

Gender in Insular Celtic:  
A functionalist account of variation and change  
in Irish and Welsh

A thesis submitted to the University of Dublin  
for the degree of Doctor of Philosophy  
Supervisors: Professor John I. Saeed and Dr. Brian Nolan

Alessio S. Frenda  
School of Linguistic, Speech and Communication Sciences

September 2011



# Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

---

September 16, 2011

Alessio S. Frenda



# Summary

In this study, I investigate linguistic variation and change in a situation of language contact and obsolescence; specifically, I look at grammatical gender in Insular Celtic. The analysis of this linguistic phenomenon is conducted using data sets pertaining to spoken Irish and Welsh, taken as the two main representatives of this language group. The data set for each language includes two main components, relative to older and contemporary varieties, respectively. These data were in part drawn from existing corpora and in part collected in the course of this study.

The sociolinguistic status of these languages is taken into account; background information is provided and the concepts of language contact, obsolescence and convergence are presented and discussed.

The two main goals of this study are (i) to characterize gender agreement in Insular Celtic, using the theoretical framework provided by Functional Discourse Grammar (FDG) and (ii) to provide a quantitative analysis of any observed pattern of variation in order to achieve a better understanding of the system and contribute to the ongoing debate on grammatical gender.

The main finding of my analysis is that whereas conservative varieties exhibit a fairly consistent agreement system across the board, i.e. across agreement targets and grammatical genders, contemporary varieties do not; in the latter, grammatical agreement is still fairly consistent between the article and the noun but significantly less frequent with other agreement targets such as the attributive adjective and the anaphoric pronoun. These findings are problematic on two different levels: typologically, distinctive agreement patterns are well known in relation to anaphoric elements but not to targets lying within the noun phrase; from a theory-internal point of view, FDG does not recognize, within the noun phrase, any smaller constituent which contains only the determiner and the noun, so the observed patterns

run contrary to what we expect given a model of gender agreement based exclusively on the notion of feature copying across syntactic boundaries.

To overcome these difficulties, I propose a model of noun-phrase agreement for Insular Celtic which is consistent with FDG and is capable of accounting for the language-internal aspects of the observed linguistic change; in this model, feature copying operates alongside a different mechanism which is responsible for the peculiar way in which agreement between the article and the noun is marked in Insular Celtic. Then, looking at anaphoric agreement in particular, I argue for the necessity that the current FDG theory be expanded in order to provide a better account of how different types of information capable of influencing pronominal agreement forms are stored and accessed.

Finally, I argue that the observed variation in gender agreement is to be interpreted as the result of a process of resemantization and address two related questions: what might have started this process, and why it has resulted in the overgeneralization of masculine agreement forms. Considering comparable cases, I argue that the resemantization of pronominal agreement may be explained as a result of the gradual demise of gender agreement marking within the noun phrase. As regards the latter question, the switch leading to the overgeneralization of masculine forms, rather than a logically equally possible switch in the opposite direction, is explained as a combination of two factors, language-internal (markedness) and contact-induced (convergence).

# Acknowledgements

I am greatly indebted to many people without whose support, advice or assistance this thesis could have not been completed.

Heartfelt thanks to my supervisors, Professor John Saeed and Dr Brian Nolan, for their expert guidance and invaluable advice, and for their time.

I am also very grateful to Eamonn Mullins in the Department of Statistics, for his advice and helpful comments (*grazie*); to Dr Anthony Harvey in the Royal Irish Academy, for his weekly Welsh classes (*diolch yn fawr*); to Professor Damian McManus and Dr Eoin Mac Cárthaigh, in the Department of Irish, for their time and expertise; to Arndt Wigger for making the Ros Muc data available to me in digital format; to Briony Williams for pointing me to the Bangor corpus of Welsh informal conversations and radio broadcasts; to Gwenllian Awbery for her bibliographic suggestions; and to Eoin Ó Céilleachair, for the many unpaid hours he spent transcribing Irish interviews (*go raibh maith agat*).

Many thanks are also due to Enlli Môn Thomas, Gearóid Ó Donnchadha, Pauline Welby and Laura Innocenti for reading parts of my thesis and for their helpful comments on it; to Professor Dalila Ayoun, Professor Frans Plank, Gunther de Vogelaer, Lien De Vos and Dave Sayers, for sharing with me their ideas, manuscripts or yet unpublished papers. I would also like to thank Dr Jeffrey Kallen and Professor Ricardo Mairal for their recommendations and suggestions.

Many useful comments and suggestions were made during seminars and presentations; I wish to thank the organizers and participants of the Fifth Celtic Linguistics Conference (Wales, 2007), the Summer School in Functional Linguistics (Copenhagen, 2008), the Workshop on The Diachrony of Classification Systems (Wassenaar, 2009), the 19th International Conference on Historical Linguistics (Nijmegen, 2009), and the research student seminar

at the Centre for Language and Communication Studies at Trinity College Dublin (2010).

The completion of this study was made possible by a Postgraduate Scholarship from the Irish Research Council for the Humanities and Social Sciences and an Exchange Scholarship from the Irish Government, which I gratefully acknowledge.

It would be impossible to name and thank all the friends and colleagues who made life in Dublin so pleasant throughout these years with their company and friendship. The list would be too long and any omission unforgivable. Thank you all.

Finally, I wish to express my deepest gratitude to my family, first and foremost my parents, Enza and Lillo, and my sister, Daniela, for their constant encouragement and support.

# Contents

|   |             |
|---|-------------|
| <b>Summary</b>  | <b>v</b>    |
| <b>Acknowledgements</b>   | <b>vii</b>  |
| <b>Abbreviations</b>  | <b>xvii</b> |
| <b>1 Introduction</b>   | <b>1</b>    |
| 1.1 Goals and scope of this study . . . . .                                   | 1           |
| 1.2 The context of this study . . . . .                                       | 3           |
| 1.2.1 The complexity of grammatical gender systems . . . . .                  | 3           |
| 1.2.2 Variation in obsolescence: the case of Insular Celtic . . . . .         | 4           |
| 1.3 Irish and Welsh today . . . . .   | 10          |
| 1.4 A note on the phonological representation of Goidelic varieties . . . . . | 12          |
| 1.5 Summary and organization of this work . . . . .                           | 14          |
| <b>2 The functional perspective</b>   | <b>17</b>   |
| 2.1 Overview . . . . .  | 17          |
| 2.2 Functional approaches to the study of language . . . . .                  | 17          |
| 2.3 Functional Grammar and beyond: the emergence of FDG . . . . .             | 22          |
| 2.4 An outline of Functional Discourse Grammar . . . . .                      | 31          |
| 2.4.1 Components, operations and representations . . . . .                    | 31          |
| 2.4.2 Primitives . . . . .  | 35          |
| 2.4.3 An example . . . . .  | 37          |
| 2.4.4 Non-linear speech production . . . . .                                  | 42          |
| 2.5 Summary . . . . .   | 43          |
| <b>3 Methodology</b>  | <b>45</b>   |
| 3.1 Overview . . . . .  | 45          |

|          |  |           |
|----------|--|-----------|
| 3.2      | The corpora . . . . .  | 45        |
| 3.2.1    | The Welsh corpus . . . . .                                   | 47        |
| 3.2.2    | The Irish corpus . . . . .                                   | 49        |
| 3.3      | Analysis of the data . . . . .                               | 53        |
| 3.3.1    | Data retrieval and statistical testing . . . . .             | 53        |
| 3.3.2    | Proper names . . . . .                                       | 55        |
| 3.3.3    | Semantic vs. non-semantic agreement . . . . .                | 56        |
| 3.3.4    | Types of pronouns and phoric relations . . . . .             | 56        |
| 3.3.5    | Complex anaphora . . . . .                                   | 58        |
| 3.4      | Summary . . . . .  | 59        |
| <b>4</b> | <b>Gender: A typological survey</b>                          | <b>61</b> |
| 4.1      | Overview . . . . .   | 61        |
| 4.2      | The gender puzzle . . . . .                                  | 61        |
| 4.3      | Gender and noun classes . . . . .                            | 65        |
| 4.4      | Assignment systems . . . . .                                 | 67        |
| 4.4.1    | Semantic systems . . . . .                                   | 71        |
| 4.4.2    | Transition between assignment systems . . . . .              | 73        |
| 4.4.3    | Formal systems . . . . .                                     | 75        |
| 4.4.4    | Overt vs. covert gender . . . . .                            | 77        |
| 4.5      | Agreement . . . . .  | 78        |
| 4.5.1    | Agreement targets in Indo-European . . . . .                 | 78        |
| 4.5.2    | Constraints on agreement . . . . .                           | 81        |
| 4.5.3    | Hybrid nouns . . . . .                                       | 83        |
| 4.6      | Summary . . . . .  | 84        |
| <b>5</b> | <b>Gender in Celtic</b>                                      | <b>85</b> |
| 5.1      | Overview . . . . .   | 85        |
| 5.2      | The Celtic language family . . . . .                         | 85        |
| 5.3      | Initial mutations and gender agreement . . . . .             | 87        |
| 5.3.1    | Historical aspects . . . . .                                 | 87        |
| 5.3.2    | Initial mutations in synchrony: theoretical issues . . . . . | 88        |
| 5.3.3    | Initial mutation and the lexicon in FDG . . . . .            | 99        |
| 5.4      | Irish . . . . .  | 101       |
| 5.4.1    | Realization and scope of agreement . . . . .                 | 102       |
| 5.4.2    | Semantic assignment . . . . .                                | 113       |
| 5.4.3    | Formal assignment . . . . .                                  | 114       |

|          |   |            |
|----------|---|------------|
| 5.4.4    | Dialectal variation and double gender . . . . .                             | 117        |
| 5.5      | Scottish Gaelic . . . . .   | 122        |
| 5.5.1    | Realization and scope of agreement . . . . .                                | 122        |
| 5.5.2    | Semantic assignment . . . . .   | 127        |
| 5.5.3    | Formal assignment . . . . .   | 129        |
| 5.6      | Manx . . . . .  | 129        |
| 5.6.1    | Realization and scope of agreement . . . . .                                | 130        |
| 5.6.2    | Gender assignment . . . . .   | 133        |
| 5.7      | Welsh . . . . .   | 133        |
| 5.7.1    | Realization and scope of agreement . . . . .                                | 134        |
| 5.7.2    | Semantic assignment . . . . .   | 145        |
| 5.7.3    | Formal assignment . . . . .   | 146        |
| 5.7.4    | Variation in agreement marking . . . . .                                    | 147        |
| 5.8      | Breton . . . . .  | 152        |
| 5.8.1    | Realization and scope of agreement . . . . .                                | 153        |
| 5.8.2    | Gender assignment . . . . .   | 157        |
| 5.9      | Cornish . . . . .   | 158        |
| 5.9.1    | Realization and scope of agreement . . . . .                                | 158        |
| 5.9.2    | Gender assignment . . . . .   | 162        |
| 5.10     | Summary . . . . .   | 162        |
| <b>6</b> | <b>Gender agreement in the noun phrase</b>                                  | <b>165</b> |
| 6.1      | Overview . . . . .  | 165        |
| 6.2      | Variation in gender agreement within the noun phrase: the<br>data . . . . . | 165        |
| 6.2.1    | Irish . . . . .   | 166        |
| 6.2.2    | Welsh . . . . .   | 172        |
| 6.2.3    | Consequences for a model of gender assignment . . . .                       | 182        |
| 6.2.4    | Consequences for a model of gender agreement . . . .                        | 183        |
| 6.3      | Modelling noun-phrase agreement in Irish and Welsh . . . .                  | 186        |
| 6.4      | Summary . . . . .   | 197        |
| <b>7</b> | <b>Anaphoric agreement and structure of the lexicon</b>                     | <b>199</b> |
| 7.1      | Overview . . . . .  | 199        |
| 7.2      | Anaphoric gender agreement: the data . . . . .                              | 200        |
| 7.2.1    | Irish . . . . .   | 200        |
| 7.2.2    | Welsh . . . . .   | 206        |

|          |  |            |
|----------|--|------------|
| 7.2.3    | Agreement with feminine antecedents: analytic issues           | 215        |
| 7.2.4    | Non-quantified variation . . . . .                             | 218        |
| 7.3      | Modelling anaphoric agreement . . . . .                        | 218        |
| 7.4      | Agreement reduction and pronominal switch . . . . .            | 226        |
| 7.5      | Markedness, convergence and directionality . . . . .           | 230        |
| 7.6      | Grammatical gender as an afunctional category . . . . .        | 233        |
| 7.7      | Gender agreement: lexical vs. contextual information . . . . . | 235        |
| 7.8      | Summary . . . . .  | 237        |
| <b>8</b> | <b>Summary and conclusions</b>                                 | <b>239</b> |
| 8.1      | Overview . . . . .   | 239        |
| 8.2      | Summary of the previous chapters . . . . .                     | 239        |
| 8.3      | Review of research goals and main findings . . . . .           | 241        |
| 8.4      | Significance of this study . . . . .                           | 245        |
| 8.5      | Directions for future research . . . . .                       | 246        |
| <b>A</b> | <b>The Irish data: further analysis</b>                        | <b>249</b> |
| A.1      | Article agreement . . . . .                                    | 249        |
| A.1.1    | Article agreement and syntactic case in OC-I . . . . .         | 249        |
| A.1.2    | Article agreement and syntactic case in CC-I . . . . .         | 250        |
| A.1.3    | The semantic factor . . . . .                                  | 251        |
| A.1.4    | Proper names . . . . .   | 251        |
| A.2      | Adjective agreement . . . . .                                  | 253        |
| A.2.1    | Proper names . . . . .   | 253        |
| <b>B</b> | <b>The Welsh data: further analysis</b>                        | <b>255</b> |
| B.1      | Article agreement . . . . .                                    | 255        |
| B.1.1    | The semantic factor . . . . .                                  | 255        |
| B.1.2    | Proper names . . . . .   | 256        |
| B.2      | Adjective agreement . . . . .                                  | 257        |
| B.2.1    | The semantic factor . . . . .                                  | 257        |
| B.2.2    | Proper names . . . . .   | 257        |
| B.3      | Agreement with demonstrative pronouns . . . . .                | 258        |
| B.3.1    | The older component . . . . .                                  | 258        |
| B.3.2    | The contemporary component . . . . .                           | 258        |
| B.4      | Agreement with numeral pronouns . . . . .                      | 259        |
| B.4.1    | The older component . . . . .                                  | 259        |

---

|  |            |
|--|------------|
| B.4.2 The contemporary component . . . . . | 259        |
| <b>Bibliography</b>                        | <b>263</b> |



# List of Figures

|     |  |     |
|-----|--|-----|
| 2.1 | Levelt's "blueprint for the speaker" . . . . .               | 30  |
| 2.2 | The architecture of FDG . . . . .                            | 32  |
| 2.3 | Lexemes and suppletive forms in FDG . . . . .                | 41  |
| 3.1 | Anaphoric relations . . . . .                                | 54  |
| 5.1 | A genealogical table of the Celtic family . . . . .          | 86  |
| 6.1 | Article agreement in Irish, OC-I vs. CC-I . . . . .          | 167 |
| 6.2 | Adjective agreement in Irish, OC-I vs. CC-I . . . . .        | 169 |
| 6.3 | Article agreement in Welsh, OC-W vs. CC-W . . . . .          | 173 |
| 6.4 | Adjective agreement in Welsh, OC-W vs. CC-W . . . . .        | 174 |
| 7.1 | Pronominal gender distributions in Irish and Welsh . . . . . | 231 |
| 7.2 | Sources of gender agreement information in FDG . . . . .     | 236 |



# Abbreviations

|                   |  |
|-------------------|--|
| #                 | word boundary                                |
| =                 | clitic boundary                              |
| 1, 2, 3           | first, second, third person                  |
| ADJ, adj.         | adjective                                    |
| ADVR              | adverbializer                                |
| AM, <sup>AM</sup> | aspirate mutation                            |
| Ap                | adjective phrase                             |
| Aw                | adjectival word                              |
| ART, art.         | article                                      |
| C                 | (any) consonant                              |
| CC-I, CC-W        | contemporary corpus component (Irish, Welsh) |
| COND              | conditional                                  |
| COP               | copula                                       |
| DAT, dat.         | dative                                       |
| DEM               | demonstrative                                |
| DIM               | diminutive                                   |
| DG                | double gender                                |
| EMPH              | emphatic                                     |
| F, f., fem.       | feminine                                     |
| FDG               | Functional Discourse Grammar                 |
| FILL              | filler, hesitation                           |
| FG                | Functional Grammar                           |
| FUT               | future                                       |
| GEN, gen.         | genitive                                     |
| Gw                | grammatical word                             |
| H                 | aspiration                                   |
| IL                | Interpersonal Level                          |

---

|                     |  |
|---------------------|--|
| IMP                 | imperative   |
| L                   | lenition   |
| lit.                | literally  |
| M, m., masc.        | masculine  |
| ML                  | Morphosyntactic Level  |
| N                   | nasal mutation   |
| NEG                 | negation, negative (particle)  |
| Np                  | noun phrase  |
| Nw                  | nominal word   |
| NT, nt.             | neuter   |
| NOM, nom.           | nominative   |
| OC-I, OC-W          | older corpus component (Irish, Welsh)                                      |
| p.c.                | personal communication   |
| PL                  | Phonological Level (in FDG analysis);<br>plural (in morphological glosses) |
| pl.                 | plural   |
| POSS                | possessive   |
| PP                  | phonological phrase  |
| PREP, prep.         | prepositional (case)   |
| PST                 | past   |
| PTC                 | particle   |
| PTCP                | participle   |
| PVA, <sup>PVA</sup> | pre-vocalic aspiration   |
| PW                  | phonological word  |
| REL                 | relativizer  |
| RL                  | Representational Level   |
| RnaG                | Raidió na Gaeltachta   |
| RnaL                | Raidió na Life   |
| ScG                 | Scottish Gaelic  |
| SG, sg.             | singular   |
| SM, <sup>SM</sup>   | soft mutation  |
| SMR, <sup>SMR</sup> | restricted soft mutation   |
| s.v.                | sub voce   |
| T                   | <i>t</i> -prefixing  |
| V                   | (any) vowel  |
| VN                  | verbal noun  |

VOC, voc.                  vocative

Interlinear glosses are provided for all examples taken from languages other than English and are set according to the Leipzig Glossing Rules.<sup>1</sup> For the sake of clarity, however, interlinear glosses were omitted in longer examples, such as dialogues containing several turns, when the relevant feature could be highlighted otherwise.

---

<sup>1</sup><http://www.eva.mpg.de/lingua/resources/glossing-rules.php>, accessed 1 March 2011.



# Chapter 1

## Introduction

### 1.1 Goals and scope of this study

This is a study of linguistic variation and change in a situation of language contact and obsolescence. It focuses on the ongoing changes in the grammatical gender systems of Irish and Welsh as representative of the two branches of Insular Celtic and it aims to achieve two main goals:

1. to characterize gender agreement in Insular Celtic within the theoretical framework of Functional Discourse Grammar (FDG), that is, to provide a model that is descriptively adequate from the cognitive, typological and diachronic point of view;
2. to quantify and analyze the variation in gender agreement which has so far been described, mostly impressionistically, in the literature,<sup>1</sup> in order to identify recognizable and systematic patterns to be characterized and modelled.

More specifically, this study provides a statistical analysis of four sets of linguistic data pertaining to different varieties of Irish and Welsh, and shows the problems they pose to a model of agreement based exclusively on the notion of feature copying.

In relation to agreement within the noun phrase, the greatest difficulty consists in the differential incidence of gender agreement between the two

---

<sup>1</sup>The exception being, as we shall see, Welsh, for which quantitative analyses of agreement variation have been carried out.

main target types, viz. the article and the attributive adjective: in contemporary varieties, the latter shows agreement rates significantly lower than the former. This is problematic on two different levels:

1. typologically, distinctive agreement patterns are well known in relation to anaphoric elements but not to targets within the noun phrase itself;
2. from a theory-internal point of view, FDG models agreement as feature copying and explains the distribution of syntactic and semantic agreement forms in terms of the number of syntactic boundaries intervening between controller and target, so it is not clear why feature copying should be to some degree impeded in the case of the attributive adjective but not in the case of the article, given the same syntactic distance from the noun.

The model of agreement which I put forward to overcome this difficulty takes into account the significance of the data, is able to accommodate both the traditional and the contemporary agreement systems of Irish and Welsh, and is in line with what we know about agreement patterns in a typological perspective; furthermore, it provides an account of the language-internal aspect of the observed change.

As regards anaphoric agreement, the type of variation emerging from the data and described in the literature is consistent with well-known typological patterns, and does not pose any particular problem to the feature-copying model of agreement. However, I will argue that an extension of the current FDG theory in this regard is necessary in order to better account for the way in which different types of information capable of influencing pronominal agreement forms are stored and accessed by the speaker.

Finally, having explained the variation in anaphoric agreement forms as the result of resemantization, I address two related questions: what might have caused the latter and why it has resulted in the overgeneralization of masculine rather than feminine forms (directionality of change).

Attention will be paid to two collateral issues: (a) whether a functional explanation for the existence of grammatical gender is possible, and (b) whether the gender affiliation of each noun is a lexically-stored piece of information or rather worked out “on line”, based on specific sets of rules, each time it must be accessed for agreement purposes.

This study therefore makes an important contribution to the development of the current model of FDG by expanding it on empirical grounds, as motivated by the data at hand, and it points to the need for further investigation into the interface between the grammatical and non-grammatical components of the same model. It also contributes to the study of Irish linguistics in particular, through the creation of a new data set in order to reflect contemporary spoken Irish and mirror an existing data set for contemporary spoken Welsh. The new data set for Irish, which I plan to expand and make available to the research community as part of a future data collection project, is referred to in what follows as CC-I and is described in §3.2.2.2.

The particular sociolinguistic context of Irish and Welsh, and the role of language contact, are given due consideration as external factors involved in the shaping of the observed variation. Irish and Welsh are minority languages now spoken almost exclusively by bilinguals, with rather restricted usage domains; this, as will emerge from the discussion, has important consequences.

Some of the considerations and analyses I present in this study (particularly in chapters 1, 3, 5, 7) will appear in *Frenda* (forthcoming) and I am indebted to its two anonymous reviewers for their comments.

## **1.2 The context of this study**

### **1.2.1 The complexity of grammatical gender systems**

Grammatical gender is a linguistic feature which defines sets of lexical items (typically nouns) on the basis of more or less obvious properties, either formal or semantic, of those items. Depending on the number and nature of the defining criteria, we may be able to tell the gender affiliation of a noun from the noun itself or we may have to infer it from the shape of other linguistic elements in the context, which are said to agree in gender with that noun. Different languages are often found to differ as to the number of gender classes, the defining criteria, and whether or not it is possible to tell the gender of a noun from its shape.

The criteria that justify the distribution of nouns across the various genders are not always transparent. The term “arbitrary” has often been used to describe gender affiliation, and several attempts have been made to reduce

to finite sets of rules the apparent unpredictability of gender assignment in many languages. Corbett's (1991) seminal monograph examines a typologically diverse corpus of evidence and demonstrates that gender affiliation can be predicted on the basis of language-specific sets of formal and/or semantic rules—algorithms which, according to Corbett, speakers “run” every time an agreement form must be produced. Others hold the alternative view that gender information is stored in the lexicon and on-line computation is a task that speakers can be spared. The different positions on this question will be reviewed in §4.4.

Gender is a classification system on a par with several others through which some degree of granularity is imposed on the continuous nature of the external world; a good example of such classification systems is the way in which the spectrum of visible light can be divided into different portions that are assigned different colour terms. There are many ways in which this spectrum may be subdivided, as the colour-term systems of different languages show (cf. Berlin & Kay, 1969): this means that there is no objectively given subdivision of the spectrum but different, culturally-determined and linguistically-enshrined ways of referring to different portions of the continuum. Along these lines, grammatical gender systems too may be considered as culture-specific instances of a fundamental cognitive faculty of our species, i.e. the ability to categorize reality, or carve it into portions, in order to make reference to it possible (cf. the works of Craig, 1986; Lakoff, 1987; Taylor, 1995). However, it has been questioned whether gender systems are (always) useful in this sense, as discussed in §4.2.

The study of variation and change in grammatical gender systems is therefore particularly interesting for those investigating the relationship between form and function in language: as a system of noun classification, grammatical gender involves syntax, morphology, phonology and semantics, so that “understanding categorization in a particular language offers us a glimpse into several levels of linguistic representation” (Polinsky, 2008: 41).

### **1.2.2 Variation in obsolescence: the case of Insular Celtic**

“Language obsolescence” is defined by Jones (1998a: 5f.) as a “gradual reduction in use, due to domain-restriction, [which] may result in the emergence of historically inappropriate morphological and/or phonological forms together with extensive lexical borrowing”. Language obsolescence may be

a consequence of language switch, which necessarily happens in a context of language contact (cf. McGahan, 2009: §1.1).

The presence of a competing linguistic variety must therefore be taken into account. It seems to be consensual that, in a situation of contact between two (or more) languages, reciprocal modification ensues:

Languages in contact—where a significant proportion of the speakers of one also have some competence in the other—gradually become more like each other. Language contact may bring about gradual convergence resulting in structural isomorphism [Aikhenvald, 2003: 2]

[...] languages spoken by bilinguals are often altered such that ensuing changes differ from the results of internal processes of change within monolingual speech communities. In other words, languages spoken by bilinguals influence each other in various ways. [Sankoff, 2001: 1]

“Convergence” is understood here as a process whereby linguistic structures that are shared by both contact varieties tend to be reinforced and maintained while those that are not will in time become disfavoured and lost (cf. Backus, 2004; Bullock & Toribio, 2004; Toribio, 2004).

However, the existence of contact effects in a given scenario does not exclude the possibility that language-internal factors, i.e., factors that are independent from the speakers’ bilingualism, may also play a role in the observed change: “Multiple causation, or ‘fusion of endogeny and exogeny’ [...] is indeed a possibility which has always to be taken into account” (Filpula, 2003: 170); markedness and complexity may be regarded as language-internal factors, while the existence of a simpler (less marked) system in the contact variety, and language dominance, as external ones (Lleó et al., 2008: 193).

It has been suggested that internally-motivated changes take place in essentially the same way in both obsolescent and non-obsolescent languages, the difference between the two scenarios being one of time scale. In other words, language change in obsolescent languages happens faster than in “healthy” ones, but is not qualitatively different (Dorian, 1976b, 1981; Jones, 1998a: 240):

The types of change in formal language structure [in obsolescent languages] [...] are not notably different from those well established in the study of language change in general. But the time-span for change seems to be compressed [...] and the amount of change seems relatively large. [Dorian, 1981: 154]

This observation is consistent with the basic functionalist assumption, to be discussed in §2.3, that at any given stage the structures of a language must typologically possible. Therefore, whether change is entirely internally motivated or has external concauses or triggers, its effect should be the same, in the sense that it should produce synchronic states that are compatible with what are considered universally valid restrictions on the occurrence of language structures.

If this is the case, then one advantage of studying obsolescent varieties is that it should give us the opportunity to observe, in a compressed time span, a sequence of changes that would take longer to become established in “healthier” languages.

The linguistic effects of obsolescence, in terms of historically unexpected forms, can be noticed in various domains; in what follows, a few examples from the Insular Celtic group will be presented. As regards syntactic constructions, Ó Curnáin (2007a: 36) observes the change from the traditional [object pronoun + connective particle + verbal noun], as in example (1a) to [verbal noun + object pronoun], as in (1b) in the Irish spoken by teenagers in Iorras Aithneach (Galway Gaeltacht).<sup>2</sup>

(1) Irish (Ó Curnáin, 2007a: 36)

a. Bhfuil tú i ndan [é a shroicheadh]?  
 be.PRS 2SG able 3SG.M PTC reach.VN

‘Can you reach it?’

b. Níl mé i ndan [*reach-áil* é].  
 NEG.be.PRS 1SG able reach-VN 3SG.M

‘I can’t reach it.’

The structure in (1b), unlike that in (1a), is syntactically aligned with that of the competing variety, English.

<sup>2</sup>Contrary to appearance, example (1b) was not uttered as the answer to (1a); they are two independent utterances, the latter recorded a number of years after the former.

Turning to phonology, we can observe the loss of certain distinctions; for instance, a [ $\pm$ tense] opposition used to distinguish the sonorants /N, R, L/ from their lenited, i.e. [ $-$ tense], counterparts /n, l, r/.<sup>3</sup> However, this opposition is virtually lost today (cf. de Bhaldraithe, 1953: 257 fnn. 1–3; Ó hUiginn, 1994: 560, 562; Ua Súilleabháin, 1994: 488; Hickey, 2003: 264; Ó Curnáin, 2007a: 188–201). Outside the traditional Gaeltacht contexts, Maguire (1991: 199f., 200, 203, 212) describes the reduction of the phonological inventory with the loss of the palatality opposition in a self-established “neo-Gaeltacht” in Belfast<sup>4</sup>—a change from which she observes even native speakers in strong Gaeltacht areas are not immune (ibid., p. 200).

The effects of obsolescence can be seen in morphological structures too; Níc Pháidín (2003: 125), in describing the variety of Irish spoken by *Gaelscoil* pupils,<sup>5</sup> observes how some speakers use analytical forms (preposition + pronoun) instead of the traditional inflected prepositions. In (2), for instance, *d(e) iad* is used instead of the synthetic third-person plural form *díobh*), and Ó Curnáin (2007b: 1278–1280) notes that in the Gaeltacht variety of Iorras Aithneach analytic expressions are used with emphatic pronouns, e.g. *air i-se* (on 3SG.F-EMPH) ‘on *her*’ instead of the synthetic equivalent *uirthi-se* (on.3SG.F-EMPH); this too can be analyzed as a contact-induced change whereby a native structure is replaced by one that is found in the competing variety.

(2) Irish (Níc Pháidín, 2003: 125)

Bhí dhá d’ iad ann.  
be.PST two of 3PL there

‘Two of them were there.’

Specific aspects of grammatical gender in Celtic have been studied quite extensively in Welsh as regards specific contexts, such as acquisition by chil-

<sup>3</sup>Cf. §1.4 below. Lenition of sonorants is not represented orthographically.

<sup>4</sup>This neo- (i.e. non-traditional) Gaeltacht, also known as the Shaw’s Road Gaeltacht, was unofficially established in 1969 by some forty Irish-speaking families; some of them had connections with the (traditional) Donegal Gaeltacht, but most of them were couples who had learned Irish as adults and wanted a supportive environment to raise their own children as Irish speakers (Maguire, 1991: 67f.).

