dwy** siopwraig yn agor y drysau yn hwyrach yfory	Bydd **dau** siopwraig yn agor y drysau yn hwyrach yfory	Byddan nhw'n agor yn gynnar **X**	F Human
Ffermwraig Female farmer	Mae'r **ddwy** ffermwraig yn mynd i weld y lloi bach	Mae'r **ddau** ffermwraig yn mynd i weld y lloi bach	Byddan nhw'n gweld y lloi ✓	F Human
Seren Star	Roedd **dwy** seren yn disgleirio'n llachar ar noson glur hyfryd	Roedd **dau** seren yn disgleirio'n llachar ar noson glur hyfryd	Noson gymylog oedd hi **X**	F Inanimate
Ffatri Factory	Bydd **dwy** ffatri fach yn cau eleni oherwydd diffyg gweithwyr	Bydd dau **ffatri** fach yn cau eleni oherwydd diffyg gweithwyr	Byddan nhw'n cau ✓	F Inanimate
Siaced Jacket	Bydd **dwy** siaced ar gael ar gyfer ein bwgan brain	Bydd **dau** siaced ar gael ar gyfer ein bwgan brain	Bydd trwser ar gael **X**	F Inanimate
Ffedog Apron	Mae **dwy** ffedog lân i ti yn y cwpwrdd pren	Mae **dau** ffedog lân i ti yn y cwpwrdd pren	Maen nhw yn y cwpwrdd ✓	F Inanimate

Non-experimental Condition: Distractor condition 2, via the gendered numeral '3' and nouns starting with letters to be AM

Noun	Item: grammatical	Item: ungrammatical	CQ	Gender / animacy
Ceffyl Horse	Mae **tri ch**effyl yn carlamu yn y cae bach heddiw	Mae **tair c**effyl yn carlamu yn y cae bach heddiw	Maen nhw'n rhedeg yn gylfym ✓	M Animal
Tarw Bull	Bydd **tri th**arw yn mynd i'r farchnad ar y penwythnos	Bydd **tair t**arw yn mynd i'r farchnad ar y penwythnos	Defaid ydyn nhw **X**	M Animal
Postmon Postman	Roedd **tri ph**ostmon yn trafod ble i fynd bore ddoe	Roedd **tair p**ostmon yn trafod ble i fynd bore ddoe	Roeddwn nhw'n sgwrsio ✓	M Human
Plismon Policeman	Bydd **tri ph**lismon yn cwrdd yn yr orsaf heddlu heddiw	Bydd **tair p**lismon yn cwrdd yn yr orsaf heddlu heddiw	Maen nhw'n cwrdd yn y bwyty **X**	M Human
Cerbyd Vehicle	Roedd **tri ch**erbyd yn blocio'r ffordd yn y stryd fawr	Roedd **tair c**erbyd yn blocio'r ffordd yn y stryd fawr	Roedd ceffylau yn y stryd fawr **X**	M Inanimate
Clwtyn Rag/cloth	Mae'r **tri ch**lwtyn brwnt yn y peiriant golchi yn barod	Mae'r **tair c**lwtyn brwnt yn y peiriant golchi yn barod	Maen nhw'n cael eu golchi ✓	M Inanimate
Pwdin Pudding	Bydd **tri ph**wdin oer ar y fwydlen newydd nos yfory	Bydd **tair p**wdin oer ar y fwydlen newydd nos yfory	Byddan nhw'n dwym **X**	M Inanimate
Tegan Toy	Mae **tri th**egan y ferch fach bert ar y llawr	Mae **tair t**egan y ferch fach bert ar y llawr	Mae hi'n ddel ✓	M Inanimate
Cwcw Cuckoo	Roedd **tair c**wcw yn canu'n hapus ar gangen y goeden	Roedd **tri ch**wcw yn canu'n hapus ar gangen y goeden	Roedden nhw'n canu✓	F Animal
Teigres Tigress	Roedd **tair t**eigres yn ymladd yn ffyrnig ynghanol y jyngl	Roedd **tri th**eigres yn ymladd yn ffyrnig ynghanol y jyngl	Llewod oedden nhw **X**	F Animal
Peint-wraig Painter (female)	Collodd **tair p**eintwraig eu brwsys gorau pan adawon eu gwaith	Collodd **tri ph**eintwraig eu brwsys gorau pan adawon eu gwaith	Roedden nhw'n gadael eu gwaith ✓	F Human
Cymraes Welsh-woman	Mae **tair C**ymraes yn perfformio yn yr Eisteddfod Genedlaethol heddiw	Mae **tri Ch**ymraes yn perfformio yn yr Eisteddfod Genedlaethol heddiw	Byddan nhw yn y cyntedd **X**	F Human
Pibell Pipe/Tube	Roedd **tair p**ibell fawr ar organ yr eglwys wedi torri	Roedd **tri ph**ibell fawr ar organ yr eglwys wedi torri	Roedden nhw yn y capel **X**	F Inanimate
Pabell Tent	Daeth **tair p**abell i lawr ar ôl y gwyntoedd cryfion	Daeth **tri ph**abell i lawr ar ôl y gwyntoedd cryfion	Roedd hi'n wyntog ✓	F Inanimate
Powlen Bowl	Bydd **tair p**owlen o uwd yn aros i Elen Benfelen	Bydd **tri ph**owlen o uwd yn aros i Elen Benfelen	Bydd jeli yno **X**	F Inanimate
Cangen Branch	Roedd **tair c**angen yn plygu'n arw dan bwysau'r afalau cochion	Roedd **tri ch**angen yn plygu'n arw dan bwysau'r afalau cochion	Roedd yr afalau yn goch ✓	F Inanimate

Non-experimental Condition: Filler condition 1 = Plural nouns

Noun	**Grammatical**	**Ungrammatical**	**Gender / animacy**
Cwn Dogs	Roedd y cŵn yn rhedeg yn egnïol yn y parc	Roeddwn y cŵn yn rhedeg yn egnïol yn y parc	M Animal
Pysgod aur Goldfish	Roedd y pysgod aur yn nofio yn y pwll newydd	Roeddet y pysgod aur yn nofio yn y pwll newydd	M Animal
Cantorion Singers	Roedd y cantorion yn nerfus cyn perfformio ar y llwyfan	Roeddech y cantorion yn nerfus cyn perfformio ar y llwyfan	M Human
Dynion Men	Rhedodd y dynion yn gyflym iawn yn y gêm rygbi	Rhedais y dynion yn gyflym iawn yn y gêm rygbi	M Human
Drychau Mirrors	Cwympodd y drychau trwm iawn a thorri mewn deilchion man	Cwympais y drychau trwm iawn a thorri mewn deilchion man	M Inanimate
Tapiau Taps	Roedd y tapiau wedi torri yn y toiledau cyhoeddus neithiwr	Roedden y tapiau wedi torri yn y toiledau cyhoeddus neithiwr	M Inanimate
Byrddau Tables	Roedd y byrddau'n barod ar gyfer arholiadau'r myfyrwyr yn Ionawr	Roeddwn y byrddau'n barod ar gyfer arholiadau'r myfyrwyr yn Ionawr	M Inanimate
Gwelyau Beds	Roedd y gwelyau yn y siop yn edrych yn gyfforddus	Roeddet y gwelyau yn y siop yn edrych yn gyfforddus	M Inanimate
Cathod Cats	Roedd y cathod yn cysgu'n sownd ar y soffa cyfforddus	Roeddech y cathod yn cysgu'n sownd ar y soffa cyfforddus	F Animal
Buchod Cows	Roedd y buchod yn gorwedd ar y glaswellt gwlyb hir	Roedden y buchod yn gorwedd ar y glaswellt gwlyb hir	F Animal
Prifathrawesau Headmistresses	Siaradodd y prifathrawesau am oriau mewn cyfarfod pwysig y sir	Siaradais y prifathrawesau am oriau mewn cyfarfod pwysig y sir	F Human
Breninesau Queens	Daeth y breninesau o bell i drafod materion pwysig iawn	Des y breninesau o bell i drafod materion pwysig iawn	F Human
Desgiau Desks	Roedd y desgiau ar werth yn y siop ddodrefn leol	Roeddwn y desgiau ar werth yn y siop ddodrefn leol	F Inanimate
Cadeiriau Chairs	Roedd y cadeiriau pinc yn fodern ac yn ffasiynol iawn	Roeddet y cadeiriau pinc yn fodern ac yn ffasiynol iawn	F Inanimate
Torthau Loaves	Roedd y torthau o fara yn edrych yn ffres iawn	Roeddech y torthau o fara yn edrych yn ffres iawn	F Inanimate
Cacennau Cakes	Roedd y cacennau yn y siop yn edrych yn flasus	Roedden y cacennau yn y siop yn edrych yn flasus	F Inanimate

Non-experimental Condition: Filler condition 1 = Plural nouns with non-local gender marked contexts