<sup>5</sup>*Gaelscoil* (pl. *Gaeilscoileanna*) is the Irish name for an Irish-language primary or secondary school, typically established outside the Gaeltacht, where children are taught through the medium of Irish.

dren (Gathercole et al., 2001), access to gender information under language impairment (Tainturier et al., 2005), the robustness of the system of initial mutations as markers of gender agreement, with specific reference to language obsolescence (Thomas, 2001; Thomas & Gathercole, 2005); their findings will be reviewed in §5.7.4. The study of gender in the context of language obsolescence was also undertaken by Dorian (1973, 1976b, 1976a, 1978a, 1978b), who studied a local variety of Scottish Gaelic (see §5.5). Grammatical gender has also entered more general discussions, for instance in monographs like Ball & Müller (1992) on the initial mutation system in Welsh. As regards Irish specifically, it is probably fair to say that grammatical gender per se has been neglected, with sporadic exceptions like Ó Siadhail (1984) (cf. §5.4); in the context of wider linguistic descriptions, a certain amount of attention has been paid to issues of variation in gender agreement assignment and marking by Ó Curnáin (2007a, 2007b) and Maguire (1991) (cf. §5.4 and this chapter). However, Celtic languages do not feature in either typological surveys of gender systems like Corbett (1991) or Unterbeck & Rissanen (1999) or theory-specific analyses like Nessel (2006) or Rice (2006), and their contribution to the study of grammatical gender in a typological perspective has in this sense remained marginal.

Some amount of variation in gender assignment and agreement in Insular Celtic has already been observed in the literature; one of the goals of this study, as already mentioned, is to establish whether this variation is systematic or random. Some evidence is provided by the cases of Manx and Cornish, which show how the gender of inanimates ceased to be distinguished and marked in the final stages of these languages (§§5.6 and 5.9); in Irish and Welsh, on the other hand, an ongoing process of simplification may be witnessed: the limited predictive power of formal assignment criteria seems to cause uncertainty in the gender assignment of inanimates, as reported by Jones (1998a) for younger Welsh speakers and by Maguire (1991: 211) for the Belfast neo-Gaeltacht, where the virtually complete collapse of the traditional assignment system has been observed. Both Jones's and Maguire's informants were schoolchildren going through Welsh- and Irish-medium education respectively, and were exposed to a more or less standardized variety of the language. But the same developments, or early-stage traces thereof, are found in the more traditional varieties of older speakers, as documented for Welsh (Jones, 1998a: 66, 171) and Donegal Irish (Ó Siadhail,

1989: 149), in which anaphoric reference to inanimates tends to be made by means of masculine pronouns irrespective of the grammatical gender of the antecedent.

Whether and to what extent such developments are contact-induced is debatable: it has been suggested that the overgeneralization of the masculine anaphor could be independent because it has been attested for quite a long time in Donegal Irish (Ó Siadhail, 1984: 175), and Comrie (1979) argues for the possibility of language-internal motivations such as phonetic distance in Welsh, where the phonetic distance between a given phoneme and its lenited counterpart would correlate with the occurrence of initial mutations.<sup>6</sup> Dorian (1976b) studied lenition in different contexts in a variety of Scottish Gaelic (East Sutherland Gaelic) and concluded that language-internal factors like functional load may explain this type of variation. She compared four environments in which lenition is traditionally required: (i) as a mark of past tense; (ii) as a mark of feminine-gender agreement after the article; (iii) after certain adverbs and numerals; and (iv) as a mark of vocative; what she found is that fluent speakers never failed to produce lenition in contexts (i) and (ii), almost never failed to lenite in (iii) and sometimes failed to lenite the vocative (iv). Semi-speakers (defined as “community members who can make themselves understood in Gaelic, but who speak a very imperfect version of the language”; Dorian, 1976b: 97) failed to lenite in (i) and (ii) only 10.5% and 28.5% of the time, respectively, and failed to lenite in (iii) and (iv) 46% and 74% of the time, respectively. Dorian explained the higher rates of mutation retention in contexts (i) and (ii) in terms of presence vs. absence of functional load: lenition provides the only way of distinguishing between the past tense form of the verb and the imperative, second person singular in (i), and the sole signal of the grammatical category of gender in (ii). On the other hand, the presence or absence of lenition after certain adverbs and numerals (iii) or after the vocative (iv) is not distinctive in any way. (The same has been observed apropos of the so-called aspirate mutation in Welsh: cf. §5.7.1.2).

However, Dorian also considered the effects of contact and noted that

---

<sup>6</sup>Phonetic distance is defined as the “number of features (within some overall framework of phonetic features) by which they differ” (Comrie, 1979: 51f.): the greater the distance, the less likely the occurrence of mutation. In the context of soft mutation (see §5.7), lenition of /b/ > /v/ is therefore more likely to be preserved than lenition of /g/ > ∅.

structures which had a functional counterpart in English appeared more resilient than those which did not—and gender distinctions within the realm of inanimates are almost completely absent from English (Dorian, 1976b: 100; 1978a); Maguire’s observation about the loss of the palatalized/non-palatalized opposition in the neo-Gaeltacht children’s variety, a development which brings the phonology of this variety closer to that of English, was mentioned above.

In this regard, Maguire explicitly suggests that speakers adopt the “lowest common denominator” between the two varieties available to them, i.e., they choose the form that most readily suits the communicative needs, often a non-standard one, “directly influenced by the English system” (Maguire, 1991: 189f.); for instance, the opposition between /s/ and /s’/ [ʃ]—paralleled by the phonemic distinction between /s/ and /ʃ/ in English—is retained, whereas the opposition between /r/ and /r’/—which is not phonemic in English—is not. Developments like these are particularly important, since the phonological opposition of palatality was instrumental to the preservation of the declension system in Irish after the loss of the inflectional endings and would be among the few formal correlates of gender left to the language if case inflection were to be lost entirely (see §5.4.3).

### 1.3 Irish and Welsh today

Irish and Welsh are two obsolescent languages which have been the object of revitalization efforts for decades; as a result, a good amount of data is available. In addition, comparisons can be made with cognate languages which are now extinct, such as Manx and Cornish.<sup>7</sup>

Both languages are spoken in communities largely comprised of English L1 speakers and have similar gender assignment systems, in which the sexual criterion determines the gender of nouns with animate referents (humans, certain higher animals); all other nouns are assigned by formal criteria and some secondary semantic criteria with rather limited predictive power (cf.

<sup>7</sup>As an anonymous reviewer of Frenda (forthcoming) points out, there is in fact a Manx-speaking community complete with a Manx-medium school on the Isle of Man; when this community of revivalists was established, the few extant native speakers of Manx were still alive, and it was from them that the revivalists learned the language (see George, 1993: 654–660). Therefore, the term “extinct” might be felt by someone as inappropriate with reference to Manx. Here I use the term as a shorthand for “no longer spoken natively”.

§§5.4 and 5.7 for Irish and Welsh, respectively). Both Irish and Welsh enjoy official or semi-official status and both have a fairly long and well-established tradition as languages used in the media (press, radio and television), and more or less extended communities of native speakers.

As spoken languages, Irish and Welsh are not monolithic entities: both have a number of dialects which coexist with standard varieties, both older literary ones and more recent ones, created for teaching and official purposes (Ball, 1993: 201; Ó Baoill, 1988: 112, 114; Borsley et al., 2007: 5–7).<sup>8</sup> In the latter case, standardization often involved seeking compromise solutions, i.e. choosing between competing dialectal variants, and the resulting varieties often met the hostility of native speakers who regarded them as “made-up”, unnatural versions of the language (see Awbery, 1987; Breatnach, 1964; Davies, 1988; Dorian, 1994; Jones, 1994; Ó Baoill, 1988).

At any rate, the education system and the media have managed to approach and win new speakers through more or less standardized forms in both the Irish and Welsh contexts (Jones, 1998a: 98; Ó Baoill, 1988; Ó Murchú, 1969; Watson, 1989: 47), which has raised some concerns about the loss of dialectal features (Coupland & Ball, 1989; Jones, 1994, 1998a; Sayers, 2004, 2005).

However, traditional dialects are now given visibility and promoted, including through dedicated radio and television stations. In the Irish case, *Raidió na Gaeltachta* (‘Radio of the Gaeltacht’) was established in 1972 and TG4 (originally *Teilifís na Gaeilge* ‘Irish-language television’) in 1996 (Akutagawa, 1987: 138). These stations broadcast a wide range of programmes, but interestingly they have chosen to regard the native varieties of the Gaeltachtaí as “pure Irish” and to employ these rather than the official standard<sup>9</sup> even in the more formal, “corporate-brand” productions such as the news programmes.<sup>10</sup> Jones (1998a: §4.2.4) extensively reports on the varieties of Welsh that are used on Welsh-speaking media, and observes that the creation of a “BBC dialect” for Welsh is actively avoided by S4C (*Sianel Pedwar Cymru*, the Welsh-medium television channel), which strives to give equal visibility to both northern and southern dialects; nevertheless, finer-

---

<sup>8</sup>Literary standards tend to be highly archaic and quite distant from all existent spoken varieties (Fife, 1986; Ó Baoill, 1988).

<sup>9</sup>See for instance Ó Murchú (1969) and Ó Baoill (1986).

<sup>10</sup>Telephone interview with Alan Esslemont, Director of Television, TG4, 22 January 2007.

grained distinctions are being lost and an increasing degree of dialect-mixing can be observed, while at the same time a new standard of abstract, non-dialectal pronunciation based on orthography is developing on the media and consequently gaining status.<sup>11</sup>

Jones (1998a: 329–331) also observes that speakers of different dialects of Welsh are still able to identify with the traditional standard, “with its long history and well-established forms which, as they can see, do not overly favour any single dialect”. Furthermore, the revivalists’ efforts (bilingual schools, evening classes for adults, etc.) and the migratory movement of native speakers from the Welsh heartlands to the south-eastern urban areas, “for employment purposes in the newly-expanded Welsh-medium professions (the media, local government, the Welsh Office, etcetera)” (Ball, 1993: 198) have brought about a situation in which

Welsh may be shifting from a set of rural heartlands where the language is dominant, to a set of social networks in both rural and urban areas where it dominates only in the network, with outside them English being the main medium of communication.  
[Ball, 1993: 198]

## 1.4 A note on the phonological representation of Goidelic varieties

Goidelic languages are characterized by the phonemic opposition between velarized and palatalized consonants. In phonological transcription (cf. Ní Chasaide, 1999), this opposition is represented by the symbol /<sup>ɣ</sup>/ for velarized consonants (e.g. /p<sup>ɣ</sup>/) and the symbol /<sup>j</sup>/ for their palatalized counterparts (e.g. /p<sup>j</sup>/). (Inherently velar or inherently palatal consonants do not receive any further marking, e.g. /k/, /j/, respectively.)

However, it is common practice in Celtic linguistics (e.g. Ó Siadhail, 1989; Ó Sé, 2000; Ó Curnáin, 2007c) to only explicitly mark the palatalized

---

<sup>11</sup>Generally speaking, differences between the regional varieties of colloquial Welsh are minimal and allow mutual intelligibility (see Fife, 1986: 148; Thomas, 1992b: 263f.); Welsh-medium education (partly by reintroducing Welsh in areas where it had died out) has “tended to level out dialect differences”, so that “increasingly, with some local exceptions, two fairly uniform spoken standards are emerging in north and south Wales respectively” (Borsley et al., 2007: 6).

member of the opposition, which is done using either /<sup>j</sup>/ or the equivalent notation /<sup>v</sup>/, so that /p/ = /p<sup>v</sup>/ and /p'/ = /p<sup>j</sup>/'. This type of notation (which I call here the “Celticists’ notation” or “C-notation” for brevity) disregards the fact that the opposition between velarized and palatalized does not necessarily involve all consonant phonemes in these varieties and has therefore been criticized by Hickey (1995: §4.8.1).

In Irish, for instance, this opposition has come to be neutralized in the case of nasal and lateral sonorants in Gaoth Dobhair (Donegal) and Conemara Irish (Ní Chasaide, 1999; Ó Raghallaigh, 2010), where different phonemic mergers have turned a historical four-way contrast in the dental to alveolopalatal region into a three-way contrast, as exemplified in (3) for the nasal series.

- (3) a. Historical four-way contrast: /n<sup>v</sup> : n<sup>j</sup> : n<sup>v</sup> : n<sup>j</sup>/

|             | Dental                | Alveolar       | Alveolopalatal        |
|-------------|-----------------------|----------------|-----------------------|
| Velarized   | <u>n</u> <sup>v</sup> | n <sup>v</sup> |                       |
| Palatalized |                       | n <sup>j</sup> | <u>n</u> <sup>j</sup> |

- b. Merger between /n<sup>j</sup>/ and /n<sup>j</sup>/

- c. Present-day three-way contrast: /n<sup>v</sup> : n<sup>j</sup> : n/

|             | Dental                | Alveolar | Alveolopalatal        |
|-------------|-----------------------|----------|-----------------------|
| Velarized   | <u>n</u> <sup>v</sup> | n        |                       |
| Palatalized |                       |          | <u>n</u> <sup>j</sup> |

In a similar way, in the lateral series, the four-way contrast /l<sup>v</sup> : l<sup>j</sup> : l<sup>v</sup> : l<sup>j</sup>/ has reduced to /l<sup>v</sup> : l<sup>j</sup> : l/. In these varieties, “[t]he alveolars /n/ and /l/ “have a fairly neutral or slightly clear quality and tend to coarticulate freely with the quality of adjacent segments” (Ní Chasaide, 1999: 114).<sup>12</sup>

While it is true that the C-notation would not be appropriate for a phonological study of the language, it has been adopted here for the immediate advantages that it has once it has been established that the phonological opposition in question is not central to our discussion. One advantage is that it permits consistent representation across the different Goidelic varieties surveyed in this study: the C-notation is the one adopted by the authors I consulted for the other two Goidelic varieties described in chapter

<sup>12</sup>In other varieties (e.g. in Corca Dhuibhne Irish, Munster), however, the velarized–palatalized opposition amongst lateral and nasals has been retained (Ó Sé, 2000; Ó Raghallaigh, 2010: 47).

5, namely Scottish Gaelic and Manx. A more detailed study of the phonology of these two varieties would have been necessary to establish the relevant facts and adopt a more precise phonological notation throughout chapter 5, but it would have been outside the scope of this study. Another advantage is that /'/ is an abstract notation and therefore more versatile than a phonologically more precise one: the relation between, e.g., *cás* 'case' (nom. sg.) and *cáis* (gen. sg.) is more evident if represented as /ka:s/ : /ka:s'/ than as /kʲa:sʲ/ : /kʲa:ɕ/.<sup>13</sup>

The four-way opposition among sonorants just characterized in terms of place of articulation and palatalization has traditionally been described in terms of tenseness and palatalization (cf., e.g., Ó Siadhail, 1989, and see Hickey, 1995 for a criticism of this tradition). In C-notation, tense sonorants are represented as capitals or small capitals, e.g.:

- /N/ = /N̄/ : /N'/ = /N̄ʲ/
- /L/ = /L̄/ : /L'/ = /L̄ʲ/

Again, for practical considerations this notation has been retained, due to its widespread use in the relevant literature and considering the fact that the phenomena in question are not central to this study.

## 1.5 Summary and organization of this work

In this chapter I anticipated the goals of this study and some of the collateral questions surrounding grammatical gender. There are two main research goals: (i) to provide an accurate description of grammatical gender in Irish and Welsh (as representative of the Insular Celtic language group) and a theoretically adequate model of the complex systems it instantiates in these languages; (ii) to analyze the variation in gender agreement which has already been observed in the literature and to determine whether it is random or systematic.

Special attention will also be paid to two collateral questions: whether a functional account of the existence of grammatical gender is possible and whether gender affiliation is recorded in the lexicon once and for all or com-

<sup>13</sup>/sʲ/ would not be a good representation for the phoneme represented as /ɕ/ by Ní Chasaide (1999) and as /ʃ/ by others.

puted on line, where necessary, based on the knowledge of the appropriate language-specific algorithm.

In §1.2 the relevant contextual information about the subject of this study was provided, with particular reference to the minority status of the languages under investigation and their being in contact with a dominant variety. The notion of convergence as a consequence of language contact was also introduced; this refers to the observation that structures shared by two languages in contact tend to be mutually reinforced, whereas those which are peculiar to the recessive variety tend to lose salience and may eventually be lost.

In §1.3 I briefly outlined the sociolinguistic setting in which Irish and Welsh are spoken and in §1.4 I discussed issues relative to the phonological representation of Goidelic varieties.

In chapter 2 I will turn to the theoretical framework of this study, Functional Discourse Grammar, and my reasons for adopting a functionalist perspective to analyze the data I present in chapter 3 together with the methodological choices I made.

In chapter 4 gender systems are discussed from a typological perspective, drawing on the work done by Corbett (especially Corbett, 1991) and paying particular attention to the Indo-European horizon and the mechanisms whereby gender systems may be transformed into typologically different ones. Two collateral issues, namely the debate about the function of gender and the opposing views on gender assignment will also be introduced there.

A detailed review of the grammatical gender system in Insular Celtic is provided in chapter 5, where I also discuss a specific agreement-marking device peculiar to Insular Celtic, the so-called initial mutations, from the historical and theoretical point of view. This chapter precedes, and prepares the ground for, the analysis of the data, contained in chapters 6 (agreement within the noun phrase) and 7 (anaphoric agreement); here the variation in gender agreement, described in the literature and observed in the data, will be presented and used to build a theoretical model of the Irish and Welsh agreement systems. Chapters 5–7 constitute therefore the analytic core of this study.

To conclude, chapter 8 summarizes the contribution made by this study in terms of the various research questions and points at areas in which future

work can be done to corroborate or reject some of the hypotheses put forward in this work and to complement the analysis on which it is based.

## Chapter 2

# The functional perspective

### 2.1 Overview

This study is based on the analysis of spoken data from actual interactions, in line with the theoretic assumption, to be discussed below, that the investigation of linguistic phenomena should always take into account their communicative functions (Dik, 1997: 6) and that the data we use should be derived from usage (Butler, 2009: 12).

In what follows I will explain why Functional Discourse Grammar (FDG) was considered an appropriate framework for this study. To do so, I will first describe in general terms the theoretical premises of functional approaches to the study of language (§2.2); then, in §2.3, I will focus on Simon Dik's attempt to formulate a descriptively adequate theory of Functional Grammar (FG), on its limitations and on how the proposal of a Functional Discourse Grammar was elaborated to address the criticisms directed at FG. Finally, in §2.4, the FDG model is outlined.

### 2.2 Functional approaches to the study of language

A functional approach can be characterized as one which sees language primarily as a means of communication between human beings in a given social and cognitive context, and therefore as a system shaped by specific cognitive and sociocultural factors (Dik, 1986: 3; Butler, 2009: 3). The natural communicative use that is made of the language is therefore of paramount

importance (Dik, 1997: 6), and samples of language use in natural communication need to be studied in order to supplement native speakers' intuition and data collected in experimental settings (Butler, 2009: 12–14).<sup>1</sup> Moreover, as the focus is on the communicative purpose, functionalists regard syntax as subservient to semantics, i.e. to the expression of some content, as motivated by the pragmatic context (Dik, 1997: 2, 8).

This defining perspective on the study of language distinguishes functional approaches from another important linguistic paradigm, usually referred to as “formal” or “generative”, which traces its origins to the theories elaborated by Chomsky in the late 1950s<sup>2</sup> and tends to confine the study of variability in language to the background, to reject the use of performance data and to regard syntax as a self-contained system of rules independent from meaning and context.

Human language, according to Chomsky (1980: 239), is a “system for free expression of thought, essentially independent of stimulus control, need-satisfaction or instrumental purpose”, and the object of linguistic study is the competence of the ideal language user “in his or her (fictitiously) homogeneous speech community” (Butler, 2009: 6). Competence is assumed to be only dimly reflected by performance (Chomsky, 1965: 4),<sup>3</sup> hence “data about usage and frequency—or even the issue of whether a given possibility is ever actually instantiated in any language” are of “marginal interest” to those interested in “what the human language faculty is”—as opposed to “what speakers do with language” (Anderson, 1999: 11).<sup>4</sup>

Van Valin & LaPolla (1997: 11) use the cover term “Communication-and-Cognition perspective” to refer to a number of functional theories which

---

<sup>1</sup>The problem with elicited data is that experimental settings are often very different from any typical scenario in which language is used for communication; moreover, depending on the degree of awareness, on the part of the subjects, that their linguistic abilities are being assessed, the response to the linguistic tasks set by the researcher may not necessarily be a valid indicator of “ordinary, everyday language behaviour” (Butler, 2009: 14). The importance of corpora as enabling the developments of functional approaches is emphasized by MacWhinney, 2010.

<sup>2</sup>Although it has become customary to apply these labels restrictively to this latter paradigm, FG and its successors are themselves generative theories which employ their own formalisms (cf. Anderson, 1999: 2 and discussion below).

<sup>3</sup>MacWhinney (2010: 18) comments that this supposition “cannot be challenged, since it is a methodological preliminary rather than a testable empirical claim”.

<sup>4</sup>But see Stokhof & van Lambalgen (2011) apropos of the complications that posing this distinction creates from the theoretical point of view.

share similar views on the issues just discussed; these include Simon Dik's Functional Grammar, Van Valin's Role and Reference Grammar, Cognitive Grammar (e.g. the work of Ronald Langacker and George Lakoff) and French functionalism (André Martinet), among others. Butler (2008: 7–10) further observes that this broad movement (to which he refers simply as “Cognitive Linguistics”) can be characterized in terms of the attention paid to various philosophical and psychological studies which question the adequacy of classical Aristotelian classification and advocate its replacement with family-resemblance or prototype-based criteria of classification (Butler, 2008: 7 and references therein).

At the heart of the definition of a functional approach lie the questions what the nature of functional explanations is and what a functionalist theory should account for (Butler, 2003: 11; 2008: 3). Given that the primary goal of a functional theory of language is to describe and explain linguistic phenomena by taking into account the instrumental nature of language as a means of social interaction, functional approaches are characterized by the acceptance of explanations that are external to a theory of syntax, be they language-internal or language-external (Van Valin & LaPolla, 1997: 13f.). Among the latter, Butler (2003: 13) includes—following Dik (1986)—such factors as communicative efficiency, physiological processing and the sociocultural context.

External factors such as contact between different varieties have also been taken into account in a number of formalist–generative studies (e.g. Ayoun, 2007; Sorace, 2004; Tsimplis et al., 2004) in explaining language change, but they have been explicitly confined to the theory-external sphere of *E-language*, i.e. to the realm of performance and variation—as opposed to *I-language*, the adult speaker's idealized competence. In this view, the linguistic input children are exposed to (*E-language*) interacts during the acquisition period with a set of species- and language-specific constraints to language acquisition called Universal Grammar which is assumed to be inherited and shared by all human beings. This interaction yields the adult speaker's grammar (*I-language*) through a process of parameter setting (cf., e.g., Roberts, 1993; Anderson & Lightfoot, 1999). The fact that certain logically possible patterns and developments are cross-linguistically unattested is therefore explained as owing to the existence of an innate and language-specific set of constraints. This process is central to language change as well,

which in the generative paradigm is attributed to changes in the composition of the external stimulus that children learning the language are exposed to (cf. Roberts, 1993; Lightfoot, 1997; Matthews, 2003).

A common position in functionalism is that language acquisition depends on the innateness of more general cognitive capacities that are not specific to language; Dik (1997: 7) maintained that only underlying principles which cannot be explained as acquired during the process of language acquisition through the interaction between the child and its environment should be attributed to genetic factors, while others have observed that recourse to an innate component is premature “as long as the exact nature and impact of the genetic component remains far from clear” and should be had only in the absence of “less controversial evidence” from areas like semantics, pragmatics and cognitive psychology (Rijkhoff, 2002: 3). Nevertheless, many authors are still open to the possibility that language-specific innate capacities might exist, but their identification remains a separate empirical issue (cf. Butler, 2003: 26).<sup>5</sup>

The positing of a Universal Grammar follows from what is known as the *argument from poverty of the stimulus*: the external input children are exposed to is not sufficient to define the adult speaker’s linguistic competence; in particular, it cannot explain why certain structures appear to never be overgeneralized by children when other structures are. If nothing in the input forbids the overgeneralization of a certain structure, its absence must be attributed to some innate constraint (Anderson & Lightfoot, 1999). However, this argument has been challenged, and models supporting the notion

---

<sup>5</sup>For instance, Anderson & Lightfoot (1999) claim that Universal Grammar must include an internal representation of the properties of a clitic (“the child is faced with a chaotic environment and scans it, looking for clitics” . . . “a general claim at the genetic level (clitics and their behavior are predefined)” . . . “a notion defined in advance of any experience”, p. 11). Anderson & Lightfoot do not indicate what typological support this particular hypothesis has (“We could have illustrated these points equally well with data from French or from Dutch, or from many other languages, because the principles apply quite generally, to pronouns in all languages”, p. 12); but apart from that, given that they conceive of Universal Grammar as the species’ “linguistic genotype”, it might be asked what the evolutionary advantage might be of transmitting the predefined notion of a clitic to one’s offspring (especially if one is born into a community where a language with no clitics is spoken). Incidentally, it would also appear to be remarkable that while actual genotypes are individual, the same “linguistic genotype” should be common to all the members of our species.

that children can learn from missing evidence have been put forward (cf. e.g. Regier & Gahl, 2004). But even if the hypothesis of a language-specific innate component were rejected, this would not imply the rejection of linguistic universals: their existence, as empirically observed, is explained in functional terms as resulting from the uses a language is put to and the external factors which determine the nature of languages. According to Dik, each natural language represents a particular way of addressing the “extremely complex problem” of making communication possible; the number and nature of possible solutions to this problem are constrained by various factors, such as the nature of the problem, the characteristics of the problem-solvers (i.e. the language users) and the environment in which they happen to interact (Dik, 1997: 7).

In the light of the foregoing it is apparent that a functional approach should be preferred, for a study of variation data such as the present one, to a theoretical approach which is not in fact concerned with performance data and whose declared object of study is the internalized competence of an ideal language user in his or her homogeneous community; as already noted by Kallen (1986: 80f.),

variation exists because speech communities are not homogeneous, and speaker-listeners in speech communities must be able to negotiate a variety of styles, dialects, registers, and the like in their use of the language. [Kallen, 1986: 81]

If the sole object of linguistic investigation is the language user’s competence, and if variation is confined to performance, the study of variation is outside the scope of interest of the formalist programme even when it is invoked, as discussed above, to explain language change (cf. Harris & Campbell, 1995: 37; Matthews, 2003).

The study of variation data from a functional-typological perspective has already proved useful in explaining the distribution of grammatical gender markers. Cases in point are Siemund (2008) and Audring (2009), whose studies, conducted on various types of corpus evidence, were able to explain the distribution of pronominal gender exponents in a number of English varieties (Siemund) and Dutch (Audring) in terms of factors that are cognitively, semantically and in some cases syntactically relevant, like degree of individuation and distinctions such as mass vs. count, concrete vs. abstract

and animate vs. inanimate.

However, explicit formulation of the facts and structures under investigation, as well as of any relevant hypothesis, is clearly desirable when descriptive adequacy is being sought. A solid formal infrastructure is therefore necessary.

Functional Discourse Grammar is what Butler (2008: 2) describes as a “structural-functional grammar”, a term which alludes to the generative character of the theory and to the fact that grammar is seen as a system. Unlike other structural-functional approaches,<sup>6</sup> structural-functional grammars are generative (in the primary sense of the term), i.e. they provide explicit “sets of rules and principles” that assign “structures and interpretations to linguistic expressions” (Butler, 2003: 30; cf. Chomsky, 1965: 4).<sup>7</sup> Structural-functional grammars therefore combine the advantages of a functional approach with the type of rigorous formalism that is necessary to attain descriptive adequacy.

The first attempt at a generative formulation of the functional approach was Simon Dik’s theory of Functional Grammar, which dates back to the 1970s (see e.g. Dik, 1978) and will be briefly presented in what follows before moving on to consider its most recent successor, FDG.

### 2.3 Functional Grammar and beyond: the emergence of FDG

Dik (1997) proposed that a theory of Functional Grammar should meet a number of criteria in order to be considered fit for purpose. Particularly important were for him the two criteria of *psychological adequacy* and *typological adequacy*. (The former is also referred to in the literature as *cognitive adequacy*, e.g. Butler, 2009: 17.) In other words, a functional theory of grammar must, on the one hand, be informed by empirically validated psychological models of linguistic competence and usage and, on the other, be applicable to the analysis of any type of language (Dik, 1997: 13–15). Psychological and typological adequacy are in Dik’s opinion deeply intercon-

<sup>6</sup>E.g. the work of Givón and West Coast Functionalism, now better known as the “usage-based” approach; see Butler (2008: 2) and references therein.

<sup>7</sup>Other current structural-functional grammars mentioned by Butler are Role and Reference Grammar and Systemic Functional Grammar.

nected, since any fundamental property of language which is universally or almost universally recurrent cross-linguistically must be a highly significant one from a psychological point of view.

For a theory to be typologically adequate, i.e. able to model the disparate structures one finds cross-linguistically, some level of abstraction is called for; however, unnecessary levels of abstraction are to be avoided. In particular, the “X is really Y” type of argument (whereby the surface realization becomes secondary to the presumed underlying structure, and the latter is analyzed instead) is criticized as leading to non-obvious analyses which are often far removed from the linguistic reality presented by the data (Dik, 1997: 16–22). On the same grounds, also ruled out are transformational analyses and all types of filtering devices: the latter in particular will be superfluous if the theory does not over-generate to begin with. In transformational analyses, the position of a constituent in the surface structure may be different from the position it occupies in the underlying representation (movement); this type of analyses, Van Valin & LaPolla (1997: ch. 2) note, was introduced by Chomsky to account for problems like case assignment to what the theory considered displaced constituents. For instance, and simplifying somewhat, under the premise that the accusative case in the affirmative clause is assigned to the noun phrase which immediately follows the verb, one can account for the assignment of the accusative case to WH-words (e.g. *Whom did Pat send to us?*) only by assuming that they were generated in a position immediately adjacent to the verb (*Pat sent whom to us?*) and then moved to the initial position (Van Valin & LaPolla, 1997: 17f.). Constituent movement is, therefore, required for theory-internal reasons rather than by any linguistic evidence (ibid.); in Functional Grammar, on the other hand, constituents are assumed to be “immediately placed in the position in which they actually occur in the final linguistic expression” (Dik, 1997: 20f.).

As was already mentioned, functional theories assume that language can be shaped by external forces, within the limits imposed by our cognitive make-up; this leads to both variation and recurrent similarities between different languages. Being open to the influence of changing and competing external forces, languages are naturally subject to change; as Butler observes,

[t]he position widely espoused within the functional literature is

that the state of a language at any particular time is the result of competing motivations. [Butler, 2003: 14]

At any particular synchronic stage of a language, a “precarious balance” is stricken between different external and internal factors (Dik, 1986: 22). Accordingly,

any non-simplistic view of the notion ‘functional explanation’ will not be committed to the view that linguistic form is a direct reflection of extragrammatical principles. [Dik, 1986: 18]

For instance, patterns of linguistic change are not constrained by the necessity to preserve meaning, as Labov (1994: ch. 19) demonstrated on the basis of empirical evidence showing that phenomena like phonetic erosion are neither blocked nor proceed at a slower pace in contexts where they delete morphological marks.<sup>8</sup>

The idea of competing motivations determining the path of linguistic change also account for the existence of “non-functional” (Nuyts, 2001: 4) or “a-functional” (Trudgill, 1999) structures (also cf. Newmeyer, 2003: 29). As Harder (2008) remarks,

[i]f there is a pervasive and well-entrenched coding pattern, it may be functionally optimal to use it even in cases where its functional motivation is weak or absent, simply to be able to stick to the same pattern in as many cases as possible. [...] Similarly, in a well-functioning transport system, you will sometimes see an empty bus. [Harder, 2008: 12f.]

Nevertheless, the kind of generalizations that typological universals capture, as well as being a reflection of our cognitive make-up, predict “possible and impossible diachronic developments with respect to the properties involved” (Dik, 1997: 32). A hierarchy is “a sequence of properties, claimed to be of absolute or statistical validity, such that a preceding property can occur without the following properties, but not the other way round” (Dik, 1989: 28), therefore, given the hierarchy  $A > B > C$ , if a language does not

---

<sup>8</sup>In fact, Labov observed “a tendency to use marks where they are least needed, and omit them where they are most needed” (Labov, 1994: 19).

have properties B and C, it will have to develop property B before property C; the possibility that C may be developed prior to B is ruled out by the hierarchy, as it would lead to a synchronic stage of the language with properties A and C but without property B (Genée, 1998: 70).

In other words, given the emergence of some strategy  $S_2$  to express some function F, which coexists and ultimately supplants a previous strategy  $S_1$ , a functional-grammatical approach should be “able to account for the existence of these possibilities [...] in accordance with the principle of diachronic adequacy” (Genée, 1998: 76); diachronic adequacy as an explanatory principle means that our explanation of a given linguistic change must not presuppose stages that are not compatible with what valid typological constraints, i.e. stages in which there is a typologically impossible linguistic structure. The factors that lead the members of a speech community to eventually prefer one strategy over the other should, on the other hand, be “captured by a different, ‘social’ component of the theory, since the acceptance by the language users of one over the other options is codependent on extra-linguistic factors” (ibid.).

However, Genée (1998: 75) argues that a Functional Grammar account can only go as far as providing a “systematic description of possible paths for language change, i.e. [...] the system-internal aspects of diachronic processes”:

What we can probably not ask of a theory such as FG is a description of actual changes which includes extra-linguistic considerations such as prestige and other factors which ultimately lead speakers to adopt or ignore a given variant. [Genée, 1998: 75]

Accordingly, Genée does not develop a social component of FG, to do which would have been beyond the scope of her study, as it would be of ours; we will nevertheless try to formulate a number of hypotheses as regards external factors such as language contact (§7.5).

Another important set of observations found in the functionalist literature in relation to typological hierarchies and language change has to do with frequency effects. There are various ways in which the frequency of linguistic expressions may affect the structural development of language. Lexicalization patterns are a case in point: the more frequently a certain

semantic content needs to be expressed (given external factors such as the environment or the sociocultural circumstances in which a language is spoken), the higher the chance that a specific expression for it will find its way into the lexicon (Dik, 1997: 33).

Frequency of use has also been shown to correlate with morphosyntactic and phonological change: at the phonetic level, higher frequencies of use have a *reducing effect*: “high-frequency words and phrases undergo phonetic reduction at a faster rate than low- and mid-frequency sequences” (Bybee, 2006: 714). At the morphosyntactic level, on the other hand, high frequency has a *conserving effect*: “High-frequency sequences become more entrenched in their morphosyntactic structure and resist restructuring on the basis of productive patterns that might otherwise occur” (ibid., p. 715). Dik (1997: 44) observes that frequency of use correlates inversely with markedness: unmarked structures are more frequent and when an opposition is neutralized it is normally the unmarked member that is overgeneralized.

Dik explicitly intended FG as a cognitively adequate linguistic theory, ultimately capable of explaining how the “natural language user” *works*; this suggests that it should ultimately provide a model of the *processes* through which language users produce and understand linguistic expressions (Dik, 1997: 1). However, FG only describes the *patterns* that are found at the various levels of linguistic description and ignores the processual aspect (pp. 57f.); very few references are made to psychological and psycholinguistic studies. Therefore, according to Butler (2009: 7), FG can only provide an incomplete answer to the question that Dik himself regarded as ultimate; to get closer to a more satisfactory account, one should first and foremost reject the idea that linguist theory should concentrate on the grammar itself: a functional theory should model verbal interaction at large and be informed by research in such relevant areas as psycholinguistic and cognitive psychology, sociolinguistics and sociology, neurolinguistic and neurology (Butler, 2008: 4). In particular, it is necessary to take into account “[w]hat we know about the storage, production and reception of language and its elements” (ibid., p. 5).

Butler (2008: 6) looks at two influential processing models, Levelt’s (1989, 1999) and Kintsch’s (1988, 1998), and observes that both agree on one important point, namely the fact that processing is *incremental*. In a multi-level model of language production (to which we shall return be-

low), “incremental processing” means that each grammatical component is triggered into activity by a minimal portion of input from its feeding component (Levelt, 1989: 24–27), i.e. that any piece of information gets encoded as soon as its formulation is completed and, crucially, “before the entire communicative intention has been fully developed” (Hengeveld & Mackenzie, 2008: 24).

An extensive critique of Dik’s model from the point of view of its psychological plausibility is provided by Hesp (1990a, 1990b); here, I will limit the discussion to two important points highlighted by Butler (2008), namely that the FG model does not distinguish between conceptual and semantic representations, assuming instead that they coincide, and that its model of clause structure derivation is bottom-up.

A distinction between the two levels of representation, conceptual and semantic, is argued for by Hesp (1990b, 1990a), Nuyts (1992, 2001) and Levinson (1997). Levinson notes that conceptual representations are language-independent while semantic ones are language-specific, and that while thought is specific, its linguistic expression is often vague and imprecise depending on the quantity and quality of different, specific concepts that each language packs together into a single lexical expression.<sup>9</sup> Nuyts (2001) shows that representations of epistemic modality may be unitary at the conceptual level but may correspond, within the same language, to a number of different representations at the semantic level.

As regards the bottom-up model of clause structure derivation, Butler observes that if the predicate must be selected before the argument slots can be filled, then the speaker would be “unable to process material which will appear before the predicate until that predicate is known”, which would be unlikely from a psychological point of view (Butler, 2008: 13), and especially problematic in the case of SOV languages, since the theory does not allow movement.<sup>10</sup>

---

<sup>9</sup>An apt example is that while speakers of French might perceive “several grades of nuance” in the formality of interpersonal relationships, linguistically they cannot express more than two, having to choose between informal *tu* and formal *vous* (Hengeveld & Mackenzie, 2008: 47).

<sup>10</sup>Hesp (1990a) further notes that in Dik’s model of the hearer, parsing is also performed by first identifying the main predicate of a linguistic expression so that the other constituents of the linguistic expression can be interpreted based on the predicate frame of the main predicate; this, Hesp observes, is problematic: in SOV languages, for instance, “the parser cannot start parsing until all the words in the utterance have been phonetically

Functional Discourse Grammar (FDG) is a theory of linguistic interaction elaborated and refined over the years (cf. Hengeveld, 2004a, 2004b; García Velasco & Rijkhoff, 2008; Hengeveld & Mackenzie, 2008) with the goal of addressing the perceived inadequacies of FG, paying particular attention to psychological and pragmatic adequacy (Butler, 2008: 14). It rests on the assumption that no purely formal account can “clarify how linguistic structures co-vary with the purposes to which they are put in communication” (Hengeveld & Mackenzie, 2008: 27f.); at the same time, it is a structural-functional theory in that while accepting that “grammar is shaped by use”, it regards it as a system when considered in synchrony (ibid., p. 29). More precisely, FDG posits that the natural language user has knowledge both of the various kinds of linguistic units and of the way in which they can be combined; as a theory of grammar, FDG aims to describe this underlying knowledge (ibid., p. 26). Such knowledge is understood to be, to a great extent, stable, which permits cross-linguistic comparisons, and to be the result of historical processes: forms that prove useful and effective through the ages “[sediment] into the repertory” of forms that are synchronically available at any given time (ibid., p. 27)—i.e., they become routinized and enter the grammar or the lexicon. The existence of variation, both inter- and intra-linguistically, is recognized and assumed to exist within the limits imposed by communicative purposes and cognitive constraints (ibid.).

While functional models typically “seek to relate language form to language function” (Hengeveld & Mackenzie, 2008: 38), FDG’s approach is more complex: it is a *form-oriented* and *function-to-form* approach, in the sense that it posits “a range of functions flowing from the Speaker’s communicative intentions”, functions that are expressed in a language-specific way; and that as a theory it is only concerned with phenomena that are “reflected in morphosyntactic or phonological form” (ibid., p. 39). It also recognizes that natural languages often display structures which cannot be aligned with any particular function or communicative intention; these are regarded as “a-functional” i.e., “autonomous characteristics of the Morphosyntactic Level and/or the Phonological Level” (Hengeveld & Mackenzie, 2008: 40): for instance, whether a language prefers/disallows head–modifier as opposed to modifier–head order (ibid.), or the existence of certain forms of grammatical

---

processed” (p. 15), yet to identify the main predicate of an utterance the system must have already started parsing it (p. 29).

gender agreement (see §§4.2 and 7.6 below).

However, and most importantly, FDG is not concerned only with “grammar itself”. Alongside the grammatical component, its architecture includes a *conceptual component*, a *contextual component* and an *output component*, as discussed in §2.4. Furthermore, it is informed by psycholinguistic studies such as Levelt (1989) and (1999), where production is modelled as a process that goes from intention to articulation through formulation; it is in this sense a top-down model, meant to mimic the sequence found in production in an incremental fashion according to the two principles of *depth first* and *maximal depth* (Hengeveld & Mackenzie, 2008: 2–6). In other words, information is transmitted from the formulation levels to the appropriate encoding levels as soon as a minimal input for the latter is available (depth first), and any level may be by-passed if it is not involved in the production of the utterance (maximal depth), as will also be illustrated in §2.4.

However, FDG remains a model of *encoded* patterns rather than *encoding* processes: its aim is to “understand how linguistic units are structured in terms of the world they describe and the communicative intentions with which they are produced” (Hengeveld & Mackenzie, 2008: 2). Although Hengeveld (2004b: 367) remarks that a pattern model need not be a static one, and can be interpreted dynamically as “mirror[ing] the language production process in individual speakers”, Hengeveld & Mackenzie (2008) explicitly state that the FDG model is informed by process models but is not one itself (*ibid.*, p. 24); in other words, FDG is not a “blueprint for the speaker” like Levelt’s model (Levelt, 1989: 8–28). Specifically, this means that FDG is not concerned with describing the algorithms which convert the input into output within each component, or how these algorithms are run; therefore, such components as Working Memory, Syntactic Buffer and Articulatory Buffer, which are central to Levelt’s account of on-line speech production (Levelt, 1989: 26), are not modelled in FDG.

Otherwise, the similarities between the two models are evident, as will become clear in §2.4; but so are the differences. An important one is that FDG does not model language comprehension or self-monitoring, which Levelt (1989: 13f.) does. In Levelt’s model (cf. Figure 2.1), a component dubbed Speech-Comprehension System feeds the speaker’s own internal speech (mental representations of planned speech not yet uttered) and overt speech (uttered speech as heard by both the speaker and the hearer)

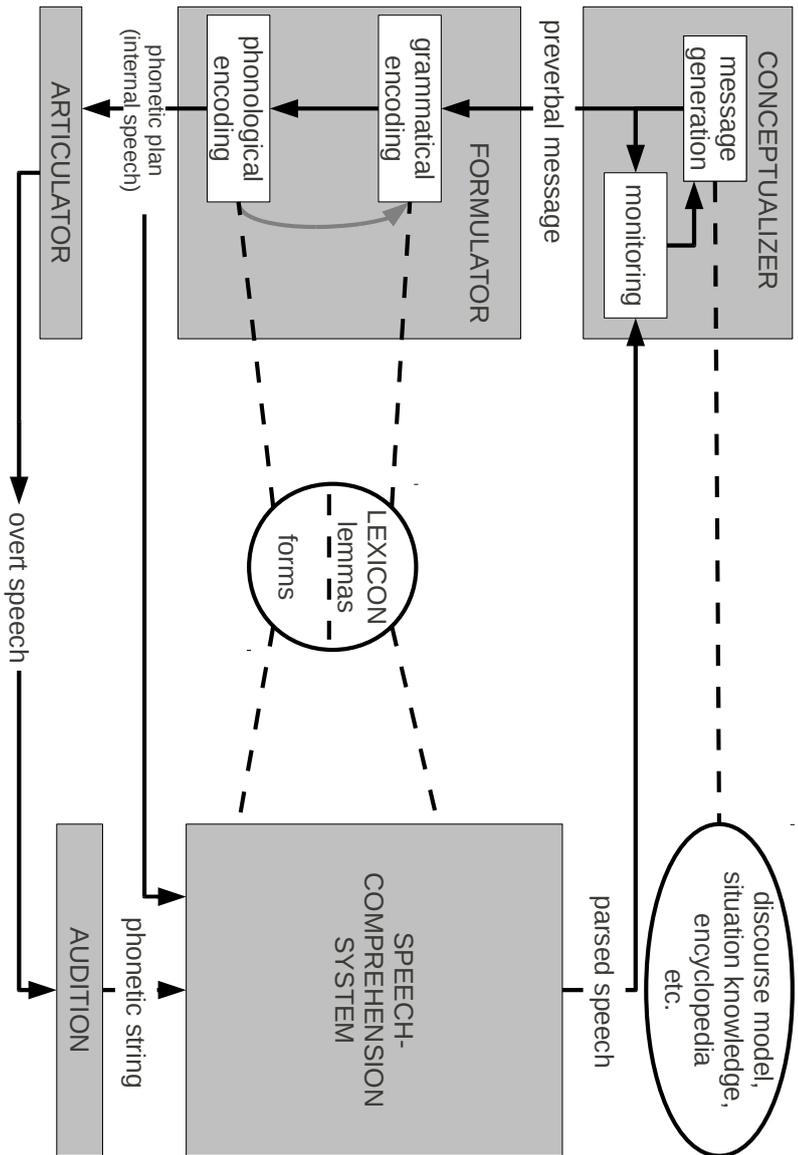


Figure 2.1: Levelt's "blueprint for the speaker"  
Boxes represent processing components, circle and ellipse represent knowledge stores (from Levelt, 1989: 9).

back into the Conceptualizer. This feedback loop allows the speaker, for instance, to produce self-corrections. In the FDG model, which does not include a comprehension system, this information loop is allowed but modelled in less specific terms by positing a constant exchange of information between the Grammatical Component and the Contextual Component: information from each level of representation contained in the former is fed into the latter, from which it can be fed back into the Grammatical Component in order to provide input for its various operations of formulation and encoding (see Figure 2.2 in §2.4 below).

## 2.4 An outline of Functional Discourse Grammar

### 2.4.1 Components, operations and representations

The model of FDG presented by Hengeveld & Mackenzie (2008) includes four components, namely the Conceptual, Contextual, Grammatical and Output Components. This is schematized in Figure 2.2, where circles represent operations (Formulation, Morphosyntactic and Phonological Encoding, Articulation) and rectangles the levels of representation resulting from each operation (Interpersonal, Representational, Morphosyntactic and Phonological Levels). Primitives and the output of Articulation are also represented by rectangles.

As is apparent from Figure 2.2, a detailed description of internal structures is given only for the Grammatical Component, but its interactions with the other components are also represented. The Conceptual Component is the primary source of input for the Grammatical Component: it is where conceptualization takes place and communicative intentions originate. Its output is fed into the Grammatical Component, where it is transformed into a linguistic representation. The Contextual Component stores “information received from the Grammatical Component concerning a particular utterance which is relevant to the form that subsequent utterances may take” (Hengeveld & Mackenzie, 2008: 9f.), e.g. a record of previous antecedents for the production of anaphoric chains (*ibid.*, p. 11).

The Grammatical Component has four different levels of processing, which are hierarchically organized as follows:

1. Interpersonal Level (IL),

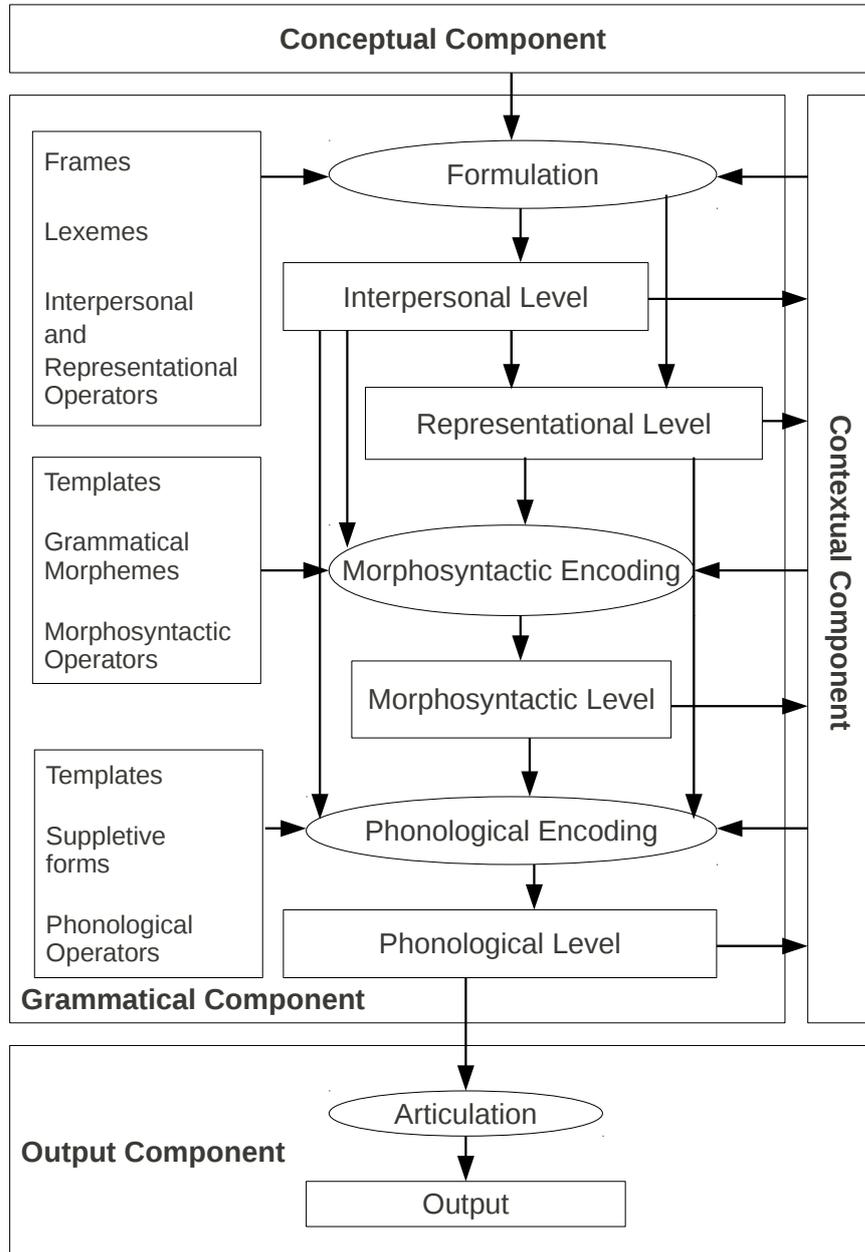


Figure 2.2: The architecture of FDG  
(from Hengeveld & Mackenzie, 2008: 13)

2. Representational Level (RL),
3. Morphosyntactic Level (ML),
4. Phonological Level (PL).

For ease of notation, these four levels are distinguished by different capitalization conventions: upper case for abbreviations pertaining to the Interpersonal Level, small case for the Representational Level, title case (e.g. Np, Vp) for the Morphosyntactic Level and small capitals for the the Phonological Level (Hengeveld & Mackenzie, 2008: 43).<sup>11</sup>

*Formulation* and *encoding* are the two types of operation involved in the top-down construction of the utterance. Formulation is the production of “underlying pragmatic and semantic representations”, while encoding is the conversion of said representations into morphosyntactic and phonological ones (Hengeveld & Mackenzie, 2008: 2).

Consistently with the criterion of typological adequacy,

[t]he rules used in Formulation are language-specific, i.e. FDG does not presuppose the existence of universal pragmatic and semantic notions. As a result, similar conceptual representations may receive different pragmatic and semantic representations in different languages. [...] This type of crosslinguistic variation may be expected to be governed by typological hierarchies, just like morphosyntactic and phonological variation. [Hengeveld & Mackenzie, 2008: 12]

The formulation module is fed the output of the Conceptual Component and produces two levels of representation: an IL representation, which pertains to the pragmatic functions of the final utterance (e.g. the type of illocution, the communicated content, acts of reference or ascription), and a semantic representation (RL), i.e. a representation of the communicated content in terms of its linguistic meaning distinctions (e.g. modality, polarity, location and time, quantification etc.).

---

<sup>11</sup>Representations at all these levels are intended as “purely linguistic in nature”, i.e., only those distinctions that are reflected in the grammar of a given language are part of the various representations (Hengeveld & Mackenzie, 2008: 15).

At the Interpersonal Level, the hierarchically highest unit of analysis is the Move, which is made up of one or more Discourse Acts;<sup>12</sup> each Discourse Act contains an Illocution, i.e. the specification of a relation between the Participants in the speech act (the Speaker and the Addressee) and some Communicated Content; the latter is comprised of one or more Subacts, which are either Referential or Ascriptive (Hengeveld & Mackenzie, 2008: 15). In accordance with the depth-first principle, linguistic expressions whose function directly maps onto a phonological form with no need for any further processing are sent immediately to the Output Component for articulation: this is the case of expressives (e.g. *wow*) and other invariable expressions (e.g. *thanks*, *congratulations*), whose phonological form is already available at the Interpersonal Level (*ibid.*, pp. 76–78).

The Representational Level pertains to the semantic representation of the utterance. The linguistic units of analysis distinguished at this level correspond to categories of meaning: the hierarchically highest unit is the Propositional Content, itself comprised of one or more episodes; each episode may in turn contain one or more descriptions of States-of-Affairs. States-of-Affairs contain one or more Properties, which may be predicated either of Individuals (e.g. first-order entities) or of other Properties (Hengeveld & Mackenzie, 2008: 15). The same category of meaning, e.g. a Property, may have different functions, i.e. it may be ascribed or referred to. The Property of being tall, for instance, is referred to by the noun *tallness* and may be predicated of some other Individual or Property by using the adjective *tall*: the formal difference between the noun *tall* and the adjective *tallness* pertains to the Morphosyntactic Level (*ibid.*, p. 16). Even Individuals, e.g. the assumedly specific referent of the phrase *my best friend*, may be ascribed or referred to, as in examples (1a) and (1b), respectively.

(1) English (Hengeveld & Mackenzie, 2008: 16)

- a. Sheila is *my best friend*. (Ascription)
- b. *My best friend* visited me last night. (Reference)

The Morphosyntactic Level is the first level of encoding. Here, linguistic

---

<sup>12</sup>The Discourse Act is considered the basic unit of analysis on the grounds of cross-linguistic evidence showing that a typologically valid grammatical theory must take into account linguistic units “larger than the individual clause” (Hengeveld & Mackenzie, 2008: 1, 3f.).

units are analyzed in terms of their syntactic constituents. Proceeding down the hierarchy from highest to lowest, these are Linguistic Expressions (Le), Clauses (Cl), the different types of Phrases—e.g. Noun Phrase (Np), Verb Phrase (Vp)<sup>13</sup> etc.—and the different types of Words that head the latter—e.g. Nominal Words (Nw), Verbal Words (Vw) etc. (Hengeveld & Mackenzie, 2008: 404).<sup>14</sup>

A Phrase may contain other Phrases, or Clauses, which within a Phrase may function either as modifiers or as arguments: cf. *the president's suite*, where the embedded Np *the president's* is a modifier of the matrix Np, vs. *the president's son* where the embedded Np is an argument of the relational Property expressed by *son* the adjective *presidential* (which can only be a modifier) may replace *the president's* in the first Np (*the presidential suite*) but not in the second (*\*the presidential son*) (Hengeveld & Mackenzie, 2008: 396).

The Phonological Level provides the representation of both segmental and suprasegmental characteristics of the Utterance (U), the largest and hierarchically highest phonological unit at this level; it contains one or more Intonational Phrases (IP); an IP may in turn contain one or more Phonological Phrases (PP), each of which is made up of one or more Phonological Words (PW) (Hengeveld & Mackenzie, 2008: 17f.).

## 2.4.2 Primitives

Formulation and encoding modules receive their input from stores of primitives, which in Figure 2.2 are represented as boxes on the left-hand side. At the level of formulation, the first type of primitives is represented by *frames*, which specify “the possible combinations of elements at the Interpersonal Level and the Representational Level for a certain language” (Hengeveld & Mackenzie, 2008: 19). The second type of primitives relevant to formulation is represented by Lexemes and the third type by IL and RL *operators* (ibid., p. 20). RL operators include linguistic categories like number and tense (ibid.), while an important IL operator is *identifiability*, which is con-

<sup>13</sup>Note that Vp is simply a phrase headed by a Vw, not “a layer of morphosyntax between the Phrase and the Clause” (as traditionally indicated by “verb phrase”), since FDG does not recognize such a layer (Hengeveld & Mackenzie, 2008: 299).

<sup>14</sup>Words are further decomposed into Morphemes which can be of three types: stems (s), roots (r), and affixes (Aff) (Hengeveld & Mackenzie, 2008: 404).

sidered a pragmatic rather than semantic operator, because it reflects the speaker's assumptions about whether the addressee will be able to identify a given referent (identifiability proper,  $\pm id$ ) and indicates whether the referent is identifiable for the speaker him/herself (specificity,  $\pm s$ ) (ibid., p. 122; Rijkhoff, 2002: 173, 227–229).

As regards encoding-level primitives, FDG posits three different types. The first type include language-specific sets of *templates*, morphosyntactic and phonological. Morphosyntactic templates specify the form of Linguistic Expressions, Clauses, Phrases and Words in terms of the combination and order of their constitutive elements; phonological templates do the same for Utterances, Intonational Phrases, Phonological Phrases, Phonological Words, Feet and Syllables (Hengeveld & Mackenzie, 2008: 20f.) The second type of encoding-level primitives includes *grammatical morphemes* and *suppletive forms*. Grammatical morphemes include elements such as auxiliaries, particles and affixes whose phonological form can already be determined at the Morphosyntactic Level because they require no further encoding. Those grammatical elements whose phonological form cannot be determined at this stage are introduced at the Morphosyntactic Level by *morphosyntactic operators*, which function as placeholders to be replaced at the Phonological Level by their actual forms, drawn from a set of suppletive forms: this is the case, for instance, when different allomorphs may be used. Morphosyntactic operators are themselves primitives: more specifically, they are the third type of primitives pertaining to the Morphosyntactic Level (Hengeveld & Mackenzie, 2008: 20f.). Finally, the necessity may arise in some languages for placeholders to be introduced at the Phonological Level as well (*phonological operators*): these are the third type of PL primitives and correspond to operations carried out by the Output Component, i.e. the Articulator, typically pertaining to intonation contours (Hengeveld & Mackenzie, 2008: 22).

In some important respects the model of FDG presented in Hengeveld & Mackenzie is informed by the work of Jan Rijkhoff (e.g. Rijkhoff, 2002, 2008), but there is an important terminological difference to be noted. Central in the parallel models of the clause and the noun phrase proposed by Rijkhoff is the distinction between *operators* and *satellites*, i.e. between grammatical and lexical modifiers, respectively. Grammatical modifiers (operators) are those which belong to closed paradigms; the distinctions they express—

unlike those expressed by lexical modifiers—tend to be absolute and non-gradable. An example of grammatical modifier is the article, which typically encodes a binary distinction ( $\pm$ id) (Rijkhoff, 2002: 85). Whereas lexical modifiers may express gradients (cf. ‘fairly heavy’ and ‘heavier’), grammatical modifiers may not: articles encode a noun as either identifiable or non-identifiable but may not be used to express degrees of identifiability (ibid., p. 86). In FDG, on the other hand, the term “operator” is used to designate a member of a specific inventory of primitives in the Grammatical Component. As just seen, FDG operators are (sets of) instructions that pertain to specific levels of formulation or encoding and may appear as placeholders in the corresponding representation; their phonological realization pertains to a successive level of encoding, where the placeholder is replaced by a phonological form from either a store of grammatical morphemes or a store of suppletive forms. Rijkhoff’s use of the term “operator” does not include the idea of instruction; instead, as we have seen, the term is applied to what in FDG would be called a grammatical element. So for instance using Rijkhoff’s terminology the article is an operator and an adjective is a satellite; in FDG, this distinction is captured by saying that the article is a Grammatical Word, whereas an adjective is an Adjectival Word at the Morphosyntactic Level, and a Lexeme at the Representational Level. In FDG, the operator in this context is (+id) at the Interpersonal Level, which is an instruction for encoding at the Morphosyntactic Level.

### 2.4.3 An example

As an example, Hengeveld & Mackenzie (2008: 23) provide a representation of the term *these bananas* (as in the sentence *I like these bananas*) at the four levels of analysis posited by the model.

(2) IL: (+id R)

RL: (prox m x: [(f: /bə'nɑ:nə/<sub>N</sub>(f)) (x)])

ML: (Np: [(Gw: this-pl (Gw)) (Nw: /bə'nɑ:nə/-pl (Nw))] (Np))

PL: (PP<sub>i</sub>: [(PW<sub>i</sub>: /ði:z/ (PW<sub>i</sub>)) (PW<sub>j</sub>: /bə'nɑ:nəz/ (PW<sub>j</sub>))] (PP<sub>i</sub>))

The IL representation captures the referential nature of the term (R) and its identifiability for the addressee as assumed by the speaker (+id). The RL representation means that reference is to more than one Individual (x) and

characterized in terms of location, as represented by the two operators plurality (m) and proximity (prox). The Property predicated of this Individual is lexically specified by a Nominal Lexeme (N), associated to a phonological form (/bə'nɑ:nə/). At the Morphosyntactic Level, the constituent is characterized as a Noun Phrase (Np) comprised of a Grammatical Word (Gw: this) and a Nominal Word (Nw).