Noun	**Item: grammatical**	**Item: ungrammatical**	**Gender / animacy**
Moch Pigs	Roedd y moch yn bwyta eu brecwast yn eu twlc	Roeddwn y moch yn bwyta eu brecwast yn eu twlc	M Animal
Dolffiniaid Dolphins	*Roedd y dolffiniaid yn dangos eu doniau yn y* môr	*Roeddet y dolffiniaid yn dangos eu doniau yn y* môr	M Animal
Tywysogau Princes	Roedd y tywysogau yn marchog eu ceffylau dros y mynyddoedd	Roeddech y tywysogau yn marchog eu ceffylau dros y mynyddoedd	M Human
Gwŷr Men	Canodd y gwŷr gân hyfryd i'w gwragedd yn y neuadd	Canais y gwŷr gân hyfryd i'w gwragedd yn y neuadd	M Human
Crysau Shirts	Roedd y crysau yn edrych yn smart gyda'u botymau sgleiniog	Roedden y crysau yn edrych yn smart gyda'u botymau sgleiniog	M Inanimate
Platiau Plates	Roedd y platiau wedi colli'u patrymau yn y peiriant newydd	Roeddwn y platiau wedi colli'u patrymau yn y peiriant newydd	M Inanimate
Bryniau Hills	Roedd y bryniau yn anodd i'w dringo oherwydd y niwl	Roeddet y bryniau yn anodd i'w dringo oherwydd y niwl	M Inanimate
Planhigion Plants	Roedd y planhigion yn edrych yn hardd gyda'u petalau pinc	Roeddech y planhigion yn edrych yn hardd gyda'u petalau pinc	M Inanimate
Llygod Mice	Rhedodd y llygod bychain nôl i'w cartref clud a chynnes	Rhedaist y llygod bychain nôl i'w cartref clud a chynnes	F Animal
Malwod Snails	Chwiliodd y malwod araf am le diogel i gysgodi'u cregyn	Chwilias y malwod araf am le diogel i gysgodi'u cregyn	F Animal
Merched Girls	Collodd y merched eu pyrsiau yn y ganolfan siopa brysur	Collais y merched eu pyrsiau yn y ganolfan siopa brysur	F Human
Mamau Mothers	Roedd y mamau'n casglu eu plant o'r ysgol gyda'i gilydd	Roedden y mamau'n casglu eu plant o'r ysgol gyda'i gilydd	F Human
Mellt Lightning	Roedd y mellt mor ddychrynllyd gyda'u fflachiadau sydyn a gwyllt	Roeddwn y mellt mor ddychrynllyd gyda'u fflachiadau sydyn a gwyllt	F Inanimate
Peli Balls	Collodd y peli siap ar ôl iddynt gael eu cicio	Collaist y peli siap ar ôl iddynt gael eu cicio	F Inanimate
Carafanau Caravans	Roedd y carafanau yn arddangos eu nodweddion newydd sbon heddiw	Roeddet y carafanau yn arddangos eu nodweddion newydd sbon heddiw	F Inanimate
Bordydd Tables	Collodd y bordydd eu coesau wrth i'r perchenogion symud tŷ	Colloch y bordydd eu coesau wrth i'r perchenogion symud tŷ	F Inanimate

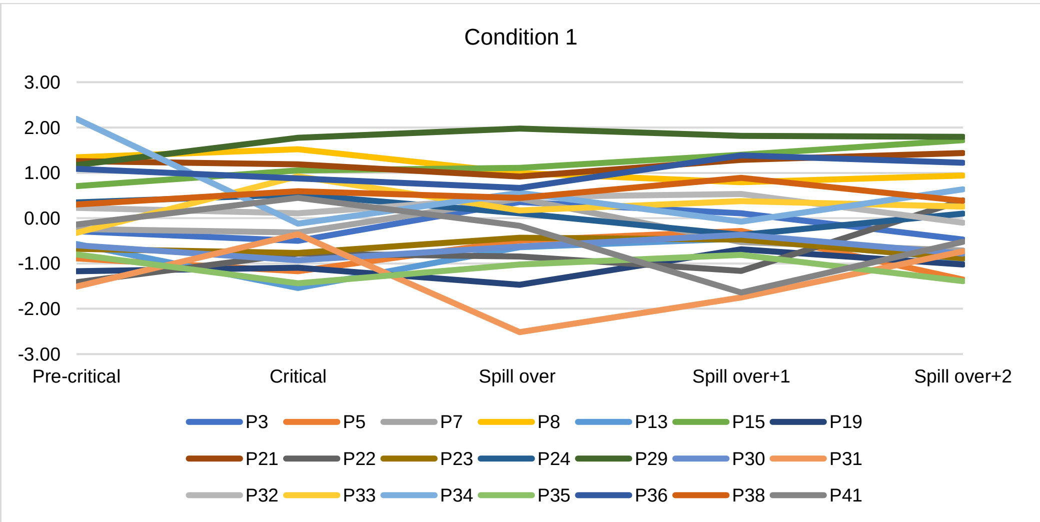
Appendix I. SPR Descriptive Results: Z-scores

The tables and figures include individual z-scores for the 4 experimental conditions, showing individual RT difference between the two levels of grammaticality, across the segments.

Descriptive Statistics for Condition 1 Individual Score Difference by Segment

ID	Pre-critical	Critical	Spill over	Spill over+1	Spill over+2
P3	-0.29	-0.50	0.36	0.11	-0.47
P5	-0.89	-1.17	-0.50	-0.28	-1.35
P7	-0.24	-0.31	0.43	-0.53	-0.71
P8	1.34	1.52	1.00	0.79	0.94
P13	-0.57	-1.54	-0.64	-0.45	-0.91
P15	0.71	1.05	1.11	1.40	1.72
P19	-1.17	-1.09	-1.47	-0.69	-1.03
P21	1.26	1.19	0.92	1.29	1.44
P22	-1.41	-0.78	-0.84	-1.16	0.39
P23	-0.68	-0.77	-0.43	-0.47	-0.89
P24	0.35	0.52	0.11	-0.37	0.10
P29	1.17	1.77	1.98	1.82	1.80
P30	-0.60	-0.93	-0.63	-0.37	-0.78
P31	-1.51	-0.35	-2.51	-1.75	-0.73
P32	0.23	0.11	0.47	0.53	-0.10
P33	-0.32	0.92	0.17	0.38	0.26
P34	2.19	-0.12	0.56	-0.07	0.64
P35	-0.80	-1.44	-1.02	-0.81	-1.39
P36	1.09	0.87	0.67	1.39	1.22
P38	0.30	0.60	0.45	0.89	0.38
P41	-0.14	0.45	-0.17	-1.64	-0.52

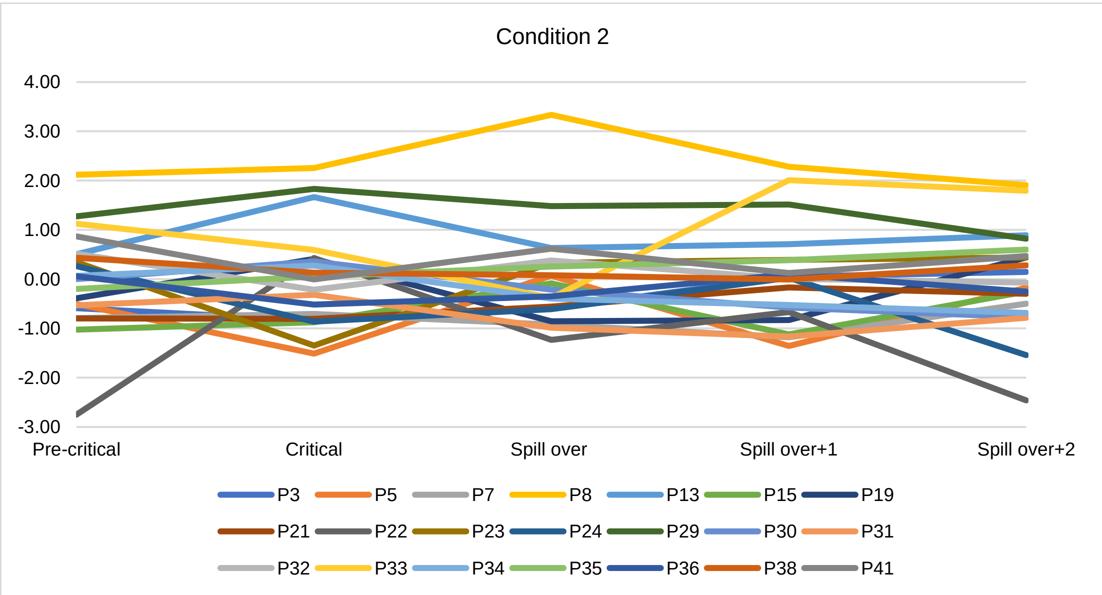
Descriptive Statistics for Condition 1 Individual Score Difference by Segment



Descriptive Statistics for Condition 2 Individual Score Difference by Segment

ID	Pre-critical	Critical	Spill over	Spill over+1	Spill over+2
P3	-0.59	-0.84	-0.60	0.05	0.15
P5	-0.50	-1.51	0.07	-1.36	-0.18
P7	-0.79	-0.71	-0.92	-1.18	-0.50
P8	2.12	2.25	3.33	2.28	1.91
P13	0.51	1.67	0.63	0.71	0.89
P15	-1.03	-0.88	-0.09	-1.12	-0.25
P19	-0.39	0.41	-0.85	-0.83	0.44
P21	-0.80	-0.81	-0.56	-0.17	-0.30
P22	-2.75	0.43	-1.24	-0.67	-2.46
P23	0.35	-1.35	0.33	0.39	0.44
P24	0.26	-0.86	-0.61	0.03	-1.54
P29	1.27	1.83	1.48	1.52	0.82
P30	0.00	0.36	-0.21	-0.58	-0.77
P31	-0.53	-0.32	-0.99	-1.17	-0.78
P32	0.51	-0.21	0.38	0.01	-0.07
P33	1.12	0.59	-0.38	2.01	1.79
P34	0.07	0.28	-0.40	-0.53	-0.69
P35	-0.20	0.05	0.25	0.39	0.60
P36	0.06	-0.52	-0.35	0.10	-0.25
P38	0.43	0.13	0.08	-0.01	0.28
P41	0.87	-0.01	0.62	0.12	0.47

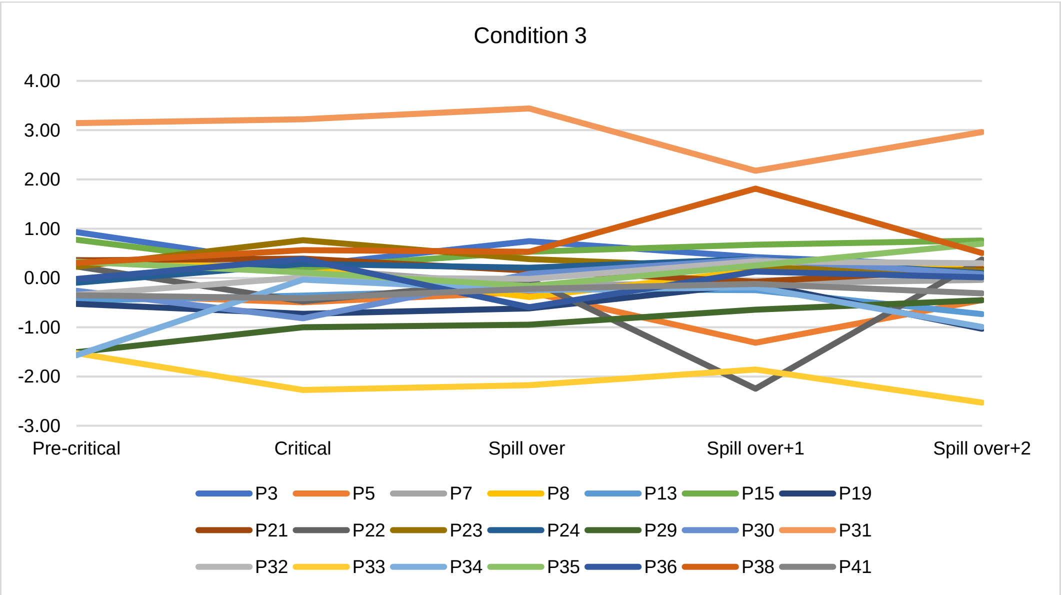
Descriptive Statistics for Condition 2 Individual Score Difference by Segment



Descriptive Statistics for Condition 3 Individual Score Difference by Segment

ID	Pre-critical	Critical	Spill over	Spill over+1	Spill over+2
P3	0.93	0.25	0.75	0.42	0.24
P5	-0.33	-0.49	-0.28	-1.31	-0.45
P7	0.37	0.20	-0.15	-0.11	-0.04
P8	0.32	0.20	-0.39	0.15	0.21
P13	-0.46	-0.36	-0.23	-0.25	-0.73
P15	0.78	0.20	0.53	0.67	0.76
P19	-0.53	-0.73	-0.62	-0.12	-1.03
P21	0.35	0.39	0.15	-0.08	0.18
P22	0.24	-0.47	-0.03	-2.25	0.38
P23	0.23	0.77	0.38	0.21	0.15
P24	-0.10	0.29	0.21	0.39	0.09
P29	-1.51	-1.00	-0.95	-0.64	-0.45
P30	-0.26	-0.82	0.08	0.38	0.08
P31	3.14	3.22	3.44	2.17	2.96
P32	-0.35	0.01	-0.02	0.34	0.30
P33	-1.53	-2.27	-2.18	-1.86	-2.53
P34	-1.57	-0.03	-0.25	-0.19	-1.00
P35	0.33	0.11	-0.16	0.25	0.69
P36	-0.02	0.38	-0.59	0.13	0.01
P38	0.30	0.57	0.53	1.81	0.51
P41	-0.35	-0.41	-0.22	-0.12	-0.32

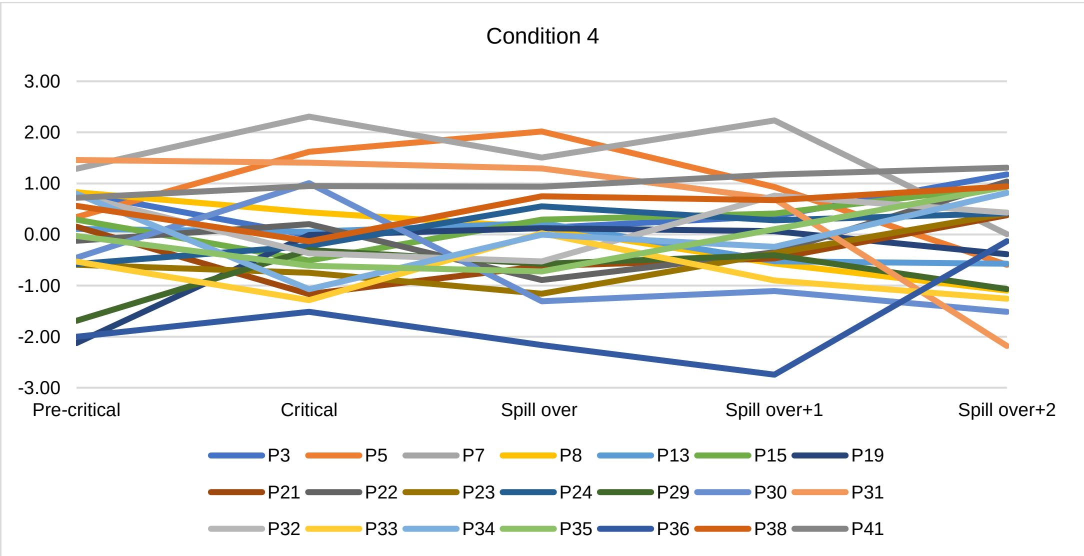
Descriptive Statistics for Condition 3 Individual Score Difference by Segment



Descriptive Statistics for Condition 4 Individual Score Difference by Segment

ID	Pre-critical	Critical	Spill over	Spill over+1	Spill over+2
P3	0.82	-0.02	0.12	0.36	1.18
P5	0.34	1.62	2.02	0.93	-0.59
P7	1.28	2.31	1.51	2.23	0.01
P8	0.83	0.43	0.14	-0.57	-1.10
P13	0.11	0.05	0.22	-0.53	-0.57
P15	0.29	-0.51	0.29	0.41	0.96
P19	-2.13	-0.01	0.12	0.06	-0.39
P21	0.15	-1.17	-0.62	-0.47	0.38
P22	-0.12	0.20	-0.89	-0.35	1.04
P23	-0.59	-0.75	-1.16	-0.37	0.41
P24	-0.58	-0.22	0.55	0.28	0.42
P29	-1.69	-0.30	-0.57	-0.40	-1.07
P30	-0.46	1.00	-1.31	-1.11	-1.51
P31	1.46	1.40	1.29	0.70	-2.18
P32	0.77	-0.37	-0.53	0.76	0.43
P33	-0.53	-1.29	0.02	-0.90	-1.26
P34	0.80	-1.07	-0.01	-0.25	0.81
P35	-0.03	-0.61	-0.72	0.10	0.93
P36	-2.00	-1.51	-2.17	-2.74	-0.13
P38	0.56	-0.13	0.75	0.67	0.94
P41	0.72	0.95	0.94	1.17	1.31

Descriptive Statistics for Condition 4 Individual Score Difference by Segment



Bibliography

Adila, W., & Ma'mun, T. N. (2020). Knowledge and use of grammar among Indonesian second language learners of Arabic: Focus on grammatical gender agreement. *Universal Journal of Educational Research*, *8*(2), 709–722. https://doi.org/10.13189/ujer.2020.080245

Alarcón, I. (2020). Early and late bilingual processing of Spanish gender, morphology and gender congruency. *Borealis – An International Journal of Hispanic Linguistics*, *9*(2), 175–208. https://doi.org/10.7557/1.9.2.5523

Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism*, *14*(3), 332–350. https://doi.org/10.1017/S1366728910000222

Amengual, M. (2015). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, *37*(5), 1221–1251. https://doi.org/10.1017/S0142716415000557

Amengual, M. (2016a). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology*, *7*(APR). https://doi.org/10.3389/fpsyg.2016.00617

Amengual, M. (2016b). The perception and production of language-specific mid-vowel contrasts: Shifting the focus to the bilingual individual in early language input conditions. *International Journal of Bilingualism*, *20*(2), 133–152. https://doi.org/10.1177/1367006914544988

Amengual, M. (2016c). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, *37*(5), 1221–1251. https://doi.org/10.1017/S0142716415000557

Antonicelli, G., & Rastelli, S. (2022). Event-related potentials in the study of L2 sentence processing: A scoping review of the decade 2010-2020. *Language Acquisition*, *00*(00), 1–38. https://doi.org/10.1080/10489223.2022.2141633

Antoniou, M., Ettlinger, M., & Wong, P. C. M. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLOS ONE*, *11*(7), 1–20. https://doi.org/10.1371/journal.pone.0158812

Arnaud, P. J. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. *In Vocabulary and Applied Linguistics*, (pp. 133-145).