Note that Lexemes and Words are not the same thing, nor is there a one-to-one correspondence between the two types of linguistic entities. Lexemes are RL constituents, Words ML constituents; one Word can correspond to several Lexemes, so for instance English *sword-swallower* is analyzed as one Nw at the Morphosyntactic Level and as a Configurational Property—with two distinct Lexemes (*swallow*, *sword*)—at the Representational Level; conversely, the English multi-word idiom *kick the bucket* corresponds to a single meaning unit (“to die”) at the Representational Level (Hengeveld & Mackenzie, 2008: 400). Also important is the fact that there are Words with no corresponding Lexeme, e.g. Grammatical Words, which are introduced at the Morphosyntactic Level in correspondence with abstract IL- or RL-features (ibid., p. 401).

At the Morphosyntactic Level, the final phonological representation of the Grammatical Word is not introduced because (this-pl) corresponds to a morphologically irregular (i.e. unpredictable) plural form: the ML operator pl(ural) represent an instruction for PL encoding. The plural form of *banana* is introduced via a placeholder for the plural morpheme (-pl), which being subject to allomorphic variation is determined at the Phonological Level.<sup>15</sup>

At the Phonological Level, finally, the plural forms of both Words are introduced: in the case of the Gw, by drawing on a repertory of suppletive forms which includes /ðɪs/ and /ðɪz/ as the two members of a paradigmatic opposition; in the case of the Nw, by adding the plural allomorph /z/. At all levels, a form is created by a rule, e.g. by appending a plural morpheme as in this case, only if there is no ready-made form stored in the lexicon, according to the *principle of lexical priority* (Hengeveld & Mackenzie, 2008: 459), which was formulated by Dik as follows:

<sup>15</sup>The choice of the allomorph /z/ (rather than, e.g., /s/ or /ɪz/) depends on the phonological context. Note that elsewhere Hengeveld & Mackenzie represent the insertion of the plural affix /s/ in the word *students* as taking place at the Morphosyntactic Level (Hengeveld & Mackenzie, 2008: 420), which looks like an oversight since in the preceding discussion (ibid., p. 418) they present it as a placeholder (Aff: Pl).

- (3) “Whenever a rule is encountered of the general form:  $M[X] = Y$ , where  $Y$  is the form of  $X$  under the operation  $M$ , first check the lexicon to see whether the  $M$  form of  $X$  is stored there ready-made. If so, select this form; otherwise apply the rule” (Dik, 1997: 345).

What this formulation explicitly suggests is that the outcome of morphosyntactic encoding (the ML representation) is scanned by the Phonological Encoder for strings which correspond to the specification of a suppletive form. For instance, if the ML representation contained the term (woman-pl), the Phonological Encoder would first check the list of suppletive forms for any item thus specified, and return /'wɪmm/; if a search in the list of suppletive forms failed to return anything, as would be the case for, say, (girl-pl), the word form /gɜlz/ would be regularly generated. What this also means, however, is that this search has to be carried out for *every single* term in the ML representation, which would seem to be an unnecessarily time-consuming exercise, considering that exceptions should, by definition, represent a small fraction of the lexicon. As a more economical alternative, we can hypothesize that Lexemes and Grammatical Words which inflect irregularly are flagged, and enter the derivation, as exceptional; being thus flagged, those elements are not assigned their phonological representation by the Morphological Encoder, and the aforementioned check is performed only on them. So, for instance, *this* and *woman* would be flagged as exceptional, while *banana* and *girl* would not. Arbitrarily adopting a superscript  $\epsilon$  as the symbol for exception, (2) can be represented as (4)

- (4) IL: (+id R)

RL: (prox m x: [(f: /bə'nɑ:nə/<sub>N</sub> (f)) (x)])

ML: (Np: [(Gw: <sup>ε</sup>this-pl (Gw)) (Nw: /bə'nɑ:nə/-pl (Nw))] (Np))

PL: (PP<sub>i</sub>: [(PW<sub>i</sub>: /ðɪ:z/ (PW<sub>i</sub>)) (PW<sub>j</sub>: /bə'nɑ:nəz/ (PW<sub>j</sub>))] (PP<sub>i</sub>))

Similarly, the derivation of the term *these women* would be represented as in (5).

(5) IL: (+id R)

RL: (prox m x: [(f: <sup>ε</sup>woman<sub>N</sub> (f)) (x)])

ML: (Np: [(Gw: <sup>ε</sup>this-pl (Gw)) (Nw: <sup>ε</sup>woman-pl (Nw))] (Np))

PL: (PP<sub>i</sub>: [(PW<sub>i</sub>: /ðiz/ (PW<sub>i</sub>)) (PW<sub>j</sub>: /'wimm/ (PW<sub>j</sub>))] (PP<sub>i</sub>))

I therefore suggest that the principle of lexical priority should be reformulated as follows:

(6) Whenever a rule is encountered of the general form:  $M[X] = Y$ , where  $Y$  is the form of  $X$  under the operation  $M$ , if  $X$  is marked as exceptional then first check the lexicon to see whether the  $M$  form of  $X$  is stored there ready-made. If so, select this form; otherwise apply the rule.

What this new formulation achieves is twofold: (i) it describes a more efficient model and (ii) it establishes a connection between the previously disjointed sets of Lexemes and of suppletive forms, clarifying their status as the two subcomponents of the lexicon. We can represent this as in Figure 2.3.

Finally, the PL representation consists of a Phonological Phrase (PP) which contains two Phonological Words (PW<sub>*i*</sub> and PW<sub>*j*</sub>) (Hengeveld & Mackenzie, 2008: 23).<sup>16</sup>

Both the Morphosyntactic and the Phonological Level have hierarchical structures and the former's map onto the latter's; even the terminology used to describe these two levels ("phrases", "words") suggests a close correspondence between them. But although mapping between the two levels tends to be on a one-to-one basis, the correspondence is not always straightforward and discrepancies are possible (Hengeveld & Mackenzie, 2008: 288, 426).<sup>17</sup>

<sup>16</sup>It is important to bear in mind that whereas the term "phonological word" applies to a word as defined using phonological criteria, e.g. stress placement, as in Dixon & Aikhenvald (2002: 13), "grammatical word" in Hengeveld & Mackenzie's terminology does not refer to a grammatically defined word but to any word in the ML representation which does not have a corresponding Lexeme and either corresponds to an operator or a function at the Interpersonal or Representational Level or is introduced as a placeholder (Hengeveld & Mackenzie, 2008: 410).

<sup>17</sup>For instance, the relationship between Morphosyntactic and Phonological Words becomes more complex when clitics are taken into account; clitics are treated as Words at the Morphosyntactic Level, where they are introduced by regular ML-templates, but as part of a larger Phonological Word at the Phonological Level (Hengeveld & Mackenzie,

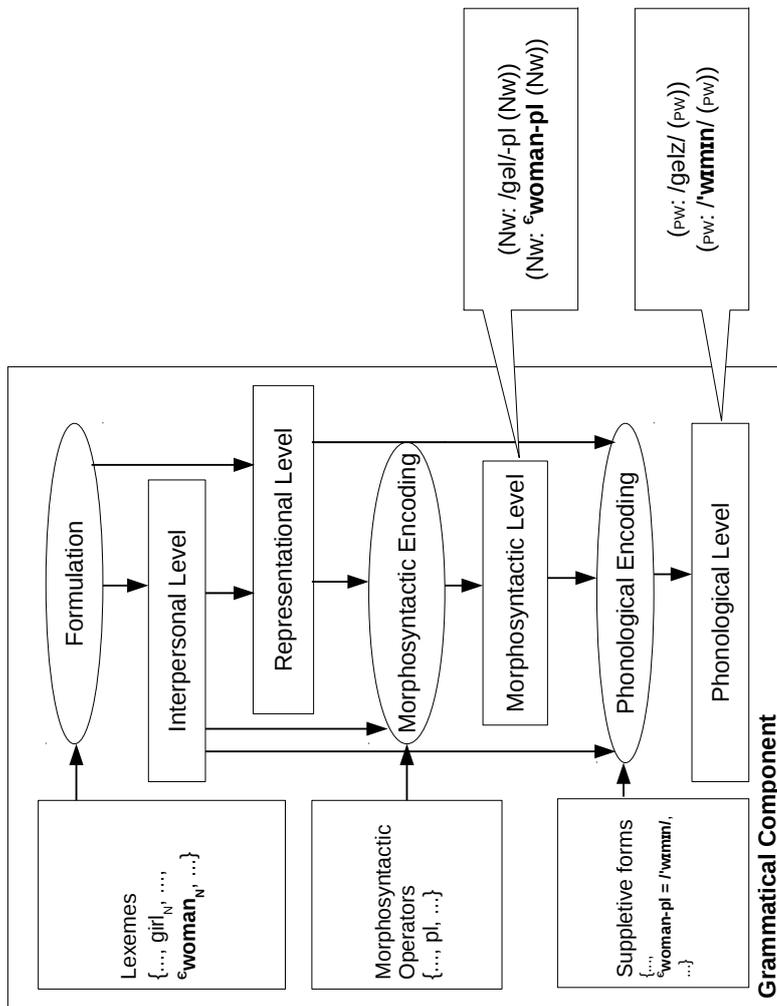


Figure 2.3: Lexemes and suppletive forms in FDG

#### 2.4.4 Non-linear speech production

FDG recognizes that speech production cannot always be modelled as a strictly left-to-right linearization process: the encoding of a segment may require access to the encoded shape of a subsequent segment, yet to be encoded; in such cases, it is assumed that the latter segment will be encoded first (“look-ahead”).

The variability of stress patterns in English provides a fairly clear example of non-linear encoding in the domain of prosody (cf. Levelt, 1989: 24f.). A word like *sixteen*, when pronounced in isolation, is stressed on the second syllable (*six'teen*), but if it is immediately followed by another stressed syllable (e.g. the first syllable of *'dollars*) it is stressed on the first syllable: so *six'teen* but *'sixteen 'dollars*. This means that for the correct stress pattern to be generated for the first word, the stress pattern of the second word must already be available (ibid., p. 25). Other articulatory-level phenomena like progressive assimilation and vowel harmony, whereby articulatory effort is reduced by anticipating the articulation of the following or subsequent segments, may also be taken as examples of look-ahead. Fortescue (2007: 340f.) observes that such contextually conditioned processes are often morphologized: he mentions examples of vowel harmony in Chukchi (a Chukotko-Kamchatkan language spoken in Northeast Siberia, Chukchi Peninsula; cf. Lewis, 2009), consonant gradation in Finnish, and *Umlaut* phenomena in Germanic languages. In Finnish, when a syllable of structure CV acquires a consonantal coda (i.e. structure CVC), the first consonant loses articulatory strength: e.g. *apu* (*a-pu*) ‘aid’, genitive *avun* (*a-vun*). This process does not apply if the following vowel is long: cf. the illative case of the same noun, *apuvun* (*a-pu:n*). In Icelandic, the quality of a root vowel may depend on the quality of the vowel in the ending: ‘cat’, for instance, is *kött-ur* in the nominative, *katt-ar* in the genitive and *kett-ir* in the plural, displaying *u-*, *a-* and *i-Umlaut*, respectively. He also provides examples of non-linear morphosyntactic processes, such as switch reference (Fortescue, 2007: 341), whereby a verb must receive a special marker if

---

2008: 446). Hengeveld & Mackenzie exemplify this with the expression *doesn't*. Its representation is (do-3sg.prs)(not) at the Morphosyntactic Level, /dʌznt/ at the Phonological Level; the form /nt/ for *not* is analyzed as part of the PL inventory of suppletive forms (alongside /ndt/); /dʌz/ and the clitic /=nt/ are joined together at the Phonological Level (ibid., pp. 458f.) and together they form one Phonological Word.





# Chapter 3

## Methodology

### 3.1 Overview

As stated in §1.1, the two main research goals of this study are (i) to characterize the system of gender agreement of Insular Celtic and (ii) to investigate the variation in gender agreement. In order to achieve the latter goal, a quantitative study was undertaken: the data sets and procedures used are explained in this chapter, in §§3.2 and 3.3, respectively.

### 3.2 The corpora

The data on which my analysis was carried out pertain to the colloquial register and come from two types of source:

1. interviews broadcast in the context of Irish- and Welsh-language radio shows;
2. dialectological studies.

Type-1 data were recorded in the last decade and represent contemporary usage; the material to include was selected on the basis of external criteria only, as recommended by best practice (cf. Sinclair, 2005), namely context (radio broadcast) and type of interaction (informal conversation). The size of the contemporary component of each corpus (in the region of 20,000 words—see below) approaches that of the “Broadcast Interviews”

component (20,000 words) in the design of the International Corpus of English.<sup>1</sup>

The use of broadcast material in linguistic inquiry is discussed by Labov (1972: 211) and applied to Welsh by Ball (1982). This type of data is representative of the language spoken in a semi-institutional and nationwide context, in which various types of speakers may be heard. Three types in particular are identified by Ó Giollagáin & Mac Donnacha (2008: 111f.):

- (a) “traditional” native speakers;
- (b) non-native (i.e. L2) speakers;
- (c) “non-traditional” native speakers, i.e., those whose parents are not traditional native speakers, but who still regard themselves as native speakers, having been raised through Irish (typically outside the Gaeltacht) and who have subsequently attended Irish-language schools.

To include non-native and “non-traditional” native varieties was deemed crucial on account of the considerable exposure that these varieties get through the media.

Type-2 data represent the variety of older generations and consist of two sets of recordings, involving speakers who were born around the beginning of the 20th century: a collection of recordings from Connemara, made in 1964 (Wigger, 2000) and a collection of transcribed recordings from a dialectological survey of Welsh (Thomas & Thomas, 1989). Because of the constraints imposed on this work in terms of time and resources, the choice of the material to include as type-2 was restricted to conversational recordings for which transcriptions were already available.

The data on traditional native speakers (older components) is meant to provide a term of comparison against which to quantify and qualify the degree of innovation exhibited by the gender system of the contemporary varieties, i.e. the varieties that are heard and accepted as contemporary Welsh and Irish, and not just the language spoken by traditional native speakers. As we shall see, this comparison yields interesting results and shows both continuity and innovation.

---

<sup>1</sup><http://ice-corpora.net/ice/design.htm>.

### 3.2.1 The Welsh corpus

The Welsh corpus includes two components. The contemporary Welsh component (CC-W) is a collection of interviews and informal conversations; it is entirely transcribed and freely available online as part of the Talkbank database, along with the original recordings.<sup>2</sup> For the present study, an hour and a half of spoken interaction (contained in five files) was analyzed.

The older component (OC-W) includes fourteen recorded conversations with different informants, along with their transcriptions, and runs for about one hour.

#### 3.2.1.1 The older component (OC-W)

OC-W contains the material published in the *Enghreifftiau* ('Examples') section of Thomas & Thomas (1989: 86–155). It includes 14 conversations, each with a different speaker and from a different location in Wales, totalling about 9,400 words. The speakers were all born between 1891 and 1922, except one speaker who was born in 1943, but whose data were included because they do not differ significantly, in pattern, from the other speakers'. His contribution is also quite limited. The date of the recordings is not indicated. The 14 speakers are listed in Table 3.1.

Each conversation is facilitated by an unnamed researcher, whose participation in the conversation is limited to a few lines per interview and has been ignored in the present analysis. Excerpts from this component are identified bibliographically and additionally by line number (e.g., "Thomas & Thomas, 1989: 86 l. 5" refers to page 86, line 5 of the transcribed text).

#### 3.2.1.2 The contemporary component (CC-W)

CC-W includes about 26,200 words and is comprised of five recordings from the BBC Welsh-language radio channel, Radio Cymru, broadcast in 2003 and 2004. There are 13 different speakers featured in this data set. Based on the estimates provided, they were born between 1945 and 1975. They are identified in Table 3.2.

Excerpts from this component are identified by the name of the file in the online repository and line number (e.g. "beti.cha 38" refers to line 38 of

---

<sup>2</sup><http://talkbank.org/data/BilingBank/Bangor/Pilot.zip>, last accessed 21 November 2010.

Table 3.1: OC-W speakers

| Year of birth | Speaker ID | Sex | Origin                            | Pages in Thomas & Thomas (1989) |
|---------------|------------|-----|-----------------------------------|---------------------------------|
| 1873          | MW         | F   | Glynogwr, Canol Morgannwg         | 151–155                         |
| 1891          | EJ         | F   | Pen-Caer, Penfro                  | 130–135                         |
| 1894          | RGT        | M   | Llangynwyd, Canol Morgannwg       | 146–150                         |
| 1900          | EIE        | F   | Blaepennal, Ceredigion            | 125–129                         |
| 1904          | EMH        | F   | Llanerch-y-medd, Môn              | 86–90                           |
| 1907          | MW         | F   | Llansawel, Dyfed                  | 136–140                         |
| 1911          | SD         | M   | Llansamnan, Goglewin Clwyd        | 96–100                          |
| 1911          | EP         | F   | Rhosllannerchrugog, Dwyrain Clwyd | 106–110                         |
| 1912          | FT         | M   | Carno, Maldwyn                    | 121–124                         |
| 1913          | CL         | M   | Treforys, Goglewin Morgannwg      | 141–145                         |
| 1914          | AJ         | F   | Llanyrnawddwy, Maldwyn            | 111–115                         |
| 1917          | OD         | M   | Ffymnongroyw, Dwyrain Clwyd       | 101–105                         |
| 1922          | EnE        | F   | Bryn-crug, Meirionnydd            | 116–120                         |
| 1943          | GWJ        | M   | Caernarfon, Gwynedd               | 91–95                           |

the file `beti.cha`). All examples from this corpus are presented as originally transcribed.

### 3.2.2 The Irish corpus

The Irish corpus also includes two components. A corpus of present-day Irish usage comparable to the Welsh one was not available and was built as part of this study using the same type of sources. The sources were both Irish-language radio stations and Irish-language shows broadcast by English-language stations.

The transcription of the Irish data set was carried out with the aid of proficient speakers of Irish. This component contains excerpts from eleven radio shows, featuring interviews and conversations with different guests and conducted by different presenters, and runs for approximately one hour and 40 minutes. Because there are in Ireland statutorily defined Irish-language districts (the *Gaeltachtaí*), this component includes both data from Dublin-based radio stations and data from the Gaeltacht-based Raidió na Gaeltachta.

The older generations' variety is represented by a selection of the transcriptions published by Wigger (2000); the corresponding audio files are freely available from the website of the Dublin Institute for Advanced Studies.<sup>3</sup>

#### 3.2.2.1 The older component (OC-I)

OC-I is comprised of six texts, totalling 12,200 words approximately. All recordings were made as informal conversations in 1964 and all speakers were males from Ros Muc in the Connemara Gaeltacht. All were born between 1896 and 1937. Going from oldest to youngest, the seven speakers were the following (with age of speaker in 1964):

1. 68-year-old M. Ó C. [M], boatman;
2. 54-year-old S. Ó M. [S], farmer;
3. 54-year-olds M. Ó N. [Ma], farmer;

---

<sup>3</sup>[http://www.celt.dias.ie/publications/online/caint\\_ros\\_muc/](http://www.celt.dias.ie/publications/online/caint_ros_muc/), last accessed 21 November 2010.

Table 3.2: CC-W speakers

| Age group in 2003–2004 | Group size | Speaker ID | Sex | Origin or dialect spoken             | Occupation                   |
|------------------------|------------|------------|-----|--------------------------------------|------------------------------|
| 30–40                  | 4          | BBA        | M   | Ceredigion, West Wales               | radio presenter              |
|                        |            | DAF        | M   | Arfon, North-West Wales              | radio presenter              |
|                        |            | GAY        | F   | speaking a northern dialect of Welsh | radio presenter              |
|                        |            | RHY        | M   | speaking a northern dialect of Welsh | camera man                   |
| 40–50                  | 3          | CAT        | F   | uncertain                            | writer and radio presenter   |
|                        |            | LIM        | F   | Eifionydd, North-West Wales          | farmer                       |
|                        |            | DYL        | M   | Conwy Valley, North Wales            | radio presenter              |
| 50–60                  | 4          | KIL        | F   | Amman Valley, South Wales            | community worker and manager |
|                        |            | JON        | M   | Arfon, North-West Wales              | radio presenter              |
|                        |            | SIA        | F   | speaking a southern dialect of Welsh | radio presenter              |
|                        |            | RHM        | F   | brought up in North Wales            | film industry                |
| 60–70                  | 2          | BET        | F   | Ceredigion, West Wales               | radio presenter              |
|                        |            | MIK        | M   | speaking a northern dialect of Welsh | retired stone mason          |

4. 52-year-old S. Ó M. [Se], farmer;
5. 50-year-old M. B. [Mi], operator;
6. 27-year-old T. Ó C. [T], teacher;
7. 27-year-old T. Ó M. [To], electrician.

The 27-year-old teacher acted as a session facilitator by asking questions and inviting other speakers to take part in the conversation. Overall, his own contribution is very limited.

Every excerpt from this component is identified by a bibliographical reference (Wigger, 2000) followed by the three-number code given to it by the corpus editor, e.g. 4-01-02. The first number (4 in this example) refers to the conversational session (referred to as *caint* by Wigger, 2000) in which the excerpt is found; the second (01 in the example) to a specific part in the *caint* (each part has a title which refers to the topic of the conversation); the third number (02 in the example) identifies the stretch of discourse in which the example is found (this subdivision is found in Wigger (2000) and refers to stretches generally not exceeding one minute in duration).

### 3.2.2.2 The contemporary component (CC-I)

CC-I contains eleven texts and about 20,000 words. It includes seven texts from Raidió na Gaeltachta (RnaG, approximately 13,000 words, broadcast in 1997 and 2007) and four from non-Gaeltacht sources (approximately 7,000 words, broadcast in late 2006 and 2007). The latter include two Dublin-based radio stations (Raidió na Life and Newstalk) and one podcast from the series *An Líonra Sóisialta* ('The Social Network'), produced in Ennis, County Clare.<sup>4</sup> There are 25 speakers, whose details are provided in Table 3.3. The recordings are representative of two of the three macro-dialects of Irish, Munster and Connacht; Ulster Irish is not represented.

Excerpts from CC-I are identified by citing the source (radio station and programme name) and the date of the recording in the format dd/mm/yyyy (e.g. RnaG "Cumarsáid" 06/11/1997).

---

<sup>4</sup><http://anlionra.com/mp3/lionra-2006-11-14.mp3>, last accessed 21 November 2010.

Table 3.3: CC-I speakers

| Year of birth<br>(estimate) | Group<br>size | Speaker<br>ID | Sex | Origin or dialect spoken                          | Occupation                               | Source    |
|-----------------------------|---------------|---------------|-----|---|--|-----------|
| 1925-'35                    | 1             | TOI           | M   | Co Kerry (Munster)                                | returned emigrant                        | RnaG      |
| 1936-'45                    | 2             | TOF           | M   | uncertain   | actor                                    | RnaG      |
|                             |               | MDM           | M   | Co Kerry (Munster)                                | manager                                  | RnaG      |
| 1946-'55                    | 6             | MNC           | F   | Co Kerry (Munster)                                | radio presenter at RnaG                  | RnaG      |
|                             |               | CTL           | F   | Co Meath, but speaking a variety of Munster Irish | secretary                                | RnaG      |
|                             |               | NNC           | F   | Co Galway (Connacht)                              | radio presenter at RnaG                  | RnaG      |
|                             |               | TMG           | M   | Co Kerry (Munster)                                | entrepreneur                             | RnaG      |
|                             |               | HNS           | F   | Speaking a variety of Munster Irish               | radio presenter at RnaG                  | RnaG      |
|                             |               | MOS           | M   | Speaking a variety of Munster Irish               | radio presenter at RnaG                  | RnaG      |
| 1956-'65                    | 7             | ANI           | F   | Co Clare (Munster)                                | n/a                                      | RnaG      |
|                             |               | MCL           | M   | Speaking a Connacht dialect                       | events organizer                         | Newstalk  |
|                             |               | DDF           | M   | Galway (Connacht)                                 | actor, writer                            | RnaG      |
|                             |               | CLS           | F   | Speaking a variety of Munster Irish               | cook                                     | RnaG      |
|                             |               | FOG           | M   | Cork (Munster)                                    | social worker                            | RnaG      |
|                             |               | CNM           | F   | Carlow (Leinster)                                 | barrister and journalist                 | RnaG      |
|                             |               | AHY           | F   | uncertain   | radio presenter at RnaG                  | RnaG      |
| 1966-'75                    | 8             | COM           | M   | Co Clare (Munster)                                | broadcaster, media entrepreneur          | An L onra |
|                             |               | JBN           | M   | uncertain   | academic researcher in Galway            | An L onra |
|                             |               | AN2           | F   | Co Cork (Munster)                                 | former nurse                             | RnaG      |
|                             |               | RSH           | F   | uncertain   | radio presenter in Dublin (RnaI)         | RnaI      |
|                             |               | SND           | F   | uncertain   | transport and traffic engineer in Dublin | RnaI      |
|                             |               | DOT           | M   | Dublin  | radio presenter in Dublin (Newstalk)     | Newstalk  |
|                             |               | RNL           | F   | Speaking a variety of Munster Irish               | receptionist                             | RnaG      |
| 1976-'85                    | 1             | COL           | M   | Co Kerry (Munster)                                | bookshop assistant                       | RnaG      |
|                             |               | NOL           | M   | uncertain   | works for a French TV channel            | Newstalk  |

### 3.3 Analysis of the data

#### 3.3.1 Data retrieval and statistical testing

After each text was checked against the corresponding recording to make sure the relevant details had been transcribed faithfully, the data were scanned manually to identify and classify the relevant diagnostics, which are the following:

1. initial mutation of the head noun after the article and of the adjective(s) after the head noun;<sup>5</sup>
2. initial mutation after a third-person possessive pronoun (anaphoric agreement);
3. form of non-possessive third-person singular pronouns (anaphoric agreement);
4. form of the adjective (local agreement);
5. form of the article (local agreement, Irish only);
6. form of the numerals 2–4 (in both local and anaphoric agreement; Welsh only).

Potential contexts for agreement were recorded and classified according to whether (i) the expected form was found, (ii) an unexpected form was found, or (iii) it was impossible to decide (e.g. if the form was unintelligible). Each observation was recorded in a database.

All observations classified as uncertain or ambiguous were discarded from subsequent analysis, which was carried out as follows:

1. the agreement rate for each target type was calculated as the proportion of realizations of grammatical gender agreement out of the total number of observations. With pronouns, anaphoric chains can be found in which a single antecedent is shared by several pronouns, as shown in Figure 3.1); when this happens, “total” refers to the number of anaphoric relations that are thus established, “agreeing” to the number of relations in which the pronoun agrees grammatically.

---

<sup>5</sup>The tense/lax opposition through which the lenition of Irish sonorants *n-*, *l-* and *r-* was realized has now been lost by all but perhaps the oldest and most conservative speakers of certain areas of Munster (see §5.4.1.1). Lenition of these segments was therefore ignored.

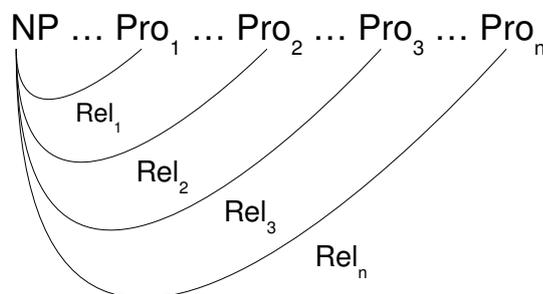


Figure 3.1: Anaphoric relations

2. Within each corpus component, statistical tests were carried out to determine if there were significant differences in agreement rate, either across different target types or within a certain target type (e.g., for initial mutation, different mutating consonants and different types of mutation; for nouns, whether they were animate or inanimate; for pronouns, whether they were stressed or unstressed forms, prepositional or non-prepositional, etc.).
3. Within each language, in diachrony, statistical tests were conducted to assess if the results obtained from the previous step were consistent.
4. Cross-linguistically, statistical tests were conducted to ascertain whether comparable agreement strategies and/or targets (e.g. agreement by mutation vs. agreement by inflection, local vs. anaphoric, etc.) were treated consistently.

The statistical test chosen for the difference between agreement rates is Fisher's exact test, since some comparisons are based on small sample sizes (cf. Baayen, 2008: 113). The software I used to perform this test (R<sup>6</sup>) returns the following statistics:

1. an estimate of the odds ratio for the difference, i.e. how many times more likely it is for the event under consideration (the occurrence of grammatical agreement) to be observed in one group than in the other;

<sup>6</sup>See <http://cran.r-project.org/>, last accessed 21 November 2010.

2. a 95% confidence interval for this ratio, which covers the true, i.e., long-run, odds ratio 95% of the time;
3. a  $p$ -value, i.e. a measure of the probability of observing an odds ratio different from 1 if its long-run value were in fact 1, i.e. if there were no real difference.

A difference was regarded as statistically significant for  $p < 0.05$ . When a difference is significant but small, or when  $p$  approximates 0.05, the confidence interval (c.i.) for the odds ratio is also presented: a c.i. which straddles 1 (e.g. 0.5 to 1.5) means that there might be no statistically significant difference between the two agreement rates being compared.

### 3.3.2 Proper names

Differential analyses were carried out for NP-internal target types (e.g. the article and the adjective) to make sure that agreement rates were not inflated by the presence of proper names. It is reasonable to hypothesize that proper names (e.g. personal or place names), constitute unanalyzed wholes that speakers learn and store in the lexicon as such. Two examples from CC-I are given in (1) and (2), a surname and place name, respectively.

- (1) Irish (RnaG “Seal Aneas” 21/10/1997)

Bean Mac                      Na                      Mara  
 Mrs    son(M).NOM.SG    ART.GEN.F.SG    sea(F).GEN.SG

‘Mrs Mac Na Mara’

- (2) Irish (RnaG “I Measc na nDaoine” 22/02/2007)

An                              Ghráig  
 ART.NOM.F.SG    village(F).NOM.SG

If these forms are actually stored in the lexicon as unanalyzed units, then we might argue that any agreement marker we find within the boundaries of such a noun phrase should not be regarded as actually representing productive agreement. All statistical tests for local agreement were therefore carried out a second time after discarding proper names.

### 3.3.3 Semantic vs. non-semantic agreement

A further distinction was made between semantic and non-semantic agreement. An observation was counted as representing semantic agreement if information (lexical or otherwise) about the sex of the referent (when relevant) was available to the speaker; otherwise it was counted as representing non-semantic agreement.

### 3.3.4 Types of pronouns and phoric relations

Pronouns were divided into three groups:

1. simple pronouns;
2. prepositional pronouns;
3. possessive pronouns.