Avery, N., & Marsden, E. (2019). A meta-analysis of sensitivity to grammatical information during self-paced reading: Towards a framework of reference for reading time effect sizes. *Studies in Second Language Acquisition*, *41*(5), 1055–1087. https://doi.org/10.1017/S0272263119000196

Ayoun, D. (2019). *3. The L2 Acquisition of French Morphosyntax by Anglophone Learners: Refocusing on the Input. L2 Grammatical Representation and Processing*. 47–67. https://doi.org/doi:10.21832/9781788925358-005

Ayoun, D. (2010). Corpus data: Shedding the light on French grammatical gender … or not. *EUROSLA Yearbook*. https://doi.org/10.1075/eurosla.10.08ayo

Ayoun, D. (2022). *The Acquisition of Gender: Crosslinguistic perspectives*. *63*.

Ayoun, D., & Maranzana, S. (2022). The second language acquisition of grammatical gender and number in Italian. In Dalila Ayoun (Ed.), *The Acquisition of Gender Cross-Linguistic Perspectives* (pp. 97–126). John Benjamins.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bachman, L. F. (1985). Performance on Cloze Tests with Fixed-Ratio and Rational Deletions. *TESOL Quarterly*, *19*(3), 535. https://doi.org/10.2307/3586277

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science*, *4*, 417–423. https://doi.org/ttps://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189–208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), 136–140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A., Hitch, G., & Allen, R. (2020). A Multicomponent Model of Working Memory. In *Working Memory: State of the Science*.

https://doi.org/10.1093/oso/9780198842286.003.0002

Ball, M. J., Müller, N., & Munro, S. (2001). The acquisition of the Rhotic Consonants by Welsh-English bilingual children. *International Journal of Bilingualism*. https://doi.org/10.1177/13670069010050010401

Bañón, J. A., Fiorentino, R., & Gabriele, A. (2012). The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance. *Brain Research*. https://doi.org/10.1016/j.brainres.2012.03.057

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/0898929052880101

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*(June), 3–4. https://doi.org/10.3389/fpsyg.2013.00328

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. Retrieved from http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bellin, W. (1984). Welsh phonology in acquisition. In *In Ball, M. L, & Jones G. E. (Eds.), Welsh phonology*. Cardiff: University of Wales Press.

Bellin, W. (1988). The development of pronunciation. In *In M. J. Ball (Ed.), The use of Welsh: A contribution to sociolinguistics.* Clevedon: Multilingual Matters.

Bernabéu, P. (2022). Language and sensorimotor simulation in conceptual processing: Multilevel analysis and statistical power. *PhD*.

Binks, H. L. (2017). Investigating the Bilingual 'Catch-Up' in Welsh-English Bilingual Teenagers. *PQDT - Global*. Retrieved from http://cyber.usask.ca/login?url=https://search.proquest.com/docview/2490534762?accountid=14739&bdid=6474&_bd=9V1L3Uu1L4s%2BB7hnBFgb1JjKEkY%3D

Binks, H. L., & Thomas, E. M. (2019). Long-term outcomes for bilinguals in minority language contexts: Welsh-English teenagers' performance on measures of grammatical gender and plural morphology in Welsh. *Applied Psycholinguistics*, *40*(4), 1019–1049.

https://doi.org/10.1017/S0142716419000110

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual language profile: An easy-to-use instrument to assess bilingualism.*

Birdsong, D. (2006). Dominance, proficiency, and second language grammatical processing. *Applied Psycholinguistics*, *27*(1), 46–49. https://doi.org/10.1017.S0142716406060036

Birdsong, D. (2016). Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In *Language domi- nance in bilinguals: Issues of measurement and operationalization* (In C. Silv, pp. 85–105). https://doi.org/https://doi.org/10.1017/CBO9781107375345.005

Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology*, *9*(MAR), 1–17. https://doi.org/10.3389/fpsyg.2018.00081

Black, M., Joanisse, M. F., & Rafat, Y. (2020). Language dominance modulates the perception of spanish approximants in late bilinguals. *Languages*, *5*(1), 1–17. https://doi.org/10.3390/languages5010007

Black, M., & Tararova, O. (2020). *Adult Acquisition of Grammatical Gender in Instructed L2 Spanish and the Role of Metacognition*.

Bonvin, A., Brugger, L., & Berthele, R. (2021). *Lexical measures as a proxy for bilingual language dominance ?* 1–29. https://doi.org/doi.org/10.1515/iral-2020-0093

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

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, 1–29. https://doi.org/10.7717/peerj.9414

Brown, J. D. (1980). Relative Merits of Four Methods for Scoring Cloze Tests. *The Modern Language Journal*, *64*(3), 311–317. https://doi.org/10.1111/j.1540-4781.1980.tb05198.x

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R.

Advances in Methods and Practices in Psychological Science, *4*(1). https://doi.org/10.1177/2515245920960351

Brown, V. A., Dillman-Hasso, N., Li, Z., Ray, L., Mamantov, E., Van Engen, K., & Strand, J. (2021). *Lipreading in Noise: Cross-modal Analysis of the Target-Masker Linguistic Similarity Hypothesis*. 1–31.

Buffington, J., Demos, A. P., & Morgan-Short, K. (2021). The Reliability and Validity of Procedural Memory Assessments Used in Ssecond Language Aacquisition Research. *Studies in Second Language Acquisition*, *43*(3), 635–662. https://doi.org/10.1017/S0272263121000127

Buffington, J., & Morgan-Short, K. (2018). Construct Validity of Procedural Memory Tasks Used in Adult-Learned Language. *The 40th Annual Meeting of the Cognitive Science Society*, (August), 1420–1425. Retrieved from https://www.researchgate.net/profile/Joshua-Buffington/publication/326689469_Construct_Validity_of_Procedural_Memory_Tasks_Used_in_Adult-Learned_Language/links/5b6ca9bc92851ca65053de01/Construct-Validity-of-Procedural-Memory-Tasks-Used-in-Adult-Learned-La

Bull, R., Espy, K. A., & Senn, T. E. (2004). A comparison of performance on the Towers of London and Hanoi in young children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(4), 743–754. https://doi.org/10.1111/j.1469-7610.2004.00268.x

Caffarra, S., Janssen, N., & Barber, H. A. (2014). Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. *Neuropsychologia*, *63*, 124–134. https://doi.org/10.1016/j.neuropsychologia.2014.08.016

Carroll, S. (1989). Second-Language Acquisition and the Computational Paradigm. *Language Learning*. https://doi.org/10.1111/j.1467-1770.1989.tb00902.x

Carstens, V. (2005). Agree and EPP in Bantu. *Natural Language and Linguistic Theory*, *23*(2), 219–279. https://doi.org/10.1007/s11049-004-0996-6

Carstens, V. (2010). *Implications of grammatical gender for the theory of uninterpretable features*. (2007), 31–58. https://doi.org/10.1075/lfab.3.03car

Chapelle, C. A., & Abraham, R. G. (1990). Cloze method: What difference does it make? *Language Testing*, *7*(2), 121–146. https://doi.org/10.1177/026553229000700201

Cho, J., & Slabakova, R. (2014). Interpreting definiteness in a second language without articles: The case of L2 Russian. *Second Language Research*, *30*(2), 159–190. https://doi.org/10.1177/0267658313509647

Chondrogianni, V. (2008). *Comparing child and adult L2 acquisition of the Greek DP: Effects of age and construction*. 97–142. https://doi.org/10.1075/lald.46.07cho

Chondrogianni, V. (2024). Gender and number agreement. *The Routledge Handbook of Second Language Acquisition, Morphosyntax, and Semantics*, 154–166. https://doi.org/10.4324/9781003412373-15

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some Notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin and Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

Corbett, G. G. (2014). *The Expression of Gender* (Volume 6 i). Berlin/Boston: De Gruyter Mouton.

Cornips, L., & Hulk, A. (2008). Factors of success and failure in the acquisition of grammatical gender in Dutch. *Second Language Research*, *24*(3), 267–295. https://doi.org/10.1177/0267658308090182

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for nonnative speakers’ sensitivity to agreement in sentence processing. *Applied Psycholinguistics*, *34*(3), 615–646. https://doi.org/10.1017/S0142716411000890

Council, B. (2023). British Council on the lookout for Welsh speakers to teach 7,000 miles from home. Retrieved from https://wales.britishcouncil.org/en/about/press/british-council-lookout-welsh-speakers-teach-7000-miles-home-programme-send-welsh

Cowan, N. (2005). *Working Memory Capacity Limits in a Theoretical Context*. *Human Lear*, 155–175.

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

Cunnings, I. (2017a). Interference in Native and Non-Native Sentence Processing. *Bilingualism*. https://doi.org/10.1017/S1366728916001243

Cunnings, I. (2017b). Parsing and Working memory in bilingual sentence Processing. *Bilingualism*. https://doi.org/10.1017/S1366728916000675

Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In *Advancing quantitative methods in second language research* (In L. Plon, pp. 159–191). Retrieved from https://shorturl.at/kvLT5

Daneman, M., & Carpenter, Patricia, A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450. https://doi.org/https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, *3*(4), 422–433. https://doi.org/10.3758/BF03214546

Daskalaki, E., Blom, E., Chondrogianni, V., & Paradis, J. (2020). Effects of parental input quality in child heritage language acquisition. *Journal of Child Language*, *47*(4), 709–736. https://doi.org/10.1017/S0305000919000850

Davidson, D. J., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience*, *21*(3), 433–446. https://doi.org/10.1162/jocn.2008.21031

Davies, J. (2014). *The Welsh Langauge: A History*. University of Wales Press.

Dekydtspotter, L., Schwartz, B. D., & Sprouse, R. a. (2006). The Comparative Fallacy in L2 Processing Research. *Proceesings of the 8th Generative Approaches to Second Language Acquisition Conference*, (Gasla), 33–40.

Dewaele, J. M., & Véronique, D. (2000). Relating gender errors to morphosyntax and lexicon in advanced French interlanguage. *Studia Linguistica*, Vol. 54, pp. 212–224. https://doi.org/10.1111/1467-9582.00061

Dokić, R., Koso-Drljević, M., & Apo, N. (2018). Working memory span tasks: Group administration and omitting accuracy criterion do not change metric characteristics. *PLoS ONE*, *13*(10), 1–22. https://doi.org/10.1371/journal.pone.0205169

Dong, Z. R., Han, C., Hestvik, A., & Hermon, G. (2022). L2 processing of filled gaps: Non-native brain activity not modulated by proficiency and working memory. *Linguistic Approaches to Bilingualism*, (March). Retrieved from https://www.jbe-platform.com/content/journals/10.1075/lab.20058.don

Dowens, M. G., & Carreiras, M. (2006). The shallow structure hypothesis of second language sentence processing: What is restricted and why? *Applied Psycholinguistics*, *27*(1), 49–52. https://doi.org/10.1017/S0142716406230030

Dowens, M. G., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish - Evidence from Event Related Potentials. *Neuropsychologia*, *49*(7), 1651–1659. https://doi.org/10.1016/j.neuropsychologia.2011.02.034

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887. https://doi.org/10.1162/jocn.2009.21304

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale. *Bilingualism*, *12*(3), 273–289. https://doi.org/10.1017/S1366728909990113

Ellis, N. C., O'Dochartaigh, C., Hicks, W., Morgan, M., & Laporte, N. (2001). Cronfa Electroneg o Gymraeg (CEG): A 1 million word lexical database and frequency count for Welsh. Retrieved April 21, 2024, from https://www.bangor.ac.uk/canolfanbedwyr/ceg.php.en

Esfandiari, L., Nilipour, R., Nejati, V., Maftoon, P., & Khosrowabadi, R. (2020). Research paper: An event-related potential study of second language semantic and syntactic processing: Evidence from the declarative/procedural model. *Basic and Clinical Neuroscience*, *11*(6), 841–854. https://doi.org/10.32598/BCN.11.6.2401.1

Ettlinger, M., Bradlow, A. R., & Wong, P. C. M. (2014). Variability in the learning of complex morphophonology. *Applied Psycholinguistics*, *35*(4), 807–831. https://doi.org/10.1017/S0142716412000586

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003). The processing of ambiguous sentences by first and second language learners of English. In *Applied Psycholinguistics* (Vol. 24). https://doi.org/10.1017/S0142716403000237

Fernandez, L. B., Bothe, R., & Allen, S. E. M. (2021). The role of L1 reading direction on L2 perceptual span: An eye-tracking study investigating Hindi and Urdu speakers. *Second Language Research*. https://doi.org/10.1177/02676583211049742

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Fhlannchadha, S. N., & Hickey, T. M. (2021). Where are the goalposts? Generational change in the use of grammatical gender in irish. *Languages*, *6*(1), 1–23. https://doi.org/10.3390/languages6010033

Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics,* 55-74.

Foerde, K., Knowlton, B. J., & Poldrack, R. A. (2006). Modulation of competing memory systems by distraction. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(31), 11778–11783. https://doi.org/10.1073/pnas.0602659103

Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, *32*(1), 187–220. https://doi.org/10.1017/S0142716410000342

Foucart, A. (2008). Grammatical Gender Processing in French as a First and a Second Language. *TIPA. Travaux Interdisciplinaires Sur La Parole et Le Langage*, (27). https://doi.org/10.4000/tipa.342

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity. *Bilingualism*. https://doi.org/10.1017/S136672891000012X

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226–248. https://doi.org/10.1016/j.jml.2011.07.007

Frenck-Mestre, C. (2002). 9 An on-line look at sentence processing in the second language. *Advances in Psychology*, *134*(C), 217–236. https://doi.org/10.1016/S0166-4115(02)80012-7

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, *21*(2), 175–198.

https://doi.org/10.1191/0267658305sr257oa

Frenda, A. (2011). Gender in Insular Celtic : A functionalist account of variation and change in Irish and Welsh. *Linguistics*, (September).

Fricke, M., Zirnstein, M., Navarro-Torres, C., & Kroll, J. F. (2019). Bilingualism reveals fundamental variation in language processing. *Bilingualism*, *22*(1), 200–207. https://doi.org/10.1017/S1366728918000482

Gabriele, A., Bañón, J. A., Hoffman, L., Covey, L., Rossomondo, A., & Fiorentino, R. (2021). Supplemental Material for Examining Variability in the Processing of Agreement in Novice Learners: Evidence From Event-Related Potentials. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* (Vol. 47). https://doi.org/10.1037/xlm0000983.supp

Gaillard, S., & Tremblay, A. (2016). Linguistic Proficiency Assessment in Second Language Acquisition Research: The Elicited Imitation Task. *Language Learning*, Vol. 66, pp. 419–447. https://doi.org/10.1111/lang.12157

Garraffa, M., Obregon, M., & Sorace, A. (2017). Linguistic and cognitive effects of bilingualism with regional minority languages: A study of Sardinian-Italian adult speakers. *Frontiers in Psychology*, *8*(NOV), 1–11. https://doi.org/10.3389/fpsyg.2017.01907

Gastmann, F., Poarch, G., & Schimike, S. (2022). Processing of relative clauses in L2 learners: No influence of cognate status, but evidence of revision effects. *EuroSLA*. Fribourg.

Gathercole, V. C. M., Thomas, E. M., & Laporte, N. (2001). The acquisition of grammatical gender in Welsh. *Journal of Celtic Language Learning*, 53–87.

Gathercole, V. C. M. (2002). Chapter 10: Monolingual and Bilingual Acquisition: Learning Different Treatments of that-trace Phenomena in English and Spanish. In *Language and Literacy in Bilingual Children*. https://doi.org/10.21832/9781853595721-011

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism*, *12*(2), 213–237. https://doi.org/10.1017/S1366728909004015

Gathercole, V. C. M., & Thomas, E. M. (2005). Minority Language Survival : Input Factors Influencing the Acquisition of Welsh. *Proceedings of ISB4*, 852–874. Retrieved from http://www.cascadilla.com/isb4.html

Gathercole, V. C. M., Thomas, E. M., & Hughes, E. (2008). *Designing a Normed*

Receptive Vocabulary Test for Bilingual Populations : A Model from Welsh Vocabulary Test for Bilingual Populations : A Model from Welsh. *0050*. https://doi.org/10.1080/13670050802149283

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). *Assessing Language Dominance with the Bilingual Language Profile.* 208–225. https://doi.org/10.21832/9781783092291-014

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment.*, 105–129.

Granfeldt, J., Schlyter, S., & Kihlstedt, M. (2007). French as cL2, 2L1, and L1 in preschool children. *PERLES: Petites Études Romanes de Lund*, (21), 6–43.

Granfeldt, J. (2018). The development of gender in simultaneous and successive bilingual acquisition of French - Evidence for AOA and input effects. *Bilingualism*, *21*(4), 674–693. https://doi.org/10.1017/S1366728916001140

Griffiths, B., & Jones, D. G. (2014). *The Welsh Academy English-Welsh Dictionary*. University of Wales Press.

Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*. https://doi.org/10.1177/0267658312437990

Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition*, *29*(3), 503–511. https://doi.org/10.3758/BF03196401

Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, *28*(6), 715–728. https://doi.org/10.1023/A:1023277213129

Hammond, M. (2016). Predicting the gender of Welsh nouns. *Corpus Linguistics and Linguistic Theory*, (1), 221–261. https://doi.org/10.1515/cllt-2015-0001

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(7), 1487–1492. https://doi.org/10.1073/pnas.1713975115

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch. *Language Learning*, *59*(1), 73–112. https://doi.org/10.1111/j.1467-9922.2009.00501.x

Hodges, R. (2024). Defiance within the decline? Revisiting new Welsh speakers' language journeys. *Journal of Multilingual and Multicultural Development*, *45*(2), 306–322. https://doi.org/10.1080/01434632.2021.1880416

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*. https://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*. https://doi.org/10.1177/0267658312461803

Hopp, H. (2014). Working Memory Effects in the L2 Processing of Ambiguous Relative Clauses. *Language Acquisition*. https://doi.org/10.1080/10489223.2014.892943

Hopp, H. (2016a). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*. https://doi.org/10.1177/0267658315624960

Hopp, H. (2016b). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, *37*(5), 1253–1280. https://doi.org/10.1017/S0142716415000569

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–27. https://doi.org/10.11431/secondlanguage.17.0_5

Hopp, H. (2022). Second language sentence processing. *Annual Review of Linguistics*, *8*, 235–256. https://doi.org/doi.org/10.1146/annurev-linguistics-030821- 054113

Howard, J. H., & Howard, D. V. (1997). Age differences in implicit learning of higher order dependencies in serial patterns. *Psychology and Aging*, *12*(4), 634–656. https://doi.org/10.1037/0882-7974.12.4.634

Jegerski, J. (2014). Self-paced reading. *Research Methods in Second Language Psycholinguistics*, (In J. Jegerski & B. VanPatten (Eds.)), 20–49.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603–634. https://doi.org/10.1017/s0142716404001298

Jiang, N. (2007). Selective integration of Linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1–33. https://doi.org/10.1111/j.1467-9922.2007.00397.x

Jiménez Catalán, R., & Moreno Espinosa, S. (2005). Promoting English vocabulary research in primary and secondary education: test review and test selection criteria. *ES: Revista de Filología Inglesa*, *25*(26), 171–188.

Jones, M. C. (1998). Language obsolescence and revitalization: Linguistic change in two sociolinguisticaliy contrasting Welsh communities. *Oxford: Clarendon Press.*

Jonz, J. (1976). Improving on the basic egg: The Multiple-Choice Cloze. *Language Learning*, 255–265. https://doi.org/10.1111/j.1467-1770.1976.tb00276.x

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, *44*(2), 137–166. https://doi.org/10.1017/S0261444810000509

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238. https://doi.org/10.1037/0096-3445.111.2.228

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment*, *24*(1), 46–53. https://doi.org/10.1037/a0025174

Kane, M. J., Tuholski, S. W., Hambrick, D. Z., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Karmiloff-Smith, A. (1979). *A functional approach to child language. A study of determiners and reference*.

Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*. https://doi.org/10.1111/j.1467-9922.2009.00516.x

Keating, G. D. (2010). *The effects of linear distance and working memory on the processing of gender agreement in Spanish*. 113–134. https://doi.org/10.1075/lald.53.05kea

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, *37*(1), 1–32. https://doi.org/10.1017/S0272263114000187

Kim, S. K., & Webb, S. (2022). Individual Difference Factors for Second Language Vocabulary. *The Routledge Handbook of Second Language Acquisition and Individual Differences*, 282–293. https://doi.org/10.4324/9781003270546-24

King, G. (2016). Modern Welsh: A Comprehensive Grammar. In *Routledge* (3rd editio). London/New York: Routledge.

Klassen, G., Ferreira, A., & Schwieter, J. W. (2021). The role of immersion learning in the acquisition and processing of L2 gender agreement. *Applied Linguistics Review*, *14*(2), 391–413. https://doi.org/10.1515/applirev-2020-0038

Knight, D., Morris, S., Tovey-Walsh, B., Fitzpatrick, T., & Anthony, L. (2020). *Yr Amliadur: Rhestri Amlder ar gyfer Cymraeg Cyfoes Frequency Lists for Contemporary Welsh*. Retrieved from http://doi.org/10.17035/d.2020.0120164107

Knight, D., Morris, S., Fitzpatrick, T., Morris, J., Lovell, A., Tovey-Walsh, B., … Muralidaran, V. (2020). CorCenCC: Corpws Cenedlaethol Cymraeg Cyfoes – the National Corpus of Contemporary Welsh. https://doi.org/http://doi.org/10.17035/d.2020.0119878310

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 699–710. https://doi.org/10.1037//0278-7393.21.3.699

Koeneman, O., & Zeijlstra, H. (2017). *Introducing Syntax*. Cambridge: Cambridge University Press.

Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, *22*(5), 564–582. https://doi.org/10.1177/1367006916654355

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research Forum*, *25*(2), 173–227.

Larrabee, G. J. (2009). Malingering scales for the continuous recognition memory test and the continuous visual memory test. *Clinical Neuropsychologist*, *23*(1), 167–180. https://doi.org/10.1080/13854040801968443

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202–226. https://doi.org/10.1191/0265532204lt277oa

Lemhöfer, K., Schriefers, H., & Indefrey, P. (2014). Idiosyncratic Grammars: Syntactic Processing in Second Language Comprehension Uses Subjective Feature Representations. *Journal of Cognitive Neuroscience*, *26*(7), 1428–1444. https://doi.org/doi:10.1162/jocn_a_00609

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means (1.7.2). Retrieved from https://cran.r-project.org/package=emmeans

Linck, J. A., & Cunnings, I. (2015). The Utility and Application of Mixed-Effects Models in Second Language Research. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*(August), 1–16. https://doi.org/10.3389/fpsyg.2015.01171

Logie, R., Camos, V., & Cowan, N. (2020). Working Memory: State of the science. *Working Memory: State of the Science*, 1–448. https://doi.org/10.1093/oso/9780198842286.001.0001

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621. https://doi.org/10.1080/20445911.2013.795574

Macwhinney, B. (1978). The Acquisition of Morphophonology. *Monographs of the Society for Research in Child Development*.

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: An introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology, 6*(JAN), 1–7. https://doi.org/10.3389/fpsyg.2015.00002

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language

Profiles in Bilinguals and Multilinguals. Journal of Speech Language and Hearing Research, 50(4), 940. https://doi. *Journal of Speech Language and Hearing Research*, *50*(4), 940. Retrieved from http://jslhr.pubs.asha.org/article.aspx?doi=10.1044/1092-4388(2007/067)

Marinis, T. (2010). *Chapter 7. Using on-line processing methods in language acquisition research*. 139–162. https://doi.org/10.1075/lllt.27.09mar

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861–904. https://doi.org/10.1017/S0142716418000036

Mathôt, S., & March, J. (2022). Conducting Linguistic Experiments Online With OpenSesame and OSWeb. *Language Learning*, *72*(4), 1017–1048. https://doi.org/10.1111/lang.12509

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*(January), 104092. https://doi.org/10.1016/j.jml.2020.104092

Meurig Evans, H., & Thomas, O. B. (2021). *Y Geiriadur Mawr: The Complete Welsh-English, English-Welsh Dictionary*. Gwasg Gomer.

Mittendorf, I., & Sadler, L. (2006). A Treatment of WElsh Initial Mutation. *Ingo Mittendorf and Louisa Sadler University of Essex*.

Monnier, C., Boiché, J., Armandon, P., Baudoin, S., & Bellocchi, S. (2022). Is bilingualism associated with better working memory capacity? A meta-analysis. *International Journal of Bilingual Education and Bilingualism*, *25*(6), 2229–2255. https://doi.org/10.1080/13670050.2021.1908220

Montrul, S. (2016). Dominance and proficiency in early and late bilingualism. *Language Dominance in Bilinguals: Issues of Measurement and Operationalization, eds C. Sil*, 15–35. https://doi.org/10.1017/cbo9781107375345.002

Montrul, S. (2011). Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition*, *33*(2), 163–192. https://doi.org/10.1017/S0272263110000720

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, *58*(3), 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x

Montrul, S., & Potowski, K. (2007). Command of gender agreement in school-age Spanish-English Bilingual Children. *International Journal of Bilingualism*. https://doi.org/10.1177/13670069070110030301

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism*, *17*(1), 56–72. https://doi.org/10.1017/S1366728912000715

Morgan-Short, K., Hamrick, P., & Ullman, M. T. (2022). Declarative and Procedural Memory as Predictors of Second Language Development. In *The Routledge Handbook of Second Language Acquisition and Individual Differences.* (pp. 67–81). https://doi.org/https://doi.org/10.4324/9781003270546

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154–193. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Morris, J. (2014). The influence of social factors on minority language engagement amongst young people: An investigation of Welsh-English bilinguals in North Wales. *International Journal of the Sociology of Language*, *2014*(230), 65–89. https://doi.org/10.1515/ijsl-2014-0027

Najjari, R., & Mohammadi, M. (2017). The Development of Reading and Operation Span Tasks in Persian as Measures of Working Memory Capacity for Iranian EFL Learners. *Journal of Teaching Language Skills (JTLS)*, *36*(2), 129–162. Retrieved from https://tesl.shirazu.ac.ir/article_4335_d11b063d293b9b4b8d41cb728e086587.pdf

Neubauer, K., & Clahsen, H. (2009). Decomposition of inflected words in a second language: An experimental study of German participles. In *Studies in Second Language Acquisition* (Vol. 31). https://doi.org/10.1017/S0272263109090354

Newcombe, L. P. (2002). “A tough hill to climb alone”-Welsh learners speak. *Hong Kong Journal of Applied Linguistics*, *7*(2), 39–56.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1–32. https://doi.org/10.1016/0010-0285(87)90002-8