Simple and prepositional pronouns differ from possessive pronouns because the latter are always unstressed clitics and mark gender and person agreement (with their antecedent) by triggering different initial mutations on the following word, as discussed in chapter 5. Prepositional pronouns in turn differ from simple pronouns in that they represent the coalescence of prepositions and personal endings; syntactically, they are prepositional phrases. Examples of prepositional pronouns are also to be found in chapter 5. This tripartition is therefore justified on the one hand by the different agreement-marking strategies employed and on the other by Ó Curnáin's (2007b: 1261) observation that in the particular variety of Irish he describes the generalization of masculine anaphors with feminine inanimate antecedents is more likely when the pronoun is prepositional.<sup>7</sup>

The investigation of gender agreement with prepositional pronouns in Irish was limited by the fact that it is often impossible to distinguish between *ann* as a prepositional pronoun (the third-person singular form of *i* 'in') and

---

<sup>7</sup>This would be at variance with the typological generalization that semantic agreement is more likely in the nominative than in oblique cases at any given point of the Agreement Hierarchy (cf. Corbett, 1991: 238).

*ann* as a deictic adverb ('there'), which does not involve agreement.

(3) Irish

níl                    aon duine ann  
NEG.be.PRS one person in.3SG.M/there

'nobody is in it/nobody is there'

It is almost always possible to interpret *ann* as meaning simply 'there'. The only contexts in which *ann* can be unequivocally identified as a pronoun is when it co-refers with both its antecedent and an intervening relative pronoun. There are two structures in which this occurs: (i) when the relative pronoun expresses an indirect (prepositional) argument of the verb in the relative clause and the prepositional pronoun (*ann*) co-refers with it, as in example (4), and (ii) in the predicative construction NP + *atá ann/inti*, exemplified in (5).

(4) Irish (Wigger, 2000 7-01-02)

duine eolais,            ar an                    abhainn<sub>i</sub>                    a<sub>i</sub>  
person knowledge.GEN on ART.PREP.SG river(F).PREP.SG REL  
ngabhfá                    ann<sub>i</sub>  
go.COND.2SG in.3SG.M

'someone familiar with the river that you would be going in'

(5) Irish

a. (Wigger, 2000 1-01-18)

Nár                    mhaith an                    fear<sub>i</sub>                    a  
NEG.COP.PST good ART.NOM.M.SG man(M).NOM.SG REL  
bhí                    ann<sub>i</sub>?  
be.PST in.3SG.M

'Wasn't he a good man?'

b. (Wigger, 2000 7-01-06)

[oíche                    Sathairn]<sub>i</sub>                    a bhí                    inti<sub>i</sub>  
night(F).NOM.SG Saturday(M).GEN.SG REL be.PST in.3SG.F

'it was Saturday night'

Therefore, only occurrences of *ann* in these two types of structures were counted.

As regards the distinction between anaphoric and cataphoric pronouns, Ó Curnáin (2007b: 1260f.) reports that in the Irish of Iorras Aithneach cataphora is a favourite context for the generalization of masculine pronouns. However, cataphoric relations with feminine antecedents are extremely rare in our corpora (only two were found in CC-I, just one in OC-I, and none in the Welsh corpus). None of them, in line with Ó Curnáin's observation, shows agreement with the co-referential noun phrase.

### 3.3.5 Complex anaphora

Excluded from the count are all instances of *complex anaphora* (in the sense of Consten et al., 2007, i.e. an expression with a propositionally structured antecedent and a conceptually complex referent). The antecedent of a complex anaphor is not lexical and has therefore no inherent grammatical gender; by default, complex anaphora is expressed by masculine pronouns, but feminine pronouns may be found if there is attraction in the gender of the predicate; the following example is from Irish.

(6) Irish (Wigger, 2000 1-01-01)

T: Bhoil anois, is faoi is mó a bhí muid ag caint an oíche cheana faoin mbádóireacht, ar ndóigh, 'sí is gaire do do chroí.

M: 'Sí ar ndóigh an chéad *turn*...

T: Agus, t'athair, agus is dóigh d'athair mór romhat.

M: Muise 'sin í an fhírinne.

*Translation:*

T: Well now, that's what we were talking about the other night for the most part, boating, of course, you're very fond of it.

M: I am, of course ...

T: As was your father, and of course your grandfather before you.

M: Indeed, that is the truth.

In (6), *sin* 'that' is a complex anaphor: it is used by the speaker to refer to a statement of which he is predicating the truth value. The pronominal form *í* is semantically pleonastic, but syntactically required in this context (because the predicate is expressed by a definite noun phrase: *an fhírinne* 'the truth'). The pronoun form is feminine by attraction in the gender of the predicate noun phrase, which is headed by a feminine noun (*fírinne*). By way of contrast, if the predicate were expressed by an adjective (e.g.

*fíor* ‘true’), a different verb would be used, with a different structure: *tá sé sin fíor* ‘that’s true’, with the complex anaphor represented by the constituent [*sé sin*] (literally, 3SG.M + distal marker: ‘that’). According to de Bhaldraithe (1953: 138), who describes another dialect of the same area as that of Ó Curnáin (2007c), gender attraction in these cases is a matter of lexical distribution: it is the norm with certain feminine nouns, e.g. *ceist* ‘question’, *cúis* ‘cause’, *fírinne* ‘truth’ etc. (so *’sin í an cheist*, *’sin í an chúis* etc.), but not with other feminines, e.g. *caoi* ‘way’, *tuairim* ‘opinion’, etc. (*’sin é an chaoi* etc.). Other studies of traditional County Galway varieties show that masculine pronouns are common in this type of constructions, regardless of the gender of the predicate noun phrase (cf. Ó Curnáin, 2007b: 1260f.).

### 3.4 Summary

In this chapter I described the data sets on which my analysis is based and their internal structure (§3.2), and discussed the methodology employed to analyze them (§3.3).

The study of agreement variation is conducted on two data sets, the Irish corpus and the Welsh corpus, each comprised of two components, referred to as “older” and “contemporary” (OC and CC, respectively). OC-speakers were born in the last decades of the 19th century or the early decades of the 20th and were recorded in the context of dialectological studies during the 20th century; CC-speakers were born for the most part after 1950 and were recorded as they took part in broadcast conversations on the radio. The purpose of this comparison was to establish if contemporary varieties are different from traditional ones in relation to gender agreement.

Variation was analyzed statistically: agreement rates for the different target types were calculated and compared within the same component and across components.

The findings of this analysis are presented in chapters 6 and 7.



## Chapter 4

# Gender: A typological survey

### 4.1 Overview

Grammatical gender represents a fascinating and multifaceted phenomenon and is quite widespread in the languages of the world (Aikhenvald, 2000: 77–80; Corbett, 1991: 1f.). The literature on the subject is extensive. Comprehensive typological studies include Aikhenvald (2000), Corbett (1991), and Craig (1986).

In this chapter, a particular emphasis will be placed on the Indo-European dimension of this phenomenon. The chapter is structured as follows: §§4.2 and 4.3 contain a few preliminary observations on gender, while its semantic and formal correlates are examined in §4.4, where two different views on assignment are also outlined. Finally, agreement is considered in §4.5 from the point of view of its realization in the Indo-European language family.

### 4.2 The gender puzzle

Students of language have been fascinated by grammatical gender for quite some time; Meillet (1948: 202) called it one of the “least logical and most unexpected” grammatical categories. The familiar terms “masculine” and “feminine” clearly refer to the biological gender of the referents, “neuter” (the Latin for ‘neither’) being characterized by opposition as neither mas-

culine nor feminine.<sup>1</sup> In the nineteenth century, Jacob Grimm's *Deutsche Grammatik* (mentioned in Zubin & Köpke, 1986) still described grammatical gender as resulting from the conceptualization of things and natural phenomena as masculine or feminine.

The view that grammatical gender cannot be explained in semantic terms is expressed by Hjelmslev (1970: 212), and Comrie (1999) observes that

[g]ender as a grammatical category must [...] be kept distinct from sex as a semantic, or perhaps conceptual, category. Turkish has no grammatical gender, but the sentence *kiral gebedir* "the king is pregnant" is as odd in Turkish as its translation is in English; compare *kiralıçe gebedir* "the queen is pregnant," which lacks this anomaly in both languages. [Comrie, 1999: 458]

Nevertheless, there seems to be experimental evidence (see e.g. Boroditsky et al., 2003) that grammatical gender can influence the way in which speakers of different languages conceive of inanimate objects by projecting onto them semantic associations with perceived masculinity and femininity.

Gender is defined by agreement: since gender distinctions need not be marked on the noun, evidence that such distinctions exist can only come from agreement (Corbett, 1994: 1348). In other words, we can only tell that two nouns belong to two different genders if, all other things being equal, other elements which are in some grammatical relation to them consistently appear in different forms. In Italian, for instance, *ponte* 'bridge' and *fonte* 'fountain' belong to different genders, as shown in the following example.

(1) Italian

- a. il        ponte    antic-o    /    modern-o  
    the.M    bridge    ancient-M /    modern-M  
    'the ancient / modern bridge'
- b. la        fonte    antic-a    /    nuov-a  
    the.F    fountain    ancient-F /    modern-F  
    'the ancient / modern fountain'

<sup>1</sup>In the fourth century BC, Aristotle observed that nouns for inanimate referents could be of any gender and referred to those that were neither masculine nor feminine as *tà metaxý*, literally "those in between" (see Lucas, 1968: 206f.; Robins, 1997: 32f.).

This characteristic is linked to the debated question whether grammatical gender has a function, on a par with other nominal categories like case and number.<sup>2</sup> Unlike case and number, gender cannot always be interpreted semantically and is defined by agreement, as we have just seen. For this reason, an oft-given answer to the functional question is that gender contributes to the success of communication by helping language users keep track of referents through pronominal agreement (Lehmann, 1988; Zubin & Köpke, 1986; Corbett, 1991, 2006), as in example (2).<sup>3</sup>

(2) German (adapted from Zubin & Köpke, 1986: 174)

a. Der Krug<sub>i</sub> fiel in die Schale, aber  
 ART.M.SG jug(M).SG fall.PST.3SG in ART.F.SG bowl(F).SG, but  
er<sub>i</sub> zerbrach nicht.  
 3SG.M break.PST.3SG NEG

‘The jug fell in the bowl, but it (the jug) didn’t break.’

b. Der Krug fiel in die Schale<sub>j</sub>, aber  
 ART.M.SG jug(M).SG fall.PST.3SG in ART.F.SG bowl(F).SG, but  
sie<sub>j</sub> zerbrach nicht.  
 3SG.F break.PST.3SG NEG

‘The jug fell in the bowl, but it (the bowl) didn’t break.’

In (2) the German expression leaves no doubt as to what did not break: the jug in (2a), the bowl in (2b). The English anaphoric pronoun *it*, on the other hand, is potentially ambiguous (although it would probably be interpreted by default as coreferential with the first subject).

To this attractive hypothesis two objections are possible: on the one hand—as pointed out by a number of authors (e.g. Leiss, 1999; Vogel, 1999; Weber, 1999)—number and case are also interested by agreement, but it is hardly ever claimed that their function is to facilitate reference tracking, presumably because they are normally assigned a semantic interpretation. On the other hand, it can be objected that if the two nouns in example (2)

<sup>2</sup>In Corbett’s (1991) monograph on the subject, only two pages out of 323 are devoted to the functional issue (as already noted by Trudgill, 1999: 140).

<sup>3</sup>It has been suggested that the function of agreement is to introduce redundant information into the discourse in order to facilitate comprehension when there is “noise” in the communicative channel, and to mark constituency, useful if the elements of a syntactic constituent are not adjacent to one another, as in languages with no fixed word order (see Corbett, 2006: 274f. and references therein).

were of the same gender, the gender of the pronoun could not disambiguate reference; this is an entirely possible and indeed probable scenario in German, where there are only three genders. Therefore, the contexts in which reference tracking is made easier by gender agreement may well turn out to be rare; in this case, we could ask if the processual load imposed by gender agreement is actually worth the trouble (Trudgill, 1999).<sup>4</sup>

As Trudgill (1999) also observes, gender is not found in pidgins—which are characterised by greater regularity and less redundancy than native varieties—and, remarkably, it is not reintroduced during the creolization process, when a pidgin becomes a native variety:

[Creolisation] involves the reintroduction of many of the grammatical categories that have been lost during pidginisation. [...] We can therefore assume that native speakers “need” some of these grammatical categories and hence reintroduce them. [...] However, and vitally for our discussion, there is not a creole language in the world which has reintroduced, during the expansion process, the category of grammatical gender on nouns or verbs. Typical creoles, moreover, do not even demonstrate natural gender. [...] *A very strong inference is that gender is a category that languages and their speakers can more readily do without than many or most other categories.* [Trudgill, 1999: 136f., emphasis mine.]

Trudgill concludes that gender is an *afunctional* category whose appearance can only be explained in historical terms as the outcome of “a series of grammaticalisation processes” and whose persistence is due to the “very high degree of frequency” with which it occurs in gendered languages (Trudgill, 1999: 148f.), i.e. to its conceptual *entrenchment* (Langacker, 1987: 59); in other words, gender is a phenomenon which has “an explanation but no function” (Trudgill, 1999: 149).

What then could be the historical explanation for the emergence of grammatical gender? Leiss (1999) suggests that, in Indo-European, gender is

<sup>4</sup>Singer (2010) describes a gender system, that of the Australian language Mawng, in which gender is used both to establish and to track reference, in a way that is more typical of classifier than gender systems. However, Singer herself stresses the exceptionality of this system—so far unattested anywhere else—which combines the formal correlates of grammatical gender with the functional versatility of a classifier system.

what is left of a morphological system which originally encoded properties of the referent such as mass distribution, configuration, individuability and so on, productively and in a derivational fashion. In the Indo-European system reconstructed by W. P. Lehmann and presented by Leiss, gender was a category with a full, three-way paradigm: virtually every stem could take all three inflections (“masculine” *-s*, “feminine” *-h* and “neuter” *-m*), thereby defining and perspectivizing its basic meaning. For instance, the stem *hima-* (with its basic idea of ‘cold’) could appear as feminine *\*hímā* (< *\*hima-h*) ‘winter’, masculine *himá-s* ‘cold, frost’ and neuter *hima-m* ‘snow’:

Lehmann re-translates *\*hímā* ‘winter’ as ‘season of cold’, pointing out the collective meaning of this feminine in contrast to *himás* ‘cold, frost’, which refers to a singular occurrence of frost. *Hímā* has the meaning [...] ‘the whole of the occurrences of frost/cold’. The so-called masculine was originally a count noun with “singulative” quality, the so-called feminine a collective noun. The two genders consequently had nothing to do with biological gender; they were much more closely related to the category number. [...] The so-called neuter *himam* ‘snow’ is, according to Lehmann, a nominal resultative; thus, *himam* could be paraphrased ‘the result of cold’. [Leiss, 1999: 241]

In this sense, Indo-European gender was originally a marker of nominal aspect or *Seinsart* (in the sense of Rijkhoff, 2002). Only later, when this derivational device lost productivity, did biological gender play a major role in the resemanticization of this category.<sup>5</sup>

### 4.3 Gender and noun classes

Three types of gender systems are found in the Indo-European language family: (i) a two-way opposition between masculine and feminine (as in most Romance languages, as well as modern Insular Celtic languages); (ii) a three-gender distinction (masculine vs. feminine vs. neuter), as in German and

<sup>5</sup>Reflexes of this hypothesized system may still be seen, albeit non-productively or less productively, in languages where specifically defined classes of concepts (e.g. the concept of individuality of reference) are represented grammatically by the marked gender value (Acquaviva, 2006).

Table 4.1: Indo-European gender systems

| No. of genders | Genders                | Examples |
|----------------|------------------------|----------|
| 2              | Masc. and fem.         | Irish    |
| 2              | Common and neuter      | Danish   |
| 3              | Masc., fem. and neuter | German   |

Russian; and (iii) a two-way distinction between common gender and neuter, as in Danish (Corbett, 1991: 102, 247). The situation can be represented as in Table 4.1. In Indo-European, type (i) historically results from type (ii) after the loss of the neuter gender.

Outside the Indo-European context, we find numerous systems of noun classification based on agreement, generally known as “noun class” systems. Structurally, gender and noun class systems are strikingly similar and it has long been observed that the distinction might be a purely terminological one, as both are defined in terms of agreement (Vendryes, 1921: 113; Fodor, 1959; Greenberg, 1978; Hjelmslev, 1970). Greenberg (1978: 49) observes that the term “gender” is generally preferred when biological gender is part of the semantic criteria of classification, but there is otherwise no substantial difference. Corbett, who generalizes the label “gender” to cover both, claims that different terminologies arise from different grammatical traditions rather than being based on linguistic differences: Tamil and Karata, for instance, resemble each other in many respects, including their noun class systems (Table 4.2), traditionally known as such in Karata but as genders in Tamil (Corbett, 1991: 146).<sup>6</sup>

<sup>6</sup>Aikhenvald & Dixon (1998: 59) observe that although the two terms are “often used interchangeably in the literature”, it may nonetheless be necessary to distinguish them, at least in those languages which show both gender and noun classes. Paumarí (a member of the Arawá language family, spoken in Southern Amazonia) is cited as an example:

Paumarí [...] has (i) a system of two genders, feminine and masculine, marked on third person singular pronouns, demonstratives, adjectives, and some verbal suffixes; and (ii) a system of two shape-based noun classes, marked by presence or absence of a verbal prefix. [Aikhenvald & Dixon, 1998: 59, note 9]

Both systems are clearly defined by the presence of agreement phenomena, although system (i) appears to involve a rather widespread range of targets, system (ii) only verbs. A similar situation is described in Corbett (1991: 184–188) for the Mba subgroup of the

Table 4.2: Gender systems in Tamil and Karata

| Tamil     | Karata | Referents                | Criterion       |
|-----------|--------|--------------------------|-----------------|
| Masculine | I      | gods, male humans        | male rational   |
| Feminine  | II     | goddesses, female humans | female rational |
| Neuter    | III    | other                    | non-rational    |

(Data from Corbett, 1991: 9)

#### 4.4 Assignment systems

While it seems uncontroversial that gender is defined by agreement, it is still not clear how speakers choose the correct agreement form every time one is required. Two hypotheses are found in the literature:

H1: speakers retrieve the information about a noun’s gender from its lexical entry;

H2: speakers determine the gender of a noun based on some known properties thereof, e.g. its shape or its meaning.

Under H1, the gender of a noun is recorded in the lexicon in addition to other information like its phonological representation, its inflectional class, its meaning etc. Under H2, on the other hand, the gender of a noun can be determined (“on line”, so to speak) based on precisely this latter kind of information, so it is not necessary to burden the speakers’ memory by separately and redundantly recording it; however, speakers must learn a set of rules (an algorithm) that map the recorded lexical information to the unrecorded gender. I will refer to H1 as the *lexical* hypothesis and to H2 as the *processual* hypothesis.

In what follows I will present the arguments adduced in support of both hypotheses and discuss the relevance of this question in the context of the present study. I will then review this question in §6.2.3 after the relevant findings have been presented.

The **processual hypothesis** is defended by Corbett (1991: 7), who argues that if there is a gender assignment system, all the speaker has to do is

---

Niger-Kordofanian language family.

learn a few assignment rules (and a limited number of exceptions), and he or she will be spared the mnemonic effort of storing and recalling the gender of several hundreds of thousands of individual nouns. Corbett adds that native speakers make very few mistakes when producing agreement forms, which would be surprisingly if they were relying only on their memory. He also observes that borrowings from foreign languages are systematically assigned a gender in the receiving language, as are invented nouns in psycholinguistic experiments on language acquisition; this would be impossible without an assignment system.

In an experiment conducted with French children (Karmiloff-Smith, 1979: 148–169), for instance, the respondents were presented with images of fictitious extra-terrestrial creatures characterized as females but referred to by the invented name *bicron*, whose phonological shape is associated in French with the masculine gender. Asked to describe pictures featuring such characters, children up to nine years of age tended to produce masculine agreement with articles and adjectives, even if they preferred feminine pronouns, as in example (3); this is taken as the demonstration that despite being also aware of the female trait, these children assigned the masculine gender to the invented noun because of its phonological shape.

(3) French (Karmiloff-Smith, 1979: 164f.)

Bon, y avait une fois un bicron vert et un bicron brun.  
 well there was one time a.M bicron green.M and a.M bicron brown.M  
 Elles étaient très amies [...] alors le bicron vert est  
 They.F were very much friends [...] then the.M bicron green.M is  
 sorti [...] et ensuite elle est allée  
 gone out [...] and then she is gone

‘Well, once there was a brown bicron and a green bicron. They were close friends [...] and the green bicron went out [...] and then she went.’

When the researcher questioned this child as to the referent of the feminine pronoun (*Elle, c’est qui?* ‘Who, she?’), the reply was as in (4).

(4) French (Karmiloff-Smith, 1979: 165)

Celle-là, le bicron vert  
 that.F-there the.M bicron green.M

‘That one(F), the(M) green(M) bicron’

Examples (3) and (4) above also show the possibility of contrasting agreement forms triggered by the same controller. In Corbett's (1991) terminology, the feminine agreement form, consistent with semantic properties of the referent, is an instance of *semantic agreement*, whereas the masculine form, consistent with the grammatical gender of the noun, represents *syntactic agreement*.<sup>7</sup>

The **lexical hypothesis** can be summarized by Dik's statement that

[g]ender classes typically have SOME semantic motivations, but are based for the greater part on rather arbitrary divisions within the full class of nominal predicates, such that in general the Gender has to be learnt for each noun. [Dik, 1997: 377, emphasis original]

Van Berkum (1996: ch. 2) argues that gender assignment to borrowings and invented nouns in similar experiments is irrelevant, as "first-time" gender assignment such as might apply to new words, invented or not, need not have anything in common with the way in which speakers access the gender of known words. He also points out that Corbett does not even explain why committing gender information to memory should lead speakers to producing more performance errors than having an assignment system.

In addition, more recent psycholinguistic research has provided experimental evidence that gender information might be stored in the lexicon in some form, at least for high-frequency nouns, a possibility contemplated by Butterworth (2004) and Stemberger & MacWhinney (2004): this could be considered a weaker version of the lexical hypothesis.

Formal regularities, where available, may play an important role in the language acquisition phase (cf. van Berkum, 1996), after which speakers access gender information directly from the lexicon. Associations between gender class and word endings (and possibly other kinds of phonological features), which are developed in the process of language acquisition, seem to

---

<sup>7</sup>A further distinction should be made on the semantic side between *lexical gender* and *referential gender* (Dahl, 1999). Semantic lexical gender is gender as codified in the lexicon, based on cultural stereotypes: it accounts for the masculine gender of such examples as Russian *vrač* 'doctor' and *sud'ja* 'judge' (the latter being expected, by purely formal criteria, to be feminine). On the other hand, when *vrač* or *sud'ja* take feminine agreement forms, Dahl (ibid., p. 110) speaks of referential gender, based on "concrete properties of referents". I return to this distinction in §7.3.

help learners acquire grammatical gender (Taft & Meunier, 1998: 41; Kempe & Brooks, 2005); however, native speakers are then able to access gender information even when they cannot access any phonological information about a noun, due for instance to specific forms of language impairment. This was shown experimentally, in single-word naming tasks,<sup>8</sup> for languages in which nouns tend to appear with an article most of the time (e.g. Romance languages); speakers of Hebrew (a language with a higher frequency of bare nominals), on the other hand, were shown not to possess the same gender-retrieving ability under the same impairment conditions (Friedmann & Biran, 2003).

Taft & Meunier (1998) further show that for native speakers of French the lexical representation of gender information might be the form of the associated article: given a set nouns which are semantically and formally associated with the feminine gender (e.g. place names ending in *-ie* like *Australie* ‘Australia’ and *Russie* ‘Russia’),<sup>9</sup> speakers take longer to name the gender of nouns with vocalic onset because the form of the definite article used before a vowel is gender-neutral (*l’*).<sup>10</sup>

This is consistent with the hypothesis that stored in the lexicon are multi-word as well as single-word units, reflecting experience and conceptual entrenchment (Erman, 2007). In the case of Insular Celtic, the lexicalization of multi-word units (and in particular article–noun sequences) may be facilitated by the phenomenon of initial mutations, which are central to agreement marking in the noun phrase (see chapter 5). As observed by Bechert,

initial variability of the stem [...] is a rather strange phenomenon because, psychologically speaking, the beginning of the word is its most salient feature and plays a prominent part in its identification. Consequently, initial variability of the stem means a reduced stability of the word, and in fact we observe in the Celtic languages the ousting of the single word by the word

---

<sup>8</sup>In single-word naming tasks, participants are asked to indicate the gender of a noun in isolation.

<sup>9</sup>Their study is based on response to visual stimuli and therefore considers orthographic rather than phonological regularities.

<sup>10</sup>The same was observed by Tucker et al. (1977: 59). The indefinite article, which would be gender-differentiated even before vowels, very rarely occurs with these nouns (Taft & Meunier, 1998: 31).

group as the elementary building block of the sentence. [Bechert, 1990: 116]

Whether one consider the gender of a noun to be computed “on demand” by an assignment system or stored in the lexicon once and for all, and then retrieved from it, it should be recognized that there are striking correlations between semantic and formal properties of nouns on the one hand and gender affiliation on the other. These systematic sets of correspondences can still be called, with Corbett, “gender assignment systems”, even if they turned out to be mere descriptions of historically-motivated regularities. Furthermore, Corbett (1991, 1994, 2003) shows that from a typological perspective these assignment systems form a continuum that goes from a maximum to a minimum of semantic transparency, and can be classified with respect to this continuum. The types identified by Corbett are presented in what follows.

#### 4.4.1 Semantic systems

This segment of the transparency continuum may be further subdivided. Corbett (1991: 8) makes a distinction between “strictly” and “predominantly” semantic systems, depending on the extent to which semantic criteria enable the speaker to univocally predict gender.

##### 4.4.1.1 Strict semantic systems

In this kind of systems there is a one-to-one relationship between the meaning of any given noun and its gender, i.e. it is always possible to predict the gender of a noun from its meaning and conversely to predict relevant semantic features of the referent from grammatical gender.

In Tamil (Table 4.3), for instance, formal features are totally irrelevant to gender assignment. There are very few exceptions, which can however be accounted for on semantic grounds (Corbett, 1991: 9): for instance, the nouns *cuuriyan* ‘moon’ and *cantiran* ‘sun’ are masculine, which is consistent with the fact that they are also the names of male deities.

##### 4.4.1.2 Predominantly semantic systems

As we have seen, the neuter gender in Tamil contains nouns which cannot be assigned to either the masculine or feminine gender, i.e. what Corbett (1991)

Table 4.3: Gender assignment in Tamil

| Criterion               | Gender    | Examples       | Gloss         |
|-------------------------|-----------|----------------|---------------|
| male human or god       | masculine | aaṅ<br>civaṅ   | man<br>Shiva  |
| female human or goddess | feminine  | peṅ<br>kaaḷi   | woman<br>Kali |
| other                   | neuter    | maram<br>viṭṭu | tree<br>house |

(Source: Corbett, 1991: p. 9, Table 2.1)

calls the *semantic residue*. The difference between strict and predominantly semantic systems is that the latter allow a significant number of “leaks” (i.e., more than a few, sporadic exceptions) from the residual gender into any of the semantically defined genders (ibid., p. 13).

Dyirbal (see Dixon, 1972: 306–11), for instance, has four genders:

- gender I is associated with male humans and non-human animates;
- gender II with female humans, water, fire, fighting;
- gender III with non-flesh food;
- gender IV covers the semantic residue.

But there are many exceptions: e.g. gender I contains the nouns for ‘moon’, ‘rainbow’, some types of spears; gender II those for most kinds of birds, some insects, sun and stars, etc. These exceptions can be explained (i) by mythological associations (e.g. birds are in gender II because they are believed to be the spirits of dead women); (ii) by conceptual associations (thus ‘fishing line’ and ‘fish spears’ are in gender I due to their association with fish); (iii) as a means of emphasizing some important property (e.g. harmful creatures and things are easily found in gender II, probably by association with fire and fighting). Such principles can interact with one another, so ‘moon’ is in gender I while ‘sun’ and ‘star’, by association with light and fire, are in gender II: the reason can be found in traditional myths and beliefs

Table 4.4: Gender assignment from Traditional to Young People’s Dyirbal

| Traditional Dyirbal |  | Young People’s Dyirbal |                       |
|---------------------|--|------------------------|-----------------------|
| Gender              | Criterion                                    | Gender                 | Criterion             |
| II                  | <i>female humans</i> , water, fire, fighting | II                     | <i>female humans</i>  |
| I                   | male humans, non-human <i>animates</i>       | I                      | <i>other animates</i> |
| III                 | non-flesh food                               | conflated with IV      | —                     |
| IV                  | <i>residue</i>                               | IV                     | <i>residue</i>        |

(After Corbett, 1991: 16, 18 and Schmidt, 1985: 154)

(the moon is the husband of the sun). Other exceptions, nevertheless, are very hard or impossible to explain (Dixon, 1972: 310).<sup>11</sup>

#### 4.4.2 Transition between assignment systems

The younger generations of Dyirbal speakers have developed a new variety of the language, which has been called Young People’s Dyirbal (Schmidt, 1985). This includes a simplified version of the traditional assignment system, “similar to the way *he, she, it* are used in English” (Aikhenvald, 2000: 390). In the new system, gender III (non-flesh food) is lost and its members re-assigned to gender IV (residue); gender II has lost its characteristic association with fire and dangerous things, now being reserved for females only; nouns formerly assigned by mythological and/or conceptual association are reassigned semantically (e.g. nouns for birds, previously feminine as seen above, are nouns in class I on account of their being animates; fishing lines and spears are no longer in gender I with ‘fish’ but in the residue, gender IV); all exceptions have been regularized, thus expanding the size of the semantic residue (Schmidt, 1985: 151–168; Corbett, 1991: 16ff.). This can be represented as in Table 4.4.

<sup>11</sup>This account of gender assignment in Dyirbal has been criticized by Plaster & Polinsky (2007), who argue that to posit an assignment system based on such culture-specific associations is inconsistent with what we know about child learners’ abilities. Plaster & Polinsky list a number of formal regularities in the system which appear to have gone completely unnoticed by Dixon and those who have tried to make sense of this apparently complex system (most remarkably Lakoff, 1987).