Oller, J. W. J. (1972). Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language. *The Modern Language Journal*, *56*(3), 151–158. https://doi.org/10.1111/j.1540-4781.1972.tb05035.x

Olson, D. J. (2017). Bilingual language switching costs in auditory comprehension. *Language, Cognition and Neuroscience*, *32*(4), 494–513. https://doi.org/10.1080/23273798.2016.1250927

Olson, D. J. (2023). Measuring bilingual language dominance: An examination of the reliability of the Bilingual Language Profile. *Language Testing*, 1–27. https://doi.org/.o0r.g1/107.711/0727/6052563552322212113399162

Oppenheim, G. M., Griffin, Z., Peña, E. D., & Bedore, L. M. (2020). Longitudinal Evidence for Simultaneous Bilingual Language Development With Shifting Language Dominance, and How to Explain It. *Language Learning*, *70*(June 2020), 20–44. https://doi.org/10.1111/lang.12398

Pereira Soares, S. M., Kubota, M., Rossi, E., & Rothman, J. (2021). Determinants of bilingualism predict dynamic changes in resting state EEG oscillations. *Brain and Language*, *223*(July). https://doi.org/10.1016/j.bandl.2021.105030

Pérez-Pereira, M. (1991). The acquisition of gender: What Spanish children tell us. *Journal of Child Language*. https://doi.org/10.1017/S0305000900011259

Perpiñán, S. (2017). Catalan-Spanish bilingualism continuum. *Linguistic Approaches to Bilingualism*, *7*(5), 477–513. https://doi.org/10.1075/lab.15004.per

Phillips, J. D. (2008). Effaith newidiadau diweddar ar hynodrwydd ieithyddol y Gymraeg. *Gwerddon*, 94–117. https://doi.org/10.61257/ztiq6569

Pignot-Shahov, V. (2012). *elal_LSWP_Vol_4_Pignot_Shahov*. *4*, 37–45.

Pizzuto, E., & Caselli, M. C. (1992). The acquisition of Italian morphology: implications for models of language development. *Journal of Child Language*, *3*(19), 491–557.

Poort, E. D., & Rodd, J. M. (2022). Cross-lingual priming of cognates and interlingual homographs from L2 to L1. *Glossa Psycholinguistics*, *1*(1), 1–33. https://doi.org/10.5070/g601147

Prentza, A., & Tsimpli, I. M. (2013). The interpretability of features in second language acquisition: Evidence from null and postverbal subjects in L2 English.

Journal of Greek Linguistics, *13*(2), 323–365. Retrieved from https://brill.com/view/journals/jgl/13/2/article-p323_7.xml

Prévost, P., & White, L. (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*(2), 103–133. https://doi.org/10.1191/026765800677556046

Quam, C., Wang, A., Todd Maddox, W., Golisch, K., & Lotto, A. (2018). Procedural-memory, working-memory, and declarative-memory skills are each associated with dimensional integration in sound-category learning. *Frontiers in Psychology*, *9*(OCT), 1–15. https://doi.org/10.3389/fpsyg.2018.01828

Quené, H., & Van Den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*(1–2), 103–121. https://doi.org/10.1016/j.specom.2004.02.004

Rattanasak, S., Pongpairoj, N., & Christianson, K. (2022). Effects of working memory capacity and distance-based complexity on agreement processing: a crosslinguistic competition account. *Applied Linguistics Review*, *0*(0), 1–26. https://doi.org/10.1515/applirev-2022-0035

Reichle, R. V., Tremblay, A., & Coughlin, C. (2016). Working memory capacity in L2 processing. *Probus*, *28*(1), 29–55. https://doi.org/10.1515/probus-2016-0003

Renaud, C. (2014). A processing investigation of the accessibility of the uninterpretable gender feature in L2 French and L2 Spanish adjective agreement. *Linguistic Approaches to Bilingualism*, *4*(2), 222–255. https://doi.org/10.1075/lab.4.2.04ren

Roberts, I. G. (2005). *Principles and parameters in a VSO language: a case study in Welsh*. Oxford: Oxford University Press.

Rodina, Y., Kupisch, T., Meir, N., Mitrofanova, N., Urek, O., & Westergaard, M. (2020). Internal and External Factors in Heritage Language Acquisition: Evidence From Heritage Russian in Israel, Germany, Norway, Latvia and the United Kingdom. *Frontiers in Education*, *5*(March), 1–17. https://doi.org/10.3389/feduc.2020.00020

Rodina, Y., & Westergaard, M. (2017). Grammatical gender in bilingual Norwegian-Russian acquisition: The role of input and transparency. *Bilingualism*, *20*(1), 197–214. https://doi.org/10.1017/S1366728915000668

Rodina, Y., & Westergaard, M. (2021). Grammatical Gender and Declension Class in Language Change: A Study of the Loss of Feminine Gender in Norwegian.

Journal of Germanic Linguistics, *33*(3), 235–263. https://doi.org/10.1017/S1470542719000217

Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in Heritage Brazilian Portuguese. *International Journal of Bilingualism*, *11*(4), 359–389. https://doi.org/10.1177/13670069070110040201

Rothman, J., Bayram, F., Deluca, V., Di Pisa, G., Duñabeitia, J. A., Gharibi, K., … Wulff, S. (2023). Monolingual comparative normativity in bilingualism research is out of control: Arguments and alternatives. *Applied Psycholinguistics*, *44*(3), 316–329. https://doi.org/10.1017/S0142716422000315

Rothman, J., Bayram, F., Deluca, V., González Alonso, J., Kubota, M., & Puig-Mayenco, E. (2023). Defining bilingualism as a continuum: Some tools and consequences for the study of bilingual mind and brain effects. In *Understanding Language and Cognition through Bilingualism* (pp. 38–67). John Benjamins.

Ruiz, S., Chen, X., Rebuschat, P., & Meurers, D. (2019). Measuring individual differences in cognitive abilities in the lab and on the web. *PLoS ONE*, *14*(12), 1–14. https://doi.org/10.1371/journal.pone.0226217

Sabourin, L. (2003). Grammatical gender and second language processing : an ERP study. *ProQuest Dissertations and Theses*, 185. Retrieved from https://shorturl.at/EKNTW

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430. https://doi.org/10.1177/0267658308090186

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621. https://doi.org/10.1006/jmla.1996.0032

Sagarra, N. (2019). Age Effects and Morphological Markedness in L2 Processing of Gender Agreement: insights from Eye-tracking. In *L2 Grammatical Representation and Processing: Theory and Practice* (Vol. 136, pp. 93–115). Retrieved from https://shorturl.at/cikCS

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022–2039. https://doi.org/10.1016/j.lingua.2010.02.004

Sarko, G. (2009). *L2 English article production by Arabic and French speakers*. (2004), 37–66. https://doi.org/10.1075/lald.49.06sar

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*(November 2018), 104038. https://doi.org/10.1016/j.jml.2019.104038

Seigneuric, A., Zagar, D., Meunier, F., & Spinelli, E. (2007). The relation between language and cognition in 3- to 9-year-olds: The acquisition of grammatical gender in French. *Journal of Experimental Child Psychology*, *3*(96), 229–246.

Serafini, E. J. (2017). Exploring the Dynamic Long-Term Interaction Between Cognitive and Psychosocial Resources in Adult Second Language Development at Varying Proficiency. *Modern Language Journal*, *101*(2), 369–390. https://doi.org/10.1111/modl.12400

Serafini, E. J., & Sanz, C. (2016). Evidence for the Decreasing Impact of Cognitive Ability on Second Language Development as Proficiency Increases. *Studies in Second Language Acquisition*, *38*(4), 607–646. https://doi.org/10.1017/S0272263115000327

Sharp, M. K. (2012). Morphosyntactic Complexity and Exposure in the Acquisition of Gender in Welsh. *Unpublished Doctoral Dissertation. Bangor University.*

Silva-Corvalán, C., & Treffers-Daller, J. (2016). *Language dominance in bilinguals: Issues of measurement and operationalization*. Cambridge University Press.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism*, *11*(2), 245–260. https://doi.org/10.1017/S1366728908003404

Simon, H. A. (1975). The Functional Equivalence of Problem Solving Skills. *Cognitive Psychology*, *7*, 268–288.

Slabakova, R., Leal, T., Dudley, A., & Stack, M. (2020). *Generative Second Language Acquisition* (A. Benati & J. W. Schwieter, Eds.). https://doi.org/10.1017/9781108762380

Solaimani, E., Myles, F., & Lawyer, L. (2023). Testing the Interpretability Hypothesis: Evidence from acceptability judgments of relative clauses by Persian and French learners of L2 English. *Second Language Research*, 1–24. https://doi.org/.o0r.g1/107.711/0727/607265768538233213116622783

Solís-Barroso, C., & Stefanich, S. (2019). Measuring language dominance in early spanish/english bilinguals. *Languages*, *4*(3), 1–22. https://doi.org/10.3390/languages4030062

Spinner, P., & Juffs, A. (2008). L2 grammatical gender in a complex morphological system: The case of German. *IRAL - International Review of Applied Linguistics in Language Teaching*, *46*(4), 315–348. https://doi.org/10.1515/IRAL.2008.014

Stefaniak, N., Baltazart, V., & Declercq, C. (2021). Processing Verb Meanings and the Declarative/Procedural Model: A Developmental Study. *Frontiers in Psychology*, *12*(September). https://doi.org/10.3389/fpsyg.2021.714523

Steinhauer, K. (2006). How dynamic is second language acquisition? *Applied Psycholinguistics*, *27*(1), 92–95. https://doi.org/10.1017.S0142716406060036

Swanson, K., & Dekydtspotter, L. (2022). A Full Parse or a Shallow Structure in L2 ? An ERP Study of Anaphora in Successive-Cyclic Wh-movement in L1-Mandarin/L2-English. *Proceedings of the 46th Annual Boston University Conference on Language Development*, (2018), 768–782.

Takano, K., Ito, M., Kobayashi, K., Sonobe, N., Kurosu, S., Mori, Y., … Niwa, S. I. (2002). Procedural memory in schizophrenia assessed using a mirror reading task. *Psychiatry Research*, *109*(3), 303–307. https://doi.org/10.1016/S0165-1781(02)00021-5

Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual Differences in the Real-Time Neural Dynamics of Language Comprehension. *Psychology of Learning and Motivation - Advances in Research and Theory*, *68*(August), 299–335. https://doi.org/10.1016/bs.plm.2018.08.007

Team, R. C. (2013). R: A Language and Environment for Statistical Computing.

Teschner, R. V, & Russell, W. M. (1984). *The gender patterns of Spanish nouns: An inverse dictionary-based analysis*. (1), 115–132.

Thomas, E. (2001). *Aspects of Gender Mutation in Welsh*. 457. Retrieved from http://e.bangor.ac.uk/4267/2/367312.pdf

Thomas, E. M., & Gathercole, V. C. M. (2005). *Minority Language Survival: Obsolescence or Survival for Welsh in the Face of English Dominance?* 2233–2257. Retrieved from http://hdl.handle.net/10242/40917

Thomas, E. M., & Gathercole, V. C. M. (2005). Minority Language Survival : Obsolescence or Survival for Welsh in the Face of English Dominance? *ISB4: Proceedings of the 4th International Symposium on Bilingualism*, 2233–2257.

Thomas, E. M., & Gathercole, V. C. M. (2007). Children's productive command of grammatical gender and mutation in Welsh: An alternative to rule-based learning. *First Language*, *27*(3), 251–278.

Thomas, E. M., & Mayr, R. (2010). Children's acquisition of Welsh in a bilingual setting: a psycholinguistic perspective. In *Welsh in the Twenty-First Century* (pp. 99-117.). University of Wales Press.

Thomas, E. M., Williams, N., Jones, L. A., Davies, S., & Binks, H. (2014). Acquiring complex structures under minority language conditions: Bilingual acquisition of plural morphology in Welsh. *Bilingualism*, *17*(3), 478–494. https://doi.org/10.1017/S1366728913000497

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263105050102

Trahan, D. E., & Larrabee, G. J. (1988). Continuous Visual Memory Test. *Odessa, FL: Psycho*.

Treffers-Daller, J. & Silva-Corvalán, C. (2016). Language dominance in bilinguals: Issues of measurement and operationalization. *Language Dominance in Bilinguals*. https://doi.org/10.1017/cbo9781107375345.014

Treffers-Daller, J. (2016). Language dominance: The construct, its measurement, and operationalization. *Language Dominance in Bilinguals*, 235–265. https://doi.org/10.1017/cbo9781107375345.012

Treffers-Daller, J.. (2019). What Defines Language Dominance in Bilinguals? *Annual Review of Linguistics*, *5*, 375–393. https://doi.org/10.1146/annurev-linguistics-011817-045554

Tremblay, A. (2007). Bridging the gap between theoretical linguistics and psycholinguistics in L2 phonology: Acquisition and processing of word stress by French Canadian L2 learners of English. *(Unpublished Doctoral Dissertation).*

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics*, *29*(4), 553–584. https://doi.org/10.1017/S0142716408080247

Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by french canadian listeners. *International Journal of Bilingualism*, *13*(1), 35–62. https://doi.org/10.1177/1367006909103528

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition*, *33*(3), 339–372. https://doi.org/10.1017/S0272263111000015

Tremblay, A., & Garrison, M. D. (2010). Cloze Tests: A Tool for Proficiency Assessment in Research on L2 French. *Second Language Research Forum*, 73–88.

Tsimpli, I. M. (2003). Clitics and Articles in L2 Greek. *In Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002), Somerville*(J. M. Liceras et al. (eds.)), 331–339.

Tsimpli, I. M., & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, *23*(2), 215–242.

Tsimpli, I. M., & Mastropavlou, M. (2007). Feature Interpretability in L2 Acquisition and SLI: Greek Clitics and Determiners. *The Role of Formal Features in Second Language Acquisition*, 142–183. https://doi.org/10.4324/9781315085340-6

Tucker, G. R., Lambert, W. E., & Rigault, A. A. (1997). The French speaker’s skill with grammatical gender: An example of rule-governed behavior. *De Gruyter*.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726. https://doi.org/10.1038/35094573

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice*, (2005), 141–178.

Ullman, M. T. (2006). The declarative/procedural model and the shallow structure hypothesis. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716406370038

Ullman, M. T. (2015). The Declarative/Procedural Model: A Neurobiological Model of Language Learning, Knowledge, and Use. In *Neurobiology of Language*. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

Ullman, M. T. (2020). The declarative/procedural model. In *Theories in Second Language Acquisition: An Introduction*. https://doi.org/https://doi.org/10.4324/9780429503986-7

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, *9*(2), 266–276. https://doi.org/10.1162/jocn.1997.9.2.266

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*(1), 39–65. https://doi.org/10.1177/0267658316675195

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. https://doi.org/https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Unsworth, S. (2008). Age and input in the acquisition of grammatical gender in Dutch. *Second Language Research*. https://doi.org/10.1177/0267658308090185

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism*, *16*(1), 86–110. https://doi.org/10.1017/S1366728912000284

Unsworth, S. (2014). *Comparing the role of input in bilingual acquisition across domains*. https://doi.org/10.1075/tilar.13.10uns

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, *35*(4), 765–805.

https://doi.org/10.1017/S0142716412000574

Unsworth, S., Chondrogianni, V., & Skarabela, B. (2018). Experiential measures can be used as a proxy for language dominance in bilingual language acquisition research. *Frontiers in Psychology*, *9*(OCT), 1–15. https://doi.org/10.3389/fpsyg.2018.01809

Unterrainer, J. M., Rahm, B., Kaller, C. P., Wild, P. S., Münzel, T., Blettner, M., … Beutel, M. E. (2019). Assessing Planning Ability Across the Adult Life Span in a Large Population-Representative Sample: Reliability Estimates and Normative Data for the Tower of London (TOL-F) Task. *Journal of the International Neuropsychological Society*, *25*(5), 520–529. https://doi.org/10.1017/S1355617718001248

Unterrainer, J. M., Rahm, B., Leonhart, R., Ruff, C. C., & Halsband, U. (2003). The Tower of London: The impact of instructions, cueing, and learning on planning abilities. *Cognitive Brain Research*, *17*(3), 675–683. https://doi.org/10.1016/S0926-6410(03)00191-5

Van der Velde, M. (2003). Déterminants et pronoms en néerlandais et en français: syntaxe et acquisition [Determiners and pronouns in Dutch and French: syntax and acquisition]. *Unpublished Doctoral Dissertation, University of Paris*.

Van der Velde, M. (2004). L'acquisition des déterminants en L1: une étude comparative entre le français et le néerlandais [The acquisition of L1 determiners: a comparative study of French and Dutch]. *Acquisition En Interaction En Langue Etrangère*, *21*, 9–46.

Vicente, M., Calandruccio, L., Miller, M. K., Browning, J. M., Oleson, J., & Leibold, L. J. (2019). Language proficiency and dominance considerations when working with Spanish–English bilingual adults. *American Journal of Audiology*, *28*(3), 724–729. https://doi.org/10.1044/2019_AJA-19-0028

Watkins. (1993). Welsh. In *In M. 1. Ball, & 1. Fife (Eds.), The Celtic languages.* (pp. 289–348). London: Routledge.

Welsh, M. C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain and Cognition*, *41*(2), 231–242. https://doi.org/10.1006/brcg.1999.1123

Wen, Z. (2016). Working memory and second language learning. In *International Journal of Applied Linguistics* (Vol. 22). https://doi.org/10.1111/j.1473-4192.2011.00290.x

Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. 1–22. Retrieved from http://arxiv.org/abs/1308.5499

Winter, B. (2020). Statistics for Linguists: An Introduction Using R. In *Statistics for Linguists: An Introduction Using R*. https://doi.org/10.4324/9781315165547

Winter, W. E., Broman, M., Rose, A. L., & Reber, A. S. (2001). The assessment of cognitive procedural learning in amnesia: Why the Tower of Hanoi has fallen down. *Brain and Cognition*, *45*(1), 79–96. https://doi.org/10.1006/brcg.2000.1257

Zalbidea, J. (2017). ‘One Task Fits All’? The Roles of Task Complexity, Modality, and Working Memory Capacity in L2 Performance. *Modern Language Journal*, *101*(2), 335–352. https://doi.org/10.1111/modl.12389