Table 4.5: Semantics of noun classes in Proto-Bantu

| Noun class | Semantics   |
|------------|---|
| 1/2        | human, person   |
| 3/4        | extended (long) (e.g. body, river)                          |
| 5/6        | fruits; non-extended (e.g. stone, spot, nose)               |
| 7/8        | utilitarian artefacts, despised objects and beings          |
| 9/10       | animal  |
| 14/6       | differentiated internal structure (e.g. bridge, bow, canoe) |

(Source: Aikhenvald, 2000: 24)

This way, as Corbett points out, a strict semantic system developed out of a less transparent one. Other examples are available: the transition from formal to semantic in the history of English is discussed in Curzan (2003), while the ongoing decline of the grammatical gender system in Dutch is the subject of Audring (2009) and De Vos & De Vogelaer (forthcoming), to which I shall return in §7.4. In both English and Dutch, the transformation is attributed at least in part to the erosion of the once-rich word-final morphology, which used to mark gender agreement.

The opposite process is also possible. Arguably, the Traditional Dyrirbal system was itself the outcome of a strict semantic system which had become increasingly opaque on account of the mythological and conceptual associations that caused the migration of a large number of nouns from their original gender to a new one (Corbett, 1991: 98).

Although no documentary evidence of this process is available in the case of Dyrirbal, other language families can provide some indirect evidence of how progressive opacification can transform a strict semantic system into a less transparent one. For instance, the reconstructed noun class system of Proto-Bantu (Denny & Creider, 1986) shows a consistent assignment system based on semantic criteria (the schematic representation in Table 4.5 is by Aikhenvald, 2000: 24).

While in Isangu class 1/2 is still reserved for humans, in other Bantu languages its membership has been somewhat extended: in Babungo it now includes novel items such as some small animals, body parts, household objects (Aikhenvald, 2000: 24); in Luvale a small number of non-human an-

imates are now part of class 1/2, while in the closely related Lunda language all animates now belong in this class (Corbett, 1991: 98). In Luvale, among the few non-human animates found in class 1/2 is the noun *muumbe* ‘jackal’, and Corbett suggests that because a jackal often appears in folk tales as a personified character (in which case the noun takes class 1/2 agreement forms), *muumbe* may have acted as a “Trojan horse”, starting the migration of an increasing number of nouns denoting animals into class 1/2.

### 4.4.3 Formal systems

If the process of opacification described in the previous section continues, eventually only a few nouns will be consistently assigned according to semantic principles. The remainder, or at least a subset of it, would be assigned according to a set of regularities of a formal nature (Corbett, 1991: 32). While diachronic evidence that formal assignment systems derive from semantic ones may not always be available, this hypothesis is corroborated by the observation that no strict formal assignment system has been identified so far. Some semantic principle always overlaps, to various extents, with the formal criteria (Corbett, 1991: 8, 63). The semantic principles at the core of noun categorization almost always include criteria like [ $\pm$ human] or [ $\pm$ animate] (Aikhenvald, 2000: 25, 275–283) and thus apply only to a restricted set of nouns.<sup>12</sup>

#### 4.4.3.1 Phonological systems

When assignment depends on just one form of the noun, we are dealing with phonological assignment (Corbett, 1991). Qafar (or Afar)—an East Cushitic language (ibid., pp. 51f.)—has two genders, labelled masculine and feminine because they include all male animates and all female animates, respectively. The position of the accent has distinctive function and is crucial

<sup>12</sup>Gender assignment within the semantic residue might also be partially semantic. Zubin & Köpke (1986) argue that this is the case in German, where inanimates seem to be consistently assigned to the neuter gender if they have a general meaning and to the masculine or feminine gender if they have a more specific meaning (so nt. *Säugetier* ‘mammal’ vs. m. *Hund* ‘dog’, *Bär* ‘bear’, f. *Katze* ‘cat’, *Antilope* ‘antelope’; or nt. *Möbel* ‘furniture’ vs. m. *Tisch* ‘table’, *Stuhl* ‘chair’, f. *Lampe* ‘lamp’; Zubin & Köpke, 1986: 147, 150). Similarly, Siemund (2008) shows that the mass/count distinction is responsible for the assignment of inanimates to non-neuter genders in some non-standard varieties of English.

Table 4.6: Accent and gender in Qafar

| Masculine |                | Feminine  |                  |
|-----------|----------------|-----------|------------------|
| bàxa      | ‘son’          | baxà      | ‘daughter’       |
| barisèyna | ‘male teacher’ | baryseinà | ‘female teacher’ |
| kùta      | ‘dog’          | kutà      | ‘bitch’          |
| gilàl     | ‘winter’       | catò      | ‘help’           |
| baànta    | ‘trumpet’      | karmà     | ‘autumn’         |

(Source: Corbett, 1991: 51f.)

for gender assignment. As can be seen from Table 4.6, male animates are masculine, female animates feminine; inanimates are feminine if they end in an accented vowel, masculine if they end in either a non-accented vowel or a consonant. In the exceptional case of conflicting criteria, the semantic criterion prevails: so for instance *abbà* ‘father’ is masculine and *gabbixeèra* ‘slender-waisted female’ is feminine.

Godie—a language of the Kru group, a sub-branch of Niger-Kordofanian (Corbett, 1991: 54–55)—has a two-way gender system (human vs. non-human) complicated by a further subdivision within the non-human gender, which is comprised of three sub-genders. Table 4.7 summarizes the criteria that determine gender assignment in Godie. Its genders are named after the form of the singular pronoun that would be used in anaphoric agreement ( $\text{ɔ}$  for humans;  $\text{ɛ}$ ,  $\text{a}$ , and  $\text{ʊ}$  for the three non-human sub-genders). A simple set of phonological correspondences permits to predict the gender of inanimates as follows: (i) nouns ending in a front vowel ( $/\text{i}, \text{ɪ}, \text{e}, \text{ɛ}/$ ) are in gender  $\text{ɛ}$ ; (ii) nouns ending in a central vowel ( $/\text{i}, \text{ʉ}, \text{ə}, \text{a}/$ ) are in gender  $\text{a}$ ; (iii) nouns ending in a back vowel ( $/\text{u}, \text{ʊ}, \text{o}, \text{ɔ}/$ ) are in gender  $\text{ʊ}$ .

#### 4.4.3.2 Morphological systems

If two or more forms of a noun’s paradigm are required in order to predict its gender, the assignment system may be called morphological. In Russian, for instance, there is a close correspondence between the inflectional paradigm of a noun (morphological information) and its gender (Corbett, 1991: 34–43). Russian has three genders: masculine, feminine, and neuter. Semantic assignment only covers nouns for male animates (masculine gender) and female

Table 4.7: Genders in Godie: animates and inanimates

| Gender | Criterion       | Example  | Gloss    |
|--------|-----------------|----------|----------|
| “ɔ̃”   | human           | nyɔ̃krɔ̃ | man      |
| “ɛ”    | large animal    | lɔ̃ɛ     | elephant |
| “a”    | small animal    | nú       | louse    |
| “ɔ̃”   | liquid          | nyú      | water    |
|        | non-solid mass  | gbaylɔ̃  | smoke    |
|        | natural element | lagɔ̃    | sky, God |

(Data from Corbett, 1991: 53f.)

Table 4.8: Morphological gender assignment in Russian

| Gender    | Criterion                       | Example   |
|-----------|---------------------------------|---|
| Masculine | declension I                    | <i>zakon</i> ‘law’  |
| Feminine  | declension II<br>declension III | <i>škola</i> ‘school’<br><i>kost’</i> ‘bone’                              |
| Neuter    | other paradigms                 | <i>vino</i> ‘wine’ (decl. IV), <i>put’</i><br>‘way’ (decl. III irregular) |

(Data from Corbett, 1991: 34–36)

animates (feminine). Russian has four declensions: declension-I nouns are masculine; declension-II and III nouns are feminine; other nouns are neuter. There are a few exceptions: *djadja* ‘uncle’ and *deduška* ‘grandfather’, for instance, belong in declension I but are masculine (semantic assignment). Table 4.8 represents morphological gender assignment in Russian.

#### 4.4.4 Overt vs. covert gender

A further distinction is made between *overt* and *covert* gender systems. According to Greenberg (1978: 53), “where a [gender] marker exists in the noun itself, the system will be called overt; where it is not found, the system will be called covert”. In fact, this distinction is not dichotomic and is better regarded as a typological continuum of possibilities (Corbett, 1991: 62f.; Aikhenvald, 2000: 57f.).

English is an example of covert gender system: agreement (which is pronominal only) entirely depends on semantic criteria. There is normally no gender marker on nouns, with the sporadic exception of feminine suffixes such as *-ess* and *-woman*. Qafar consistently allows gender assignment to be predicted from formal clues and can be taken as a particularly good example of overt gender system.

## 4.5 Agreement

The term “agreement” covers a rather wide range of phenomena, which can be characterized essentially as forms of “systematic covariation between a semantic or formal property of one element and a formal property of another” (Steele, 1978: 610; Corbett, 2006: 4).

As discussed in chapter 6, agreement is modelled in Functional Grammar in terms of feature-copying from a source to one or more targets (Dik, 1997: 377), i.e. from the constituent to which the feature is inherent to other constituents. Corbett (1991) employs the term “controller” for Dik’s “source”, a terminological choice which will be adopted in what follows.

### 4.5.1 Agreement targets in Indo-European

Gender agreement in Indo-European may involve a number of targets, from determiners to personal pronouns. The following is a review accompanied by some examples.

#### 4.5.1.1 Articles

e.g. German *der Mann* ‘the.M man’ vs. *die Frau* ‘the.F woman’ vs. *das Kind* ‘the.NT child’.

#### 4.5.1.2 Adjectives

e.g. Spanish *hombre hermos-o* (man beautiful-M) ‘handsome man’ versus *mujer hermos-a* (woman beautiful-F) ‘beautiful woman’.

#### 4.5.1.3 Demonstratives

e.g. Russian *tot žurnal* (that.M magazine) ‘that magazine’ vs. *ta kniga* (that.F book) ‘that book’ vs. *to pis'mo* (that.NT letter) ‘that letter’ (ex-

amples from Corbett, 1991: 106).

#### 4.5.1.4 Numerals

e.g. Zürich German *zw-ee Mane* (two-M men) ‘two men’ vs. *zw-oo Fraue* (two-F women) ‘two women’ vs. *zw-ai Chind* (two-NT children) ‘two children’ (examples from Hurford, 1988: 590).

#### 4.5.1.5 Possessives

possessives may agree in gender locally, i.e. with the head of the noun phrase they modify, or anaphorically with their antecedent (possessor); or with both. Italian is an example of the first possibility.

(5) Italian

|               |        |                    |             |
|---------------|--------|--------------------|-------------|
| a. su-o       | figlio | b. su-a            | figlia      |
| POSS:3SG-M    | son(M) | POSS:3SG-F         | daughter(F) |
| ‘his/her son’ |        | ‘his/her daughter’ |             |

English is an example of the second possibility (anaphoric agreement: cf. *his son* vs. *her son*), and so are Irish and Welsh (see chapter 5); German represents the third option (double agreement marking, local and anaphoric).

(6) German

|           |                   |           |
|-----------|-------------------|-----------|
| a. sein-∅ | Freund / ihr-∅    | Freund    |
| his-M     | friend(M) / her-M | friend(M) |
| b. sein-e | Macht / ihr-e     | Macht     |
| his-F     | power(F) / her-F  | power(F)  |

#### 4.5.1.6 Verbs

in analytic forms comprised of a finite form of the verb in conjunction with an adjectival form (e.g. a participle), the latter may inflect for gender.

(7) Italian

|          |            |          |
|----------|------------|----------|
| Maria    | è andat-a  | a casa   |
| Maria(F) | is gone-F  | to house |
| ‘Maria   | went home’ |          |

In some languages, e.g. Russian and Modern Hebrew, a participle may be used on its own, without the finite verb (Aikhenvald & Dixon, 1998: 70f.; Corbett, 1991: 125f.), so it agrees with its subject in gender and number but not in person, as the Russian past tense in example (8).

(8) Russian (Corbett, 1991: 100)

- a. žurnal        ležal-∅ na stole  
    magazine(M) lay-M    on table  
    ‘the magazine lay on the table’
- b. kniga        ležal-a na stole  
    book(F) lay-F    on table  
    ‘the book lay on the table’
- c. pis'mo        ležal-o na stole  
    letter(NT) lay-NT on table  
    ‘the letter lay on the table’

#### 4.5.1.7 Relative pronouns

e.g. in German.

(9) German

- a. der    Mann,    der    nach Berlin ging  
    the.M man(M) who.M to    Berlin went  
    ‘the man who went to Berlin’
- b. die    Frau,        die    nach Berlin ging  
    the.F woman(F) who.F to    Berlin went  
    ‘the woman who went to Berlin’
- c. das    Kind,        das    nach Berlin ging  
    the.NT child(NT) who.NT to    Berlin went  
    ‘the child who went to Berlin’

#### 4.5.1.8 Personal pronouns

it is normal for Indo-European languages to have gender distinguished, to various extents, in the personal pronoun series. In some languages, like German or English, separate gender forms are only available in the third person

singular (*er* ‘he’ vs. *sie* ‘she’ vs. *es* ‘it’); in Spanish, gender distinctions are found in all plural persons as well as in the third singular (see Table 4.9).

The question whether there actually is agreement in gender between personal pronouns and their antecedents is a controversial issue on which Corbett (1991: 112) maintains that the selection of the appropriate gender form of the anaphoric pronoun cannot be considered different from “undisputed cases of agreement”. Therefore, he regards English as having three genders (masculine, feminine, and neuter), an opinion shared by Comrie (1999: 458f.). Others (e.g. Fodor, 1959: 2; Aikhenvald & Dixon, 1998: 59; Aikhenvald, 2000: 21) do not regard the choice of the anaphoric pronoun’s form as a case of agreement. Siemund (2008: 7–9) distinguishes pronominal gender from gender proper, and argues that languages like English have the former but not the latter. However, the existence of an Agreement Hierarchy (see §4.5.3 below and chapter 7), with the possibility of similar agreement choices with different targets, including pronominal ones, suggests a unified treatment of agreement.

## 4.5.2 Constraints on agreement

Various factors may limit the realization of gender agreement.

### 4.5.2.1 Syntactic constraints

In German, attributive adjectives agree with the head of the noun phrase, but predicative adjectives do not (Corbett, 1991: 124).

(10) German (Corbett, 1991: 124)

Table 4.9: Spanish personal pronouns

| Person | Gender    |           |
|--------|-----------|-----------|
|        | M         | F         |
| 3SG    | él        | ella      |
| 1PL    | nosotr-os | nosotr-as |
| 2PL    | vosotr-os | vosotr-as |
| 3PL    | ell-os    | ell-as    |

- a. warm-er Tee / warm-e Milch / warm-es Wasser  
 warm-M tea / warm-F milk / warm-NT water  
 ‘warm tea / milk / water’
- b. der Tee / die Milch / das Wasser ist warm  
 the.M tea / the.F milk / the.NT water is warm  
 ‘the tea / the milk / the water is warm.’

#### 4.5.2.2 Phonotactic constraints

In Italian, definite and indefinite articles are marked for gender, as in (11), but this distinction is commonly neutralized when the noun begins with a vowel, as shown in (12).

(11) Italian

- |   |  |
|---|--|
| a. un/il gatto<br>a.M/the.M cat(M)<br>‘a cat/the cat’ | b. una/la gatta<br>a.F/the.F cat(F)<br>‘a she-cat/the she-cat’ |
|---|--|

(12) Italian

- |   |  |
|---|--|
| a. un/l' amico<br>a/the friend(M)<br>‘a friend/the friend (male)’ | b. un'/l' amica<br>a/the friend(F)<br>‘a friend/the friend (female)’ |
|---|--|

#### 4.5.2.3 Morphological constraints

The inflectional paradigm of the target may block agreement. In Spanish, for instance, adjectives that inflect like *rojo* ‘red’ must agree in gender with the controller, while those that inflect like *verde* ‘green’ cannot.

(13) Spanish

- a. el coche roj-o / verde  
 the.M car(M) red-M / green  
 ‘the red car / the green car’
- b. la bici roj-a / verde  
 the.F bike(F) red-F / green  
 ‘the red bike / the green bike’

#### 4.5.2.4 Lexical constraints

It is typically only the lower numerals that mark gender agreement (Corbett, 1991: 135; Hurford, 1988: 590).

#### 4.5.2.5 Syncretism

Failure to realize gender agreement as a result of interactions with other grammatical categories falls within the scope of what has traditionally been called syncretism or neutralization (Aikhenvald & Dixon, 1998: 61; Baerman et al., 2005: 4–7, 81–91; Moravcsik, 2003).

Syncretism of gender with number is particularly common.<sup>13</sup> Other categories which may be syncretic with gender are tense, person, and case (Aikhenvald & Dixon, 1998: 66–71; Corbett, 1991: 125–33). Gender–person neutralization in Indo-European may involve personal pronouns without affecting the targets agreeing with them (Corbett, 1991: 128): cf. French *je suis heureux* (/œrø/) ‘I am happy.M’ vs. *je suis heureu-se* (/œrø-z/) ‘I am happy-F’, where the form of the adjective agrees with a gender feature that the pronoun (*je*) is underspecified for (Corbett, 1994: 58). Interactions of gender with case are also quite common in Indo-European: in Latin, for instance, the genitive and dative case of demonstrative pronouns and adjectives neutralize gender distinction in both the singular and plural.

#### 4.5.3 Hybrid nouns

Nouns with which different types of targets display different agreement forms are called by Corbett (1991: 183f., 228) “hybrid” nouns. For example, German *Mädchen* ‘girl’ (neuter) consistently takes neuter agreement forms within the noun phrase, but may be referred to by feminine pronouns. In Russian, hybrid nouns like *vrač* ‘doctor’ tend to take masculine agreement forms within the noun phrase, but other targets (predicates, relative pronouns and personal pronouns) may display feminine agreement forms con-

<sup>13</sup>“Where gender marking is morphologically distinct between singular and non-singular, gender syncretism is more the rule than the exception” (Baerman et al., 2005: 82).

sistent with the referent's sex (*ibid.*, p. 183).

(14) Russian (Corbett, 1991: 180)

Naš- $\emptyset$  vrač prišl-a.  
our-M doctor came-F

'Our doctor came.'

From a typological point of view, the probability of multiple agreement forms with different targets can be described as in (15).

(15) The Agreement Hierarchy (Corbett, 1991: 226)

attributive < predicate < relative pronoun < personal pronoun

The Agreement Hierarchy predicts that if multiple agreement forms are allowed, semantic agreement will be more likely for targets on the right-hand side of the hierarchy. It also implies that there will only be one cut-off point at which the split may occur: we would not for instance expect to find a language in which semantic agreement is possible with predicates only, while attributive adjectives and pronouns must agree syntactically.

## 4.6 Summary

In this chapter, I introduced the grammatical category of gender, explaining how it is defined and illustrating its typical structural properties. In doing so, I paid particular attention to the structural and semantic properties of gender systems in the Indo-European language family, in which Insular Celtic belongs.

The crucial notions of agreement, assignment system and conflicting assignment criteria were also presented and discussed.

I also introduced and discussed the two debated questions whether grammatical gender information is recorded and stored in the lexicon or assigned by rules, and whether it has a function. I will return to these questions in chapters 6 and 7, respectively.

## Chapter 5

# Gender in Celtic

### 5.1 Overview

In this chapter I will introduce the linguistic domain of this study (§5.2) and discuss the important class of phenomena known as initial mutations, which play a prominent role in the realization of gender agreement in Insular Celtic. Initial mutations will first be considered from the historical point of view (§5.3.1); then the theoretical debate about these phenomena will be summarized (§5.3.2) and an FDG-based treatment proposed (§5.3.3). Finally, in §§5.4–5.9, I will describe the grammatical gender systems of the six modern representatives of Insular Celtic (Irish, Scottish Gaelic, Manx, Welsh, Breton and Cornish).

### 5.2 The Celtic language family

The six languages presented in this chapter continue the Insular Celtic branch of Proto-Celtic (Figure 5.1), itself a (reconstructed) member of the Indo-European language family. The other branch of Proto-Celtic, known as the Continental subgroup, is not represented by any modern language (Fife, 1993: 6; Watkins, 1999). Insular Celtic comprises two subgroups, Goidelic (or Gaelic) and Brythonic (or Brittonic or British). To the former belong the modern continuators of Old Irish, i.e. Irish, Scottish Gaelic and Manx. Welsh, Breton and Cornish all stem from Brythonic.

All of these languages have a two-way gender system which includes masculine and feminine. The neuter was still attested in Old Irish but had

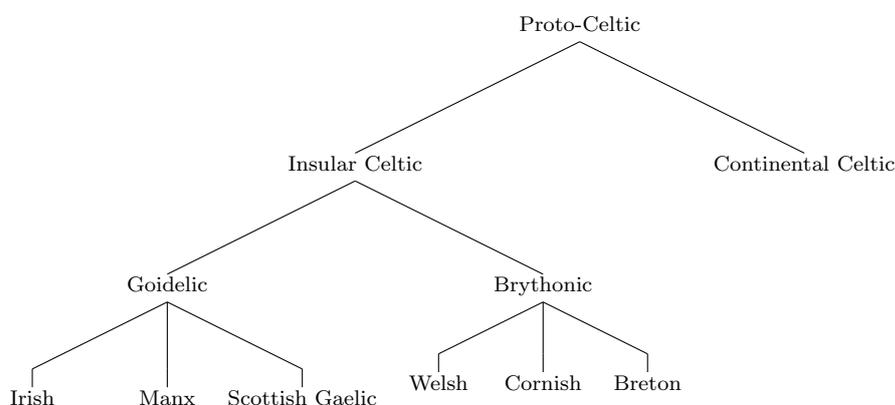


Figure 5.1: A genealogical table of the Celtic family  
(from Fife, 1993: 6, adapted)

completely disappeared by the tenth century (Thurneysen, 1980: §245). In the languages of the Brythonic branch the neuter was lost before the earliest available documents (Lewis & Pedersen, 1974: 159). The reduction to schwa that interested the vowel of all unstressed proclitics in the ninth century is commonly held responsible for the loss of the neuter gender in Old Irish: as the neuter article became less distinguishable from the masc./fem. form, the latter began to be used in lieu of the former (Thurneysen, 1980: §245); most Old Irish neuters became masculines.<sup>1</sup> In Brythonic, old neuters could become either masculine or feminine; some became masculine in one language and feminine in another, e.g. *tro* ‘turn’ (masculine in Welsh, feminine in Breton); others would take different agreement forms in different contexts, like *tra* ‘thing’, which in modern Breton takes feminine agreement forms with articles and attributive adjectives, masculine agreement forms with numerals (Lewis & Pedersen, 1974: 159).<sup>2</sup>

<sup>1</sup>Before the neuter had ceased to exist, a number of nouns were already oscillating between neuter and masculine assignment (e.g., *recht* ‘law’, *áis/óis* ‘age’) or between neuter and feminine (e.g., *ré* ‘time’ Lewis & Pedersen, 1974: 159).

<sup>2</sup>Less systematic reflexes are observed in fossilized exceptions to synchronic rules: for instance, in Welsh the masculine numeral *dau* ‘two’ lenites the following word, but lenition may fail to apply with certain old neuters: e.g., *dau cant* (unlenited) vs. *dau gant* (lenited) ‘200’, *dau tu* (unlenited) vs. *dau du* (lenited) ‘two sides’ (Lewis & Pedersen, 1974: 159).

## 5.3 Initial mutations and gender agreement

### 5.3.1 Historical aspects

Initial mutations are a particularly salient feature of Insular Celtic (Hamp, 1951: 230). Historically, they represent the fossilized continuation of previously productive sandhi phenomena, like intervocalic lenition. An extensive account of initial mutations in Celtic, from a typological point of view, may be found in Hickey (1995).

Intervocalic lenition is a typologically well-represented type of sound change whereby the articulation of a consonant between two vowels is weakened (Giannelli & Cravens, 1997: 35; Kirchner, 2004: 313). In the variety of Italian spoken in Florence, for instance, the word *prato* ‘lawn’ may be pronounced as either [ˈpraːto] or [ˈpraːθo], without any change in meaning (Giannelli, 1997; Kirchner, 2004: 316–20; Loporcaro, 1997). The lenition of /t/ (> [θ]) is triggered by its intervocalic position and does not encode any grammatical distinction. Intervocalic lenition in Florentine may also occur across word boundaries. The sequence *la casa* ‘the house’ may be pronounced as either [laˈkaːsa] or [laˈxaːsa]. However, it may be blocked in certain circumstances: the sequence *a casa* ‘at home’, for instance, is realized as [aˈkaːsa], without lenition and with a stronger articulation of /k/ (*Raddoppiamento Fonosintattico*), because *a casa* continues Latin *ad casa(m)* /ad kasa(m)/ (/dk/ > /kk/; cf. Loporcaro, 1997: 42; Rohlfs, 1966: 235–238). No grammatical distinction is encoded by *Raddoppiamento Fonosintattico*, either.

In Insular Celtic, on the other hand, the allophonic variation resulting from erstwhile sandhi phenomena was phonemicized and put to grammatical use after the triggering environment had disappeared due to the systematic loss of unstressed word-final syllables (Hamp, 1951: 239–42; Kortlandt, 1982). For instance, the initial /b/ of the reconstructed adjectival stem *\*bodi-* in Proto-Goidelic was lenited to [v] after a vowel, as shown in example (1).

(1) Proto-Goidelic (from Hamp, 1951: 241)

- a. \* wir-as                      bodi-as  
     man(M)-NOM.SG blond-NOM.M.SG  
     ‘a blond man’

- b. \*wir-u:                      vodi-u:  
       man(M)-DAT.SG blond-DAT.M.SG  
       ‘to a blond man’

In Old Irish, *fer buide* continued \**wiras bodias* and *fur buidiu* (/v/*uidiu*)<sup>3</sup> continued \**wiru vodiu*. The noun now ended in a consonant in both cases, so the initial mutation of the adjective was no longer triggered by the phonological context; moreover, /v/ (originally an allophone of /b/) had acquired phonemic status (Hamp, 1951: 241).

### 5.3.2 Initial mutations in synchrony: theoretical issues

Once phonemicized, initial mutations became available as a means of marking grammatical distinctions, including agreement. However, patterns of initial mutation have persisted in Insular Celtic even when they serve no apparent purpose. In example (2a), lenition has a distinctive function, being used to mark the gender of the possessor; its presence or absence causes different interpretations; in (2b), on the other hand, it has no distinctive function: its absence would be unexpected, but not meaningful.

(2) Irish *béal* /b'e:l/ ‘mouth’

- a. *a bhéal* /ə ve:l/ ‘his mouth’ : *a béal* /ə b'e:l/ ‘her mouth’  
 b. *mo bhéal* /mə ve:l/ ‘my mouth’ : ?*mo béal* /mə b'e:l/ ‘my mouth’

In example (3), lenition has a distinctive function not at the semantic but at the morphosyntactic level: in the nominative singular, it affects feminine nouns (but not masculine ones) after the definite article.

(3) Irish

- a. masc. sg. *béal* /b'e:l/ ‘mouth’ : *an béal* /ən b'e:l/ ‘the mouth’  
 b. fem. sg. *beatha* /b'ahə/ ‘life’ : *an bheatha* /ən vahə/ ‘(the) life’

This distinctive use continues erstwhile sandhi lenition, since there were historically two synchronic forms of the article in the nominative singular: a consonant-final masculine form and a vowel-final feminine form, the latter causing the the following consonant to be in intervocalic position.

<sup>3</sup>Old Irish orthography does not always note lenition.

Lenition is also used to mark grammatical gender with adjectives, as in (4), but in this context it does not always continue erstwhile sandhi lenition, since it is found even after feminine nouns which originally terminated in a consonant. Its application was generalized later, after it had been grammaticalized as an agreement marker (McCone, 1994: 120f.).

## (4) Irish

bean                      mhaith  
 woman(F).NOM.SG    good.NOM.F.SG

‘a good woman’

(cf. *fear* maith ‘a good man’)

Equally disjoined from their historical causes are examples where the initial mutation is distinctive and occurs independently of any preceding word, as in example (5).

## (5) Irish

a. /bʳʰisʰ              e:/    ⟨bris é⟩  
     break-IMP.2SG 3SG.M

‘break it’

b. /vrʰisʰ              sʰe:/    ⟨bhris sé⟩  
     break-PST 3SG.M

‘he broke’

Although the case of adjective and past-tense lenition should leave no doubt as to the fact that initial mutations may be employed, from a synchronic point of view, as a morphosyntactic device, the attempts to provide a unified and purely phonological explanation to this phenomenon have been numerous, based either on abstract (non-surfacing) phonemes or on diacritics and readjustment rules. Kallen (1979: 19–27) surveys a number of them, pointing out their ad-hoc nature, the fact that they lead to a proliferation of underlying abstract forms, and the impossibility of subjecting any of these hypotheses to empirical testing. Furthermore, they fail to provide any type of generalization or further insight “beyond that found in a traditional list of mutating environments” (p. 26). This, Kallen observes, is ultimately a consequence of the “lack of any predictable surface phonological environment

for initial mutation” (p. 19). Further phonological models are examined and criticized by Stewart (2004: ch. 3), and will not be reviewed here; it should be further noted that the arguments Green (2006) presents against the morphophonological approach (see below) apply by extension to any purely phonological model.<sup>4</sup>

An attempt to overcome the limitations of a purely phonological analysis by overtly involving morphology is represented by the morphophonological approach, first proposed by Hamp (1951), who posits distinctive elements (*morphophonemes*) whose presence triggers the initial mutation of the following word. These elements are quasi-segmental in that they do not have autonomous phonological realization; rather, their presence is revealed only by the modification of the following sound. In this sense their operation pertains to phonology, but their distribution does not, since it may be either lexical or morphological. Thus the two synchronically distinct forms of the article may be represented as *an* (masculine singular) and *an<sup>L</sup>* (feminine singular), as in (6).

|           |                   |                           |                     |
|-----------|-------------------|---------------------------|---------------------|
| (6) Irish |                   |                           |                     |
|           | NP                | Underlying representation | Surface realization |
|           | <i>an béal</i>    | /əN b'eɪl/                | /əN b'eɪL/          |
|           | <i>an bheatha</i> | /əN <sup>L</sup> b'ahə/   | /əN vahə/           |

Following Hamp (1951) and Oftedal (1962), superscript *L* represents the morphophoneme which triggers lenition. It should be noted, as it will be relevant for the discussion to follow, that this type of analysis allows to model article-induced mutations—which would otherwise be considered a clear case of head marking since it is the head of the noun phrase which is marked—as dependent marking, by positing two different underlying forms of the article at a synchronic level.

As Kallen (1979: 28) further points out, the morphophonological approach does not seek to offer a unified account, which had been the aspiration of phonological models; rather, a distinction between different types of initial mutations is explicitly articulated by Oftedal (1962). Simplify-

<sup>4</sup>A rather different “phonological” approach, notes Stewart (2004: 37), has involved expanding the domain of phonological explanation by allowing into it conditioning factors that are in fact syntactic, morphological or lexical. This, as already pointed out by Janda (1987, cited *ibid.*), points to the “absurd logical conclusion of permitting anything that touches sound structure to be categorized as phonology”.

ing somewhat, Oftedal distinguishes between *projected* mutations, i.e. those triggered by a morphophoneme of the preceding word, as in examples (2–4), and *incorporated* mutations, which depend on the morphosyntactic context but not necessarily on any preceding morpheme, as in example (5b).

As both Kallen (1979: 29) and Stewart (2004: 39) further observe, morphophonemes are positioned *a posteriori*, based on the observed distribution of mutated segments, rather than *a priori*, based on some independent motivation for their presence. A similar analysis is proposed by Lieber (1983), who calls the trigger a “floating autosegment” which attaches to, or is located at, the end of a word and instructs the phonological component to change the initial segment of the following word.

Green (2006: 1948) points out two major problems with morphophonological approaches: (i) there is no single phonological feature or bundle of features in whose terms “the wide variety of alternations found within a single mutation” may be described; (ii) the phonological component does not treat all morphophonemes equally; sometimes it just ignores them, or skips them.

Irish lenition provides good examples of both problems. As will be discussed in §5.4.1.1, lenition “turns oral stops and *m* (but not *n*) into fricatives, debuccalizes coronal obstruents, ‘laxes’ tense coronal sonorants, and deletes *f*” (Green, 2006: 1959):

[+continuant] alone will trigger only the spirantization, not the debuccalization, sonorant laxing, or *f*-deletion. If a case can be made that what distinguishes *L L' N N'* from *l l' n n'* is the feature [tense] [...] then [–tense] could account for the sonorant laxing, but not the other cases. And it is difficult to conceive of any feature that could be added to *f* to induce deletion. [Green, 2006: 1959]

In other words,

[u]nlike truly phonological processes [...], the mutations do not target natural classes of sounds or have uniform effects. [ibid.]

The second problem is exemplified in (7).

## (7) Irish (Green, 2006: 1966)

- a. *cuid* ‘part’
- b. *ár<sup>N</sup> gcuid* ‘our part’
- c. *dhá<sup>L</sup> chuid* ‘two parts’
- d. *ár<sup>N</sup> dhá<sup>L</sup> gcuid* ‘our two parts’

As this example shows, *ár<sup>N</sup>* and *dhá<sup>L</sup>* trigger nasal mutation and lenition, respectively, as expected; but when both are present, as in (7d), the noun is nasalized rather than lenited, despite being immediately adjacent to *dhá<sup>L</sup>*. While there might be possible explanations for (7d), e.g. that the semantic expression of the possessor overrides adjacency, these cannot be purely phonological.<sup>5</sup>

Borsley et al. (2007) note another flaw of an autosegment analysis à la Lieber (1983), i.e. that in a sequence of adjectives each mutation would have to be caused by a separate occurrence of the mutating autosegment, which originates from the feminine noun and “reattaches itself to the following word” (Borsley et al., 2007: 188), and that therefore “the mutation on all but the first adjective is actually triggered by the preceding adjective” (p. 189), whereas the mutation of the first adjective is triggered by the preceding noun, as in the following examples (SM = soft mutation; see §5.7.1.1 below).

## (8) Welsh (Borsley et al., 2007: 189)

gadd<sup>SM</sup>      fawr<sup>SM</sup>      breifat<sup>SM</sup>      (cf. *mawr*, *preifat*)  
 garden(F).SG large.F.SG private.F.SG

‘a large private garden’

## (9) Irish (Ó Sé, 2000: 65)

an<sup>L</sup>                      fhlead<sup>L</sup>                      bheag<sup>L</sup>                      dheas<sup>L</sup>  
 ART.NOM.F.SG whistle(F).NOM.SG little.NOM.F.SG nice.NOM.F.SG  
 chaol<sup>L</sup>  
 slender.NOM.F.SG

‘the nice, small, slender whistle’

<sup>5</sup>In spoken Irish, furthermore, a certain amount of variation may be found: a small survey of Gaeltacht speakers revealed that *ár<sup>N</sup> ndá cuid* (with no “trigger skipping” and no mutation after the numeral) and *ár dhá<sup>L</sup> chuid* (with lexicalized mutation on the numeral and the regular mutation after the numeral) are also accepted forms (Denis Kelliher, personal communication).

Borsley et al. (ibid.) find that having different sources of SM on the different adjectives is per se an “undesirable feature” of this analysis, which is also problematic when it comes to accounting for the occurrence of certain mutation patterns. Consider example (10).

(10) Welsh (Borsley et al., 2007: 191)

eich siop<sup>SM</sup> [lyfr-au<sup>SM</sup> Cymraeg] leol  
 POSS:2PL shop(F).SG book-PL Welsh local.F.SG

‘your local Welsh-language book shop’

If the mutating autosegment reattaches itself to the right edge of each mutated word, then it is not clear why mutated *lyfrau* (< *llyfrau*) fails to mutate the adjacent adjective *Cymraeg*, which in turn, albeit unmutated itself, causes the following adjective *lleol* to mutate (> *leol*). The analysis would lead us to expect \**siop<sup>SM</sup> lyfrau<sup>SM</sup> Gymraeg<sup>SM</sup> leol<sup>SM</sup>*, but what we actually see is *siop<sup>SM</sup> lyfrau Cymraeg<sup>SM</sup> leol<sup>SM</sup>* (Borsley et al., 2007: 192).

A similar problem arises in Irish. As shown in example (11), attributive nouns are not mutated when modified by an attribute, which we cannot explain on purely phonological grounds.

(11) Irish (Ó Siadhail, 1989: 121)

oíche<sup>L</sup> [gaoithe móire]  
 night(F).NOM.SG [wind.GEN.SG great.GEN.SG]

‘a night with a high wind’

Nor can we explain in phonological terms why, in Welsh, mutation is triggered on attributive noun phrases (cf. *siop lyfrau* above) but not on those which denote a possessor, as in example (12) (cf. §5.7.1.6 below);

(12) Welsh (Borsley et al., 2007: 192)

siop<sup>SM</sup> [mab y meddyg]  
 shop(F).SG son(M).SG ART doctor

‘the doctor’s son’s shop’

or why the trigger fails to operate beyond the noun phrase boundary, as shown in examples (13a) and (13b).

## (13) Irish

- a. tá [an<sup>L</sup> b<sup>h</sup>ean<sup>L</sup> c<sup>h</sup>liste] ann  
 be.PRS ART.NOM.F.SG woman(F).SG clever.NOM.F.SG there  
 ‘the clever woman is there’
- b. tá [an<sup>L</sup> b<sup>h</sup>ean<sup>L</sup>] c<sup>l</sup>iste  
 be.PRS ART.NOM.F.SG woman(F).SG clever  
 ‘the woman is clever’

In other words, the observed mutation patterns need to be explained in syntactic and possibly semantic terms: the constituent [*llyfrau Cymraeg*] in (10) mutates as a whole, and the attributive noun in (11) does not mutate if it is followed by an adjective; and a dependent noun phrase will mutate if it has an attributive meaning (10), but not if it denotes a possessor (12); and so on. Phonology alone cannot account for all the data (Borsley et al., 2007: 192). However, the superscript notation is a convenient shorthand to describe mutation patterns, and as such it will be employed in this chapter; where it is, it should not to be taken as representing morphophonemes.

Attempts to explain initial mutations not as purely phonological, but rather as the phonological reflex of syntactic processes, have also been made; such attempts share the assumption that there is a “direct interface between syntax and phonology” (Mittendorf & Sadler, 2006: 345), i.e. no room for morphological processes proper;<sup>6</sup> Stewart (2004: ch. 4) contains a comprehensive survey of some such attempts, and the problems they raise, so they will not reviewed here. As he remarks,

to attempt to reduce the mutations to fully determined, and therefore fully redundant, syntactic “effects” in phonology [...] glosses over the nature of the morpholexical properties that the mutation regularly realize. [Stewart, 2004: 69]

Initial mutations, continues Stewart, may be determined by syntactic categories like case, but also by lexical categories like gender and discourse properties like definiteness (ibid.; cf. §§5.4.1.6 and 5.5.1.5 below). Syntax

<sup>6</sup>As Kallen observed in 1979, most of the work in generative grammar had until then concentrated on syntax or phonology, and morphology occupied “a rather uncertain position” in this field (Kallen, 1979: 1). As Stewart (2004: ch. 4) shows, however, similar attempts continued to be made for some time after the publication of Aronoff (1994).

may condition the morphological realization of single constituents and determine whether specific lexical categories can act as agreement targets (cf., e.g., the case of German adjectives discussed in §4.5.2); but, continues Stewart (2004: 69f.), “[t]o attribute to syntax the origin of everything that must pass through it” would be as mistaken as to try to attribute to phonology everything that pertains to sound structure, as already observed.

The morpholexical nature of initial mutations has been demonstrated by a number of authors who share the view that they should be regarded as an inflectional phenomenon; among them, the aforementioned Kallen (1979), Stewart (2004), and Green (2006), as well as Mittendorf & Sadler (2006). Their proposals have important consequences for the internal organization of lexical entries.

Kallen observes that “the aspirated [= lenited] and unaspirated [= unlenited] forms of words in Irish are clearly related to each other” in the same way as two word forms like *woman* and *women* are in English, and “must be united by some kind of rule that creates a phonological change in a specified use of a lexical entry” (1979: 39); he proposes that the information necessary to generate these different forms is found in the lexicon (*ibid.*). Words which belong to the relevant categories (e.g. nouns) and begin with an appropriate consonant (i.e. one that can be mutated) will be marked in their lexical entry with a special diacritic, e.g.  $[\pm L]$  for lenition;<sup>7</sup> only the radical form is listed in the lexical entry (p. 40). Given the appropriate conditions, which may be syntactic or morphological, this diacritic will be realized as  $[+L]$  in the underlying phonological representation and thus trigger the corresponding initial mutation (p. 41); a “special kind of readjustment rule” (p. 59) produces the expected phonological representation (e.g., if the radical form of X has initial /p/ and X is marked as  $[+L]$ , then /p/ → /f/, etc.; p. 61). Exceptions are also specified in the relevant readjustment rule rather than in the lexical entry (e.g., if X is  $[+L]$  and X is one of a set of exceptional lexical items, then  $[+L] \rightarrow [-L]$ : cf. pp. 59f.).

In a similar vein, Green (2006) suggests that mutated forms (“mutation allomorphs”) of a word should be considered in the same way as its inflected forms are, and proposes that “the generalizations regarding the effects of the mutation are stored [in the lexical entry] as morphological word formation

<sup>7</sup>Kallen (1979) actually represents this diacritic as  $[\pm asp]$  since he uses the term “aspiration” instead of “lenition”.

strategies” (Green, 2006: 1975).

Stewart also regards initial mutations as the expression of morphological processes, more precisely as a case of “non-concatenative inflectional marking [...] realizing morphosyntactic or morpholexical properties” (2004: 3). The hypotheses he formulates specifically refer to Scottish Gaelic (see §5.5 below), which presents strong analogies with Irish, but abstracting from the details we can extend them to cover initial mutations in Insular Celtic in general. Stewart proposes that inflected forms of noun, verb and adjective lexemes in Scottish Gaelic “have three of four logically possible shapes available for selection by morphological rules” (Stewart, 2004: 106). Because these stems are morphemes (in the sense of Aronoff, 1994), i.e. pure forms rather than meaningful units, they are given abstract labels, e.g. “P-stem” rather than “plural stem”; also because “one or more of these stems may be segmentally identical to the other stems”, as shown below, abstract labels like “L-stem” are preferred to concrete ones like “lenited stem” (Stewart, 2004: 109). So the four labels “R-stem”, “L-stem”, “P-stem” and “X-stem” are adopted, and exemplified using the SG lexemes *cat* ‘cat’, *làmh* ‘hand’ and *sgoil* ‘school’ in (14).

(14) Scottish Gaelic stems (Stewart, 2004: 109)

|               |              |              |              |
|---------------|--------------|--------------|--------------|
| <i>R-stem</i> | <i>cat</i>   | <i>làmh</i>  | <i>sgoil</i> |
| <i>L-stem</i> | <i>chat</i>  | <i>làmh</i>  | <i>sgoil</i> |
| <i>P-stem</i> | <i>cait</i>  | <i>làimh</i> | <i>sgoil</i> |
| <i>X-stem</i> | <i>chait</i> | <i>làimh</i> | <i>sgoil</i> |

Those listed in (14) are not the nouns’ paradigms, but the stems on which the latter would be based, as shown in (15) for *cat*.<sup>8</sup>

<sup>8</sup>Stewart makes a further distinction between *stem mutations* and *shape mutations*. The former have to do with how a paradigm cell is realized, while the latter with morphosyntactic processes depending on a trigger, so the distinction is one between incorporated and projected mutations, respectively. To illustrate the difference, he takes the form of the genitive plural in ScG. In the absence of the article, the genitive plural form is always lenited, i.e. based on the L-stem (*chat* /xat<sup>h</sup>/). However, if the noun is preceded by the article, its genitive plural form is not lenited but nasalized (*nan cat* /nəŋ gat<sup>h</sup>/). There is no “N-stem” in this model, because nasalization does not result from the interaction of case and number, so it does not pertain to nominal inflection. Furthermore, nasalization is always a projected mutation in ScG, and its shape formation is “sufficiently regular” that Stewart does not see it as necessary to have it listed in the lexicon (2004: 140, endnote 6).

(15) Scottish Gaelic *cat* ‘cat’

| Case | Singular            | Plural                    |
|------|---------------------|---------------------------|
| Nom. | R-stem <i>cat</i>   | P-stem <i>cait</i>        |
| Gen. | P-stem <i>cait</i>  | L-stem <i>chat</i>        |
| Dat. | R-stem <i>cat</i>   | P-stem <i>cait</i>        |
| Voc. | X-stem <i>chait</i> | L-stem + /ə/ <i>chata</i> |

Mittendorf & Sadler (2006) also favour the hypothesis that “initial mutation is close to inflection in nature and is essentially a morphosyntactic phenomenon” (p. 345), but their analysis differs from Stewart’s in that it is based on word forms rather than on inflectional stems: in their model of Welsh mutations, the entire mutation paradigm of a word is exhaustively listed in the lexical entry, irrespective of whether a mutation type is incorporated or projected. In some cases, this can lead to homophony between two or more forms, as illustrated in (16): this may be due to exceptions or to the non applicability of a particular mutation to a specific consonant.<sup>9</sup>

## (16) Welsh mutation paradigm (simplified) (Mittendorf &amp; Sadler, 2006: 350)

|     | <i>gardd</i> ‘garden’ | <i>gêm</i> ‘game’ |
|-----|-----------------------|-------------------|
| Rad | <i>gardd</i>          | <i>gêm</i>        |
| AM  | <i>gardd</i>          | <i>gêm</i>        |
| SM  | <i>ardd</i>           | <i>gêm</i> [!]    |
| SMR | <i>ardd</i>           | <i>gêm</i> [!]    |
| NM  | <i>ngardd</i>         | <i>ngêm</i>       |

(AM = Aspirate Mutation, SM = Soft Mutation, SMR = Reduced Soft Mutation, NM = nasal mutation, [!] = exception.)

As shown in (16), the Rad and AM forms are identical for both nouns because /g/-initial words have no distinct AM form (only /p,t,k/-initial words do); the SM and SMR forms are identical for *gardd* because only initial /ɹ,r<sup>h</sup>/-initial nouns have distinct SMR forms (cf. §5.7.1.4); finally, the SM and SMR forms are identical to the Rad form for *gêm* because the latter represents an exception (cf. again §5.7.1.4). Their analysis shares with Kallen’s (1979) the assumption that the information necessary to define

<sup>9</sup>(16) is a simplified presentation of the full nominal mutation paradigm presented by Mittendorf & Sadler (2006). Forms and details which are not relevant to the present discussion have been left out.

mutation patterns is found in the lexicon, but, unlike in Kallen's account, exceptions are listed in the relevant lexical entry as part of its mutation paradigm.

Mittendorf & Sadler also distinguish two different mutation mechanisms which they term lexical and syntactic, corresponding to projected and incorporated, respectively. Lexical mutations (as in (17)) are conditioned by a particular lexical item, whereas syntactic ones depend on the morphosyntactic context and cover cases of agreement (e.g. mutation of a feminine singular noun after the article or of an adjective after a feminine singular noun).

In the case of lexical mutations, the lexical entry of the triggering word contains a selector which specifies the shape of the following word. For instance, in Welsh the third-person singular masculine possessive pronoun *ei* 'his' (cf. §5.7.1.9 below) will include in its lexical entry a selector +SM+ which has to be matched by the +SM+-indexed form of the following word, as in example (17).

(17) Welsh (Mittendorf & Sadler, 2006: 353)

- a.
- Lexical specification for *ei* 'his': +Rad+ ei +Pron +Pers +Proclit +3Sg +M +SM+
  - Lexical specification for *cath* 'cat': +Rad+ cath +Noun +F +Sg
  - Lexical specification for *gath* 'cat': +SM+ cath +Noun +F +Sg
- b.
- Grammatical sequence: +Rad+ ei +Pron +Pers +Proclit +3Sg +M +SM+ +SM+ cath +Noun +F +Sg → *ei gath* 'his cat'
  - Ungrammatical sequence: \*+Rad+ ei +Pron +Pers +Proclit +3Sg +M +SM+ +Rad+ cath +Noun +F +Sg → \**ei cath* 'his cat'

In the case of syntactic mutations, on the other hand, there is no lexical trigger and the selection of the mutated form is dictated by a rule, stating for instance that the +SM+ form of an adjective must be selected when it follows a feminine singular noun in the same noun phrase (Mittendorf & Sadler, 2006: 359).

### 5.3.3 Initial mutation and the lexicon in FDG

While FDG-inspired research in the structure of the lexicon has concentrated on issues of semantic and syntactic relevance such as agreement features, predication frames, and combinatorial constraints (e.g. Dikker, 2004; Honselaar & Keizer, 2009; Keizer & Honselaar, 2009), the paradigmatic dimension of the individual lexical entry does not appear to have been the object of any detailed analysis.

In what follows, I suggest that a morpholexical hypothesis akin to those examined in §5.3.2 can be integrated in FDG to account for the Insular Celtic phenomena under investigation. To summarize, two morpholexical models were presented:

1. Kallen's (1979) model, whereby only the radical form is listed in the lexical entry; special readjustment rules produce the expected phonological representations and specify relevant exceptions;
2. the mutation paradigm model, whereby a word's lexical entry specifies either (a) inflectional stems only (Stewart, 2004) or (b) every possible mutation form (Mittendorf & Sadler, 2006), and exceptions are listed in the paradigm.<sup>10</sup>

If we take the ScG form *chata* mentioned in example (15) above, we can formulate three hypotheses:

H1: the word form *chata* /xat<sup>h</sup>ə/ is listed in the lexical entry for *cat*, indexed as “vocative plural”;

H2: the word form *chata* /xat<sup>h</sup>ə/ is generated by an abstract rule like (18) based on an L-stem *chat* /xat<sup>h</sup>/, which is listed in the lexical entry together with the R-stem *cat* /k<sup>h</sup>at<sup>h</sup>/, the P-stem *cait* /k<sup>h</sup>at'/, etc.

$$(18) \text{ Vocative plural} = \text{L-stem} + /ə/$$

---

<sup>10</sup>According to Stewart, the range of possibilities implied by 1 and 2 are not mutually exclusive: he does not exclude the possibility that the lexicon may be organized by word forms rather than by inflectional stems and regards as still open the question whether speakers keep storing full paradigms in their lexicon based on their experience or only forms that are clearly irregular or based on less-common morphological rules (Stewart, 2004: 140, endnote 6).

The rule in (18) is abstract as it generalizes over all similarly-inflected nouns;

H3: only the radical stem *cat* /k<sup>h</sup>at<sup>h</sup>/ is specified in the lexical entry, with a diacritic like [±L], as suggested by Kallen, indicating that it may be subject to lenition;<sup>11</sup> assuming complete inflectional regularity, /xat<sup>h</sup>ə/ is generated by the application of two rules: an abstract morphological rule such as (19), valid for all nouns which inflect like *cat*, selects a positive value for [±L] and adds the suffix /ə/, and a readjustment rule like (20) converts [+L]/k<sup>h</sup>at<sup>h</sup>ə/ to /xat<sup>h</sup>ə/.

(19) Vocative plural = [+L]stem + /ə/

(20) [+L]/k<sup>h</sup>/ → /x/

From the discussion of §2.4, it will be recalled that in FDG fully lexicalized exceptions (suppletive forms) are considered part of a store of primitives feeding into the Phonological Encoder, while Lexemes are part of a separately accessed store of primitives which feeds into the Formulation component. Full or partial lexical specification, as in H1 and H2, respectively, would be considered necessary only in the case of irregular word forms (principle of lexical priority, cf. §2.4.3). All regular word forms are generated by the Morphosyntactic and the Phonological Encoder; H3 can therefore account for regular word-form generation in FDG, but an FDG-based account would differ from Kallen's in that exceptional word forms would be lexically stored rather than rule-generated. For instance, Welsh *gêm* 'game' /ge:m/, as illustrated in (16), keeps its initial /g/ under soft mutation. In Kallen's account, a rule would be posited whereby [+SM]*gêm* → [-SM]*gêm* based on the fact that *gêm* is flagged as exceptional in this regard; in FDG, however, the list of suppletive forms would contain a form /ge:m/ specified as [+SM]*gêm*, which would be selected by the Phonological Encoder in response to the ML-level term (Nw: [+SM]<sup>ε</sup>*gêm* (Nw)).

Consistent with the theoretical framework described in §2.4, [+SM], [+L] etc. would be considered Morphosyntactic Operators. The distinction between lexical and syntactic mutations which is found in Mittendorf

<sup>11</sup>And, for consistency, a diacritic like [±P] to signal that it also may be subject to morphological palatalization, therefore [±L]/k<sup>h</sup>at<sup>h</sup>/[±P].

& Sadler's (2006) account and which continues Oftedal's (1962) classification still holds, in the sense that a Morphosyntactic Operator like [+SM] can be inserted by the Morphosyntactic Encoder in response to a lexical item containing a specific selector, like Welsh *ei* 'his' (see (17) above), or as a morphosyntactic marker, as in the case of [+L] in the ScG vocative *chata*.

This account diverges from Mittendorf & Sadler's when it comes to the listing of regular forms. As illustrated by example (16) above, Mittendorf & Sadler propose that each mutated form of a word, whether regular or irregular, is listed in full; since in FDG this is assumed of suppletive forms only, regular correspondences between radical phonemes and their counterparts under different types of mutation would rather be treated as rule-generated. However, it should be noted that Mittendorf & Sadler do regard full listing of syntactically mutated word forms as "basically equivalent" to treating them "as a type of edge inflections", and ultimately disprefer the latter option only because it is more difficult to implement in their framework (Mittendorf & Sadler, 2006: 360).

## 5.4 Irish

Irish as a community language is still spoken in the Gaeltacht areas of western Ireland. The main dialects are three: Ulster, Connacht, and Munster (Ó Siadhail, 1989; Watson, 1989). In what follows, reference will be made to the relevant dialectal differences when necessary. Language revitalization policies are in place to maintain and promote the status of Irish as a community language in these areas (Akutagawa, 1987; Breatnach, 1964; Harris & Murtagh, 1987; Hindley, 1990; Macnamara, 1971; Watson, 1989). Irish speakers are by now all bilingual with English (cf. Ó Murchú, 1988: 248) and represent a minority of the population: the available figures from the 2006 Census tell us that 40.8% of the population aged three years and over identified themselves as Irish speakers; however, only about 44% of the latter (or 18% of the total) self-reported that they ever used Irish outside the education system and only 3% of them (or 1% of the total) that they did so daily (CSO, 2006: Tables 1 and 40; also see Punch, 2008).

### 5.4.1 Realization and scope of agreement

There are two basic types of initial mutations in Irish (Ó Siadhail, 1989: §6.2): lenition and nasalization (or “eclipsis”); *h*-prefixing (or “provection”) is also commonly referred to as an initial mutation although it is a sound insertion rather than a sound mutation; likewise *n*-prefixing, which is traditionally considered as the application of nasalization to vowel-initial words, and *t*-prefixing, which also applies to vowel-initial nouns. Each mutation type only applies to a specific set of phonemes.

#### 5.4.1.1 Lenition

Under lenition, labial and velar plosives are replaced by homorganic fricatives, dental plosives /t,t'/ and /d,d'/ by the back fricatives /h,h'/ and /ɣ,ɣ'/, respectively; lenition of initial *l,r,n* (described as [+tense]) consists in a loss of tension (Ó Siadhail, 1989: 111); fricative *s* is also replaced by *h*, while *f* is deleted. A prospect of lenition in Irish is given in Table 5.1.<sup>12</sup>

The following observations are made by Ó Siadhail (1989: 112–14) apropos of Irish lenition:

1. In the late 1980s, phonemic opposition between [+tense] /N N', L L', R R'/ and [–tense] /n n', l l', r r'/ was retained only by the oldest speakers of some dialects of Munster Irish. Elsewhere, distinctive use of this feature was no longer found.<sup>13</sup>
2. The lenition of initial *f*- is rather unstable in some dialects, e.g. Connacht Irish.
3. The lenition of initial *s*- depends the following phoneme. Lenition is possible when *s*- is followed by a vowel or by *l* or *r*; in some Munster dialects, *s*- may also lenite when followed by *m*. Consonant clusters in which *s* is followed by a plosive are not lenited.

<sup>12</sup>In what follows, initial mutations will be illustrated with reference to the two phonemically distinctive realizations that are possible with all consonants in the Goidelic languages, namely palatalized and non-palatalized (cf. §1.4). When no distinction is implied, the conventional orthographic notation is used: e.g. *t* may refer to either /t/ or /t'/.

<sup>13</sup>Cf. de Bhaldraithe (1953: 257, fnn. 1–3), Ó hUiginn (1994: 560, 562), Ua Súilleabháin (1994: 488), Hickey (2003: 264), Ó Curnáin (2007a: 188–201).

Table 5.1: Lenition in Irish

|                   | Basic consonant |         | Mutated consonant |                   |
|-------------------|-----------------|---------|-------------------|-------------------|
| <b>Plosives</b>   | ⟨c⟩             | /k k' / | ⟨ch⟩              | /x x' /           |
|                   | ⟨g⟩             | /g g' / | ⟨gh⟩              | /ɣ ɣ' /           |
|                   | ⟨t⟩             | /t t' / | ⟨th⟩              | /h h' /           |
|                   | ⟨d⟩             | /d d' / | ⟨dh⟩              | /ɣ ɣ' /           |
|                   | ⟨p⟩             | /p p' / | ⟨ph⟩              | /f f' /           |
|                   | ⟨b⟩             | /b b' / | ⟨bh⟩              | /w v /            |
| <b>Fricatives</b> | ⟨s⟩             | /s s' / | ⟨sh⟩              | /h h' /           |
|                   | ⟨f⟩             | /f f' / | ⟨fh⟩              | /∅ <sup>j</sup> / |
| <b>Nasals</b>     | ⟨n⟩             | /N N' / | ⟨n⟩               | /n /              |
|                   | ⟨m⟩             | /m m' / | ⟨mh⟩              | /w̃ v /           |
| <b>Laterals</b>   | ⟨l⟩             | /L L' / | ⟨l⟩               | /l l' /           |
| <b>Vibrants</b>   | ⟨r⟩             | /R R' / | ⟨r⟩               | /r' /             |

(Source: Ó Siadhail, 1989: 112)

4. Lenition between two dental segments (i.e. lenition of a dental preceded by another dental) may be blocked in certain contexts; the segments involved include *t*, *d*, *r*, *l*, *n*. A relevant example is the blocking of lenition after the article, as shown in (21).

(21) Irish (Wigger, 2000 4-03-10)

an<sup>(L)</sup> dúchan  
 ART.NOM.F.SG potato blight(F).NOM.SG

‘the potato blight’

However, lenition between two dentals might be preserved between the head noun and an attributive dependent (adjective or noun), as in (22).

(22) Cois Fhairrge Irish (Ó Siadhail, 1989: 113)

a. móin<sup>L</sup>                      dhhubh  
 peat(F).NOM.SG black.NOM.F.SG  
 ‘black peat’

b. sloitín<sup>L</sup>                      dhraíocht(a)  
 wand(F).NOM.SG magic(F).GEN.SG  
 ‘magic wand’

5. Lenition in loanwords is of limited application; in particular, initial *t*- and *d*-, as well as *s*- and *f*-, tend to resist lenition.

#### 5.4.1.2 Nasalization or “eclipsis”

Under eclipsis (Table 5.2), voiced plosives are replaced by homorganic nasals, voiceless plosives by their voiced counterparts, and /f/ and /f’/ by /w/ and /v’/, respectively. Also described as eclipsis is the prefixing of *n*- to vowel-initial words.

Table 5.2: Nasalization or “eclipsis” in Irish

|                   | Basic consonant |        | Mutated consonant |        |
|-------------------|-----------------|--------|-------------------|--------|
| <b>Plosives</b>   | c)              | /k k’/ | <gc>              | /g g’/ |
|                   | <g>             | /g g’/ | <ng>              | /ŋ ŋ’/ |
|                   | <t>             | /t t’/ | <dt>              | /d d’/ |
|                   | <d>             | /d d’/ | <nd>              | /N N’/ |
|                   | <p>             | /p p’/ | <bp>              | /b b’/ |
|                   | <b>             | /b b’/ | <mb>              | /m m’/ |
| <b>Fricatives</b> | <f>             | /f f’/ | <bhf>             | /w v/  |

(Source: Ó Siadhail, 1989: 112)

#### 5.4.1.3 *h*-prefixing or “provection”

According to Ó Siadhail (1989: 122–124), provection is the occurrence of hiatus between two vowels, one of which would otherwise be elided, and is represented in writing by the letter *h* prefixed to the second vowel. (Its realization need not therefore be an audible /h/ sound as long as the hiatus is

maintained.) Provection is exploited grammatically in a number of meaningful distinctions, including gender agreement. Remarkably, in dialects where /h/ is regularly dropped between two vowels (cf. Ó Siadhail, 1989: 81f.), it is retained if a grammatical opposition depends on it: so *ní hiad* ‘it is not them’ (where the presence of /h/ is not meaningful) is often heard as *ní iad* in the Irish of Cois Fhairrge (Connacht), whereas /h/ is likely to be heard in *a hathair* ‘her father’ (cf. *a athair* ‘his father’)—where it distinguishes the gender of the possessor.

#### 5.4.1.4 Article

There is agreement in number, gender and case between the definite article and the head noun.<sup>14</sup> The forms of the article are shown in Table 5.3, where the superscript letters indicate the co-occurrence of mutation on the following noun. The following should be observed:

1. superscript *T* indicates the prefixing of *t-*, which only applies to nouns that begin with a vowel, e.g. *an t-asal* ‘the donkey’.
2. Lenition is blocked if the noun begins in a dental (see §5.4.1.1 above).
3. Superscript *L+T* represents ordinary lenition except in the case of lenitable /s s’/, which is deleted (but still written) and replaced by /t t’/ (e.g. *súil* /sul’/ ‘eye’ : *an tsúil* /ən’tu:l’/ ‘the eye’, *srón* /sro:n/ ‘nose’ : *an tsrón* /ən’tro:m/ ‘the nose’, cf. Mac Eoin, 1993: 113; Ó Siadhail, 1989: 127).<sup>15</sup>
4. *h-* in the genitive feminine is prefixed to a noun beginning with a vowel, e.g.

(23) Irish (Wigger, 2000 5-01-05)

i rith na                    haimsir-e  
in flow the.GEN.F.SG time(F)-GEN.SG

‘during that time’

<sup>14</sup>There is no indefinite article in Goidelic.

<sup>15</sup>Historically, this *t-* results from the encounter of the ancient ending *-d* of the article with the *h-* resulting from the lenition of *s-*: /d#h/ > [t] (cf. Ball & Müller, 1992: 48).

5. “Prepositional I” and “prepositional II” refer to the two mutation patterns which occur when the article follows a preposition, and do not distinguish between grammatical genders, except when the noun begins with lenitable *s-*. Lenition is blocked before dentals and so is (in standard Irish and in Connacht) nasalization (Ó Siadhail, 1989: 127).

Table 5.3: Article paradigm in Irish: singular forms

|                  | Masculine         | Feminine                           |
|------------------|-------------------|------------------------------------|
| Nominative       | an <sup>T</sup>   | an <sup>L+T</sup>                  |
| Genitive         | an <sup>L+T</sup> | na <sup>H</sup>                    |
| Prepositional I  | an <sup>N</sup>   | an <sup>N</sup> /an <sup>L+T</sup> |
| Prepositional II | an <sup>L</sup>   | an <sup>L+T</sup>                  |

(Cf. Mac Congáil, 2004: 20–23)

To recap, gender agreement between the noun and the article may only be found in the singular; in the nominative, agreement is possible only if the noun begins in either a lenitable consonant or a vowel; after a preposition, only if the noun begins with a lenitable *s-*.

#### 5.4.1.5 Adjectives

Gender agreement with adjectives is limited to the attributive position in the singular: predicative adjectives are not inflected (Mac Eoin, 1993: 115). Adjectives may mark gender agreement either by initial mutation and, depending on their morphological class, also by final inflection, as illustrated in Table 5.4. The relevant distinctions can be summarized as in Table 5.5. Gender agreement is not marked in the vocative; after certain prepositions (prepositional II) adjective agreement may be neutralized by the optional lenition of the adjective after a masculine noun.<sup>16</sup>

In a sequence of adjectives, mutation is normally carried over, as in (9)

<sup>16</sup>However, after a masculine noun an adjective is more commonly left unlenited if the noun itself is not lenited: e.g. *ar an hata dubh* ‘on the black hat’, *san uisce te* ‘in the hot water’, *ag an doras mór* ‘at the big door’ (NIG, 2004: 18).

Table 5.4: Adjective inflection in Irish

|              | Masculine ‘the great man’ |        |               | Feminine ‘the great woman’ |       |              |
|--------------|---------------------------|--------|---------------|----------------------------|-------|--------------|
|              | Art.                      | Noun   | Adj.          | Art.                       | Noun  | Adj.         |
| Nom. sg.     | an                        | fear   | mór           | an                         | bhean | <u>mh</u> ór |
| Gen. sg.     | an                        | fhir   | <u>mh</u> óir | na                         | mná   | móire        |
| Prep. I sg.  | (leis) an                 | bhfear | mór           | (leis) an                  | mbean | <u>mh</u> ór |
| Prep. II sg. | (do)n                     | fhéar  | m(h)ór        | (do)n                      | bhean | <u>mh</u> ór |
| Voc. sg.     | (a)                       | fhir   | <u>mh</u> ór  | (a)                        | bhean | <u>mh</u> ór |

(Cf. Mac Congáil, 2004: 89f.)

Table 5.5: Features of adjective inflection in Irish

|              | Masculine                  | Feminine                      |
|--------------|----------------------------|-------------------------------|
| Nom. sg.     | unmarked                   | lenition                      |
| Gen. sg.     | lenition + palatal. switch | palatal. switch + suffixation |
| Prep. I sg.  | unmarked                   | lenition                      |
| Prep. II sg. | unmarked or lenition       | lenition                      |
| Voc. sg.     | lenition                   | lenition                      |

(repeated below as (24)), but it may also be blocked by an intervening non-mutable adjective, as in (25).

(24) Irish (Ó Sé, 2000: 65)

an                    fhead                    bheag                    dh eas  
 ART.NOM.F.SG whistle(F).NOM.SG little.NOM.F.SG nice.NOM.F.SG  
chaol  
 slender.NOM.F.SG

‘the nice, small, slender whistle’

(25) Irish (ibid.)

oíche            bhreá            réilteannach            geal  
 night(F).SG fine.NOM.F.SG starry.NOM.F.SG bright.NOM.M.SG

‘a fine, bright, starry night’

Adjective classes are defined by the different realization of the genitive singular, masculine and feminine; different grammars list different numbers of adjective classes, according to what criteria are used (cf. NIG, 2004: 63ff.; Mac Congáil, 2004: 86ff.). Simplifying, these two forms are derived from the nominative in one or two steps, involving the palatalization (or depalatalization) of the final consonant and/or the addition of a further syllable by suffixing /ə/.

To summarize, gender agreement with adjectives is distinguished, in the nominative, only by the initial lenition. In the genitive, both lenition and final inflection may play a role in gender agreement, depending on both the initial consonant and the morphological class of the adjective. After a preposition, unmarked masculine and lenited feminine are the norm, although dialectal variation also plays an important role (see §5.4.4).

However, it is important to observe that final inflection of the adjective is disappearing:

[it] continued to be the norm in the literary language until very recent times, though there was a considerable reduction of form in use. In the modern spoken language, declension of the adjective in the singular has largely been abandoned, though old forms survive in many set phrases [Mac Eoin, 1993: 115f.]

Furthermore, continues Mac Eoin, what little adjective inflection is left is threatened by the gradual decline of the attributive position altogether, which is felt to be “bookish” and tends to be avoided in spoken Irish, where “[v]arious stratagems are used to ensure that the adjective is usually in predicative construction” (H. Wagner, cited in Mac Eoin, 1993: 116).

As already noted (in §1.3 above), the phonological system in use by the younger generations in certain geographical varieties tends to abandon the distinction between palatalized and non-palatalized consonants, thus contributing to further undermine agreement marking in the adjective, which seems set to become a completely indeclinable element even in the attributive position.

#### **5.4.1.6 Attributive nouns**

Dependent nouns which occupy the same position as the attributive adjective, i.e. the postnominal position, may under certain circumstances undergo

lenition, as in example (26).<sup>17</sup>

(26) Irish (Wigger, 2000 5-01-14)

drochoíche      bháistí  
bad night(F).SG rain.GEN

‘a bad rainy night’

When, as in this example, the dependent noun is indefinite, its lenition is conditional on the head noun being feminine; its occurrence is therefore considered as an agreement phenomenon.

When the dependent noun is definite, on the other hand, it is expected to show up as lenited always, regardless of any agreement feature, as in (27), where the proper name is mutated after both feminine *bean* and masculine *teach*.

(27) Irish

a. (Wigger, 2000 7-01-14)

go leor de                      mhuintir      Chonamara  
a lot of    people(F).NOM.SG Connemara

‘a lot of Connemara people’

b. (Wigger, 2000 6-01-01)

fear                      Chonamara  
man(M).NOM.SG Connemara

‘a Connemara man’

A definite noun is either a proper name (27) or a noun which heads a genitive noun phrase (28) (cf. NIG, 2004: 25f.).<sup>18</sup>

(28) Irish (NIG, 2004: 16)

a. mac                      [fhear                      an  
son(M).NOM.SG man(M).NOM.SG ART.GEN.M.SG  
tí]  
house(M).GEN.SG

‘the landlord’s son’

<sup>17</sup>The description that follows is based on Ó Siadhail (1989: 119–21), Mac Congáil (2004: 58–60), and the *New Irish Grammar* (NIG, 2004: 15f.); where possible, examples are taken from OC-I.

<sup>18</sup>A noun phrase is also made definite by the article, but this is irrelevant here as the presence of the article means that the two nouns are not adjacent.

- b. obair                    [b]hean                    an  
 work(F).NOM.SG woman(F).NOM.SG ART.GEN.M.SG  
 tí]  
 house.GEN.M.SG  
 ‘the housewife’s work’

Indefinite nouns, on the other hand, are only lenited after a feminine noun in the nominative singular, as mentioned above. Further examples are shown in (29–30).

(29) Irish (NIG, 2004: 15)

- a. an    ghaoth    Mháirta  
 the.F wind(F).SG March.GEN  
 ‘the March wind’
- b. cúis                    gháire  
 cause(F).SG laughter  
 ‘a reason for laughter’

(30) Irish (Wigger, 2000 1-01-14)

- diarthrá [= díthrá] mhara  
 low-tide(F).NOM.SG sea.GEN  
 ‘low-tide’

Certain contexts in which this rule may not apply, some semantically and some syntactically motivated, are sometimes mentioned. Semantically motivated contexts include the following:

1. if the modified noun “denotes excess, part, want” (NIG, 2004: 15) or in general “quantity” (cf. Mac Congáil, 2004: 59): e.g. *a ndóthain móna* (Wigger, 2000 4-01-03) ‘enough turf for them’ (lit. ‘their sufficient amount of turf’); a notable exception is represented by *beirt*, a feminine noun used to refer to pairs of persons (e.g. *an<sup>L</sup> bheirt bhádóirí*, Wigger, 2000 1-01-25 ‘the two boatsmen’), after which mutation is commonly found;
2. if it denotes body parts, or parts of a thing or apparatus (Mac Congáil, 2004: 59), e.g. *cois gasúir* (Wigger, 2000 6-01-10) ‘a boy’s foot’, *teanga sionnach* (Wigger, 2000 6-01-08) ‘a fox’s tongue’;

3. if it is followed by a partitive, e.g. *a gcuid Gaeilge* ‘their Irish’ (lit. ‘their share of Irish’) (Wigger, 2000 2-04-10),<sup>19</sup> or appositive genitive, e.g. *óinseach mná* ‘a fool of a woman’ (Ó Siadhail, 1989: 121);
4. if it denotes something owned by, or meant for, the referent of the following noun: e.g. *culaith fir* ‘a man’s suit’, *bróg páiste* ‘a child’s shoe’ (cf. Mac Congáil, 2004: 60);
5. and normally, but not necessarily, if it expresses the logical subject or object of a verbal noun,<sup>20</sup> e.g.

(31) Irish (Wigger, 2000 1-01-23)

ag tabhairt                      misneach don      dream  
at giving(F).NOM.SG courage      to.ART group

‘encouraging the group’

(32) Irish (Mac Congáil, 2004: 60)

beannacht                      baintrí  
blessing(F).NOM.SG widow(F).GEN.SG

‘a widow’s blessing’

As regards syntactic contexts, an attributive noun may or may not be lenited if it is followed by an attributive adjective (cf. example (11) in §5.3.2), or if it is the argument of a prepositional phrase (33–34).

(33) Irish (Mac Congáil, 2004: 60)

i láthair                      múinteora  
in presence(F).NOM.SG teacher.GEN.SG

‘in the presence of a teacher’

<sup>19</sup>Note that *Gaeilge* is not a definite noun in Irish and is made definite by the article, e.g. *an Ghaeilge*.

<sup>20</sup>Ó Siadhail (1989: 121) observes that “there is a core of phrases common to all dialects” in which the logical object of a verbal noun is lenited: e.g. *ag fáil bháis* ‘dying’ (*fáil* is a feminine verbal noun), *ag cur fhataí* ‘sowing potatoes’ etc. Note that this does not depend on the gender of the verbal noun: *cur* ‘putting’, for instance, is masculine. According to Ó Siadhail, lenition after the verbal noun mainly occurs with high-frequency, common verbs.

(34) Irish (Wigger, 2000 2-13-17)

le haghaidh deoch  
with face(F).NOM.SG drink.NOM.SG

‘for a drink’

The occurrence of lenition with indefinite dependent nouns seems to suggest a distinction between modifiers and arguments (see §2.4.1 above). Lenition after a feminine noun seems to be found with modifiers, including adjectives and nouns used as attributes, but not with arguments. This does not seem to be, however, a clear-cut distinction; there seems to be prototypical and peripheral cases, as is confirmed by qualifications such as “frequently”, “normally but not always” and the like, which are normally found in the traditional descriptions of these phenomena, as is also apparent from the above discussion.

Prototypical arguments in the examples above would clearly be the logical subjects and objects of verbal nouns, as in examples (31–32), as well as the arguments of prepositional phrases as in (33–34). Other clear examples of arguments are relational nouns indicating possession, body parts, parts of wholes, portions of masses etc., as in, respectively, the aforementioned *bróg páiste* ‘a child’s shoe’, *cos gasúir* ‘a boy’s foot’.

Central examples of attributive nouns (modifiers) are those in which the dependent’s denotatum is fairly low in specificity and/or individuation, as in (26) (*drochoíche bháistí* ‘a bad rainy night’), *an ghaoth Mhárta* ‘the March wind’ (29a) (any March, not referring to the March of a particular year) and *cúis gháire* ‘a reason for laughter’ (29b), i.e. something funny or ridiculous (again not referring to a specific burst of laughter).<sup>21</sup>

#### 5.4.1.7 Possessives

Irish possessives are genitive personal pronouns agreeing in gender with their antecedent (the “possessor”). They are characterized by a mutational pattern which in the third person distinguishes masculine from feminine and singular from plural (Table 5.6).

<sup>21</sup>However, other cases are less straightforward: *gaoithe* in example (11) (*oíche gaoithe móire* ‘a night with a high wind’) is used attributively but left unlenited; being modified by an adjective, however, its degree of specificity is greater than would be conveyed by the simple noun.

Table 5.6: Possessives in Irish (third person)

| Person  | Possessive          | Example            | Gloss        |
|---------|---------------------|--------------------|--------------|
| 3sg. m. | a /ə <sup>L</sup> / | a <u>b</u> hád     | his boat     |
|         |                     | a athair           | his father   |
| 3sg. f. | a /ə <sup>H</sup> / | a bád              | her boat     |
|         |                     | a <u>h</u> athair  | her father   |
| 3pl.    | a /ə <sup>N</sup> / | a <u>m</u> bád     | their boat   |
|         |                     | a <u>n</u> -athair | their father |

#### 5.4.1.8 Personal pronouns

Non-possessive pronouns agree in gender in the third person singular and can be divided into simple and prepositional (Table 5.7). Additionally, emphatic forms can be derived by means of suffixation (masculine /sən, fən/ vs. feminine /sə, fə/).

Table 5.7: Examples of third-person pronouns in Irish

|       | Non-emphatic    |                    | Emphatic                     |                               |
|-------|-----------------|--------------------|------------------------------|-------------------------------|
|       | Simple          | Prepositional      | Simple                       | Prepositional                 |
| masc. | sé<br>'he, it'  | dó<br>'to him, it' | seisean<br>'himself, itself' | dósan<br>'to himself, itself' |
| fem.  | sí<br>'she, it' | di<br>'to her, it' | sise<br>'herself, itself'    | dise<br>'to herself, itself'  |

(Cf. NIG, 2004: 84)

#### 5.4.2 Semantic assignment

Gender is highly predictable in Irish for animate referents. Names for male animates tend to be masculine, names for female animates feminine. There are very few exceptions, e.g. *stail* 'stallion' (feminine) and *cailín* 'girl' (masculine).

Semantic assignment has been suggested to work at another level as well:

Mac Congáil (2004: 27), for instance, observes that the names of most continents, countries, rivers, and languages are feminine, while Acquaviva (2006: 1864f.) points out that most units of measure have feminine designations: *bliain* ‘year’, *pingin* ‘penny’, *seachtain* ‘week’, *scilling* ‘shilling’, *uair* ‘time, occasion’, etc.

When semantic assignment is at variance with formal assignment, semantic agreement tends to prevail with anaphoric pronouns. This is always the case with humans but not always with other animates: agreement with f. *stail* ‘stallion’, for instance, is at least traditionally realized by a feminine pronoun (Ó Siadhail, 1984: 175; 1989: 148). There are also dialectal idiosyncrasies: *francach* ‘rat’ and *capall* ‘horse’ (both masculine) are commonly referred to by feminine pronouns in the Irish of Cois Fhairrge, and in Kerry m. *rón* ‘seal’ may systematically be referred to by feminine pronouns (Ó Siadhail, 1984: 175). In general, in native varieties, pronominal agreement might still be consistent with grammatical gender even when the sex of an animal is known, unless the emphasis is on the “reproductive or sexual difference” (Ó Siadhail, 1984: 175).

Finally, feminine pronouns are also consistently used with a number of masculine antecedents, roughly falling within three classes of meaning (machines, vehicles and garments): e.g. *measín* ‘machine’, *bád* ‘boat’, *leoraí* ‘lorry’, *carr* ‘car’, *geansaí* ‘jersey’, *caipín* ‘cap’. Ó Siadhail (1984: 175) observes that this is the case in all major dialects and regards the phenomenon as a case of personification. Two very basic terms, both feminine—*áit* ‘place’ and *uair* ‘hour; time, occasion’—are also commonly referred to by a masculine anaphoric pronoun (Ó Siadhail, 1984: 175).

In Gaoth Dobhair Irish, masculine pronouns are used anaphorically for all inanimate referents (Ó Siadhail, 1984: 175), a usage which appears to be well established even among odler speakers.

### 5.4.3 Formal assignment

From a formal point of view, certain morphological and/or phonological features of Irish nouns permit, more or less reliably, to predict their gender, and it is common to describe gender assignment in terms of inflectional paradigms.

Traditionally, Irish nouns are divided into five declensions, distinguished by the forms of the nominative and genitive singular (cf. Ó Dochartaigh,

1992: 65; Mac Eoin, 1993: 114f.; NIG, 2004: chapters 8–12), as shown in Table 5.8.<sup>22</sup>

Table 5.8: Irish declensions

| Decl. | Ending       |                    | Gender    | Example            | Gloss  |
|-------|--------------|--------------------|-----------|--------------------|--------|
|       | Nom. sg.     | Gen. sg.           |           |                    |        |
| I     | C            | C'                 | m. only   | crann, crainn      | tree   |
| II    | C or C'      | C' + /ə/           | f. only   | cearc, circe       | hen    |
| III   | C or C'      | C + /ə/            | m. and f. | dochtúir, dochtúra | doctor |
| IV    | /i:n'/ or V  | = nom. sg.         | m. and f. | buille             | stroke |
| V(a)  | /l', r', n'/ | /l, r, n/ (+ /əx/) | mostly f. | riail, rialach     | rule   |
| V(b)  | V            | V (+ /n/ or /d/)   | mostly f. | pearsa, pearsan    | person |

As can be seen from Table 5.8, the first and second declensions correlate with the masculine and feminine gender respectively, and exclusively,<sup>23</sup> and most fifth-declension nouns are also feminine.<sup>24</sup> Within the third declension, it is possible to tell masculines from feminines on a morphological or phonological basis (number of syllables and shape of word ending; cf. NIG, 2004: 47–50): nouns ending in the agentive suffixes like /ər'/ (-*aeir*, -*eoir* etc.) and /e:r'/ (-*éir*) are masculine (e.g. *rinceoír* ‘dancer’); polysyllables ending in /əxt/ (-*acht*, -*ocht*) are feminine; monosyllables tend to be feminine if they end in a palatalized consonant, masculine if they end in a non-palatalized one.

In the fourth declension, nouns ending in /i:n'/ (-*ín*) are normally masculine, except when -*ín* is used to derive a diminutive, in which case the gender of the lexical primitive is retained: e.g. *beainín* ‘little woman’ (< *bean* ‘woman’, f.), *páircín* ‘little field’ (< f. *páirc* ‘field’) (Ó Siadhail, 1989: 145), and cf. (35);

<sup>22</sup>There are also a number of irregularly inflected nouns, both masculines and feminines.

<sup>23</sup>With very few exceptions in the second declension, e.g. masculine *im* ‘butter’, *sliabh* ‘mountain’ and *teach* ‘house’ (NIG, 2004: 42).

<sup>24</sup>There are very few masculine nouns in the fifth declension, most notably *athair* ‘father’, *bráthair* ‘brother (in a religious congregation)’, *deartháir* ‘brother (a male sibling)’, which have male referents.

(35) Irish (Wigger, 2000 2-04-09)

an                      bhó-ín                      bheag  
 ART.NOM.F.SG cow(F)-DIM.SG little.NOM.F.SG

‘the little cow’

if the derived noun is not a diminutive of the base, on the other hand, its gender will be masculine: e.g. *céirín* (m.) ‘poultice’ (< *céir* ‘wax’, f.), *paidrín* (m.) ‘rosary’ (< *paidir* ‘prayer’, f.).<sup>25</sup>

In contemporary spoken Irish the marking of the genitive is now rare in all dialects (cf. Hughes, 1994: 628–635; Ó hUiginn, 1994: 563–567; Ua Súilleabháin, 1994: 492–493). Since the five declensions cannot be kept distinct in the absence of genitive marking, the total collapse of the case system would make morphological gender assignment no longer possible, and there are indications that phonology-based gender re-assignment may be under way in some native varieties. In particular, since non-palatalized consonant endings tend to be associated with the masculine gender and palatalized ones with the feminine gender, two developments have been observed whose outcome reinforce the existing tendency:

1. the quality of the final consonant is changed and the traditional grammatical gender retained (Ó Siadhail, 1984: 174 and 1989: 145). *Fuinneog* ‘window’ and *muc* ‘pig’, for instance, are feminine nouns ending in a non-palatalized consonant whose traditional nominative forms are replaced by forms with palatalized endings (*fuinneoig* and *muic* respectively) in some Connacht dialects and among younger Kerry Irish

<sup>25</sup>However, some diminutives in *-ín* deriving from feminine nouns may trigger feminine agreement in the nominative and masculine agreement in the genitive (NIG, 2004: 51–53): cf. *páircín* ‘small field’ (from f. *páirc* ‘field’) in example (i).

i. Irish (NIG, 2004: 51–53)

a. an<sup>L</sup>                      pháircín                      bheag  
 ART.NOM.F.SG small field.NOM.SG small.NOM.F.SG  
 ‘the small little field’

b. bun                      [an<sup>L</sup>                      pháircín                      bhig]  
 bottom(M).NOM.SG ART.GEN.M.SG small field.GEN.SG small.GEN.M.SG  
 ‘the bottom of the small little field’

speakers. Likewise, the masculine agentive suffix /e:rə/ (-*éara*) is replacing /e:r'/ (-*éir*) in Connacht Irish (e.g. *tincéara* instead of *tincéir* ‘tinker’, *búistéara* instead of *búistéir* ‘butcher’, cf. Wigger, 2000 1-05-20, 4-05-02).

2. Alternatively, the quality of the final consonant is retained but the gender re-assigned: e.g. in some Connacht dialects feminine monosyllables ending in /əxt/ are treated as masculines (e.g. *fuacht* ‘cold’ in Cois Fhairrge Irish), while other nouns, like *deatach* ‘smoke’, oscillate between the two genders, as in example (36).

(36) Connemara Irish

- a. (Wigger, 2000 6-17-05)

deatach            trom            dubh  
 smoke.NOM.SG heavy.NOM.M.SG dark.NOM.M.SG

‘dark, heavy smoke’

- b. (Ó Siadhail, 1989: 146)

baladh            na            deataí  
 smell(M).NOM.SG ART.GEN.F.SG smoke.GEN.SG

‘the smell of smoke’

Therefore, Ó Siadhail’s (1989: 143–45) prefers to describe formal gender assignment not in terms of declensions but combining phonological features (word-final segments) with morphosemantic ones (derivational suffixes), as shown in Tables 5.9 and 5.10. The decisive factor is whether the final consonant is palatalized or not; the exceptions are masculine agentive derivations (mostly referring to functions historically performed only by men), derivations in *-ín* (see above) and the aforementioned non-palatalized feminine endings in /əx/ and /əxt/.

#### 5.4.4 Dialectal variation and double gender

Dialectal variation involves two distinct aspects: (i) morphosyntactic variation and (ii) idiosyncratic gender assignment. The two will be considered in turn.

Table 5.9: Formal gender assignment in Irish (masculine)

| Type           | Ending                    | Example             | Gloss       |
|----------------|---------------------------|---------------------|-------------|
| Agent suffixes | -óir /or'/, -eoir /er' /  | badóir              | boatman     |
|                | -(a)ire /ar'e /           | cabaire             | prattler    |
|                | -(a)í /i' /               | scéalaí             | storyteller |
|                | -éir /er' /, -éara /erə / | búistéir, búistéara | butcher     |
| Ethnonyms      | -nach /nex /              | Éireannach          | Irishman    |
| Various        | -as /as /                 | doras               | door        |
|                | -án /an /                 | cupán               | cup         |
|                | -(e)an /an /              | leagan              | version     |
|                | -ín /in' /                | caíín               | girl        |
|                | -(e)adh /e /              | moladh              | praise      |
|                | -(e)amh /av' /, /u /      | caitheamh           | use         |
|                | -ás /as /                 | sólás               | consolation |
|                | -úr /ur' /                | casúr               | hammer      |
|                | -ín /u:n /                | naisiún             | nation      |

(After Ó Siadhail, 1989: 143–145)

Table 5.10: Formal gender assignment in Irish (feminine)

| Type                     | Ending                           | Example    | Gloss               |
|--------------------------|----------------------------------|------------|---------------------|
| Mass-noun suffixes       | -(e)ach /əx/                     | báisteach  | rain                |
| Various                  | -ó(i)g /o:g(ˈ)/, -eo(i)g /əg(ˈ)/ | lasóg      | small light         |
|                          | -(e)acht /əxt/                   | Gaeltacht  | Irish-speaking area |
|                          | -seach /ʃəx/                     | cláirseach | harp                |
|                          | -éis /e:ʃ/                       | móiréis    | haughtiness         |
|                          | -áil(t) /a:l'(tˈ)/               | spáráil    | sparing             |
|                          | -(a)irt /ər'tˈ/                  | bagairt    | threat              |
|                          | -(a)ilt /əl'tˈ/                  | meilt      | grinding            |
| -(a)int /ən'tˈ/          | caint                            | talk       |                     |
| Various (non-productive) | -aois /i:f/, -ís /i:ʃ/           | mailís     | malice              |
|                          | -áid /a:dˈ/                      | úsáid      | use                 |
|                          | -óid /o:dˈ/                      | neascóid   | boil                |

(After Ó Siadhail, 1989: 143–145)

#### 5.4.4.1 Morphosyntactic variation

The treatment of prepositional phrases largely depends on the specific dialect (Jackson, 1942: 272f.). In Connacht Irish, as in the Standard, the mutation pattern we called prepositional I is found after all prepositions except *de* ‘from/of’ and *do* ‘to/for’; after the latter two, the prepositional-II pattern is used instead, and in neither case are initial *t-* and *d-* affected. This is also the case in most of County Clare (Munster), with the only difference that initial *t-* and *d-* are not immune from eclipsis. In northern dialects, i.e. in Ulster and West Mayo, prepositional II is found with all prepositions, whereas (according to T. F. O’Rahilly, cited in Jackson, 1942: 273) the opposite is true in Waterford and some parts of Kerry (Munster dialects), as well as in some parts of northern Connacht, where it is prepositional I that is found after all prepositions, including *de* and *do*. Lenition rather than eclipsis is the norm after the preposition *i(n)* ‘in’ in West Munster dialects.

This dialectal variation does not affect the marking of gender agreement between article and noun, but might be expected to affect the realization of gender agreement with adjectives: as observed in §5.4.1.5 (Table 5.4), prepositional I has a stronger potential for gender distinction than prepositional II, since adjectives may be lenited regardless of gender if the noun is, so that in dialects where prepositional II is generalized we should expect more instances in which adjective agreement is neutralized. However, two studies conducted on Munster dialects (Breatnach (1958) on the Irish of West County Waterford and South Tipperary, and Jackson (1942) on Blasket Irish) paint a more irregular picture. Both focus on the variable occurrence of eclipsis and lenition in prepositional phrases, and show that the same prepositions may occur with different mutation patterns depending on phonological rather than syntactic factors; namely, whether the preposition itself ends in a vowel or a consonant and whether the noun begins with a voiced or a voiceless consonant; which may in turn affect the marking of the attributive adjective (Jackson, 1942: 274f.). A particularly interesting observation is made apropos of prepositional phrases with no article, where the noun mutates after each preposition depending on historical factors and regardless of gender, while the mutation of the attributive adjective still depends on both lexical gender and the preposition itself: after *de* and *do*, the adjective is lenited always and regardless of gender; with all other preposi-

tions, lenition of the adjective is found only in feminine noun phrases. The only exception seems to be represented by the preposition *i(n)*, after which the lenition of the adjective is optional and independent of gender.

In summary, gender agreement with adjectives is not always a straightforward matter, and may depend on a range of different factors: dialectal, lexical and phonological.

#### 5.4.4.2 Idiosyncratic gender assignment

Another aspect of dialectal variation has to do with what Ó Siadhail (1984: 174; 1989: 145) calls double gender (henceforth, DG). He distinguishes two types of DG nouns: DG1 nouns, which show masculine agreement with the article and optionally feminine agreement forms with the adjective, as in (37), and DG2 nouns, which show different agreement forms depending on the syntactic case, as in (38).

(37) Irish

- a. (Wigger, 2000 7-02-13)
- |                 |                  |
|-----------------|------------------|
| an <sup>T</sup> | t-eolas          |
| ART.NOM.M.SG    | knowledge.NOM.SG |
| ‘the knowledge’ |                  |
- b. (Wigger, 2000 7-01-12)
- |                  |                |
|------------------|----------------|
| eolas            | <u>m</u> haith |
| knowledge.NOM.SG | good.NOM.F.SG  |
- c. (Wigger, 2000 7-02-07)
- |                    |               |
|--------------------|---------------|
| eolas              | <u>m</u> aith |
| knowledge.NOM.SG   | good.NOM.M.SG |
| ‘a good knowledge’ |               |

(38) Gaoth Dobhair Irish (Ó Siadhail, 1984: 175)

- a. an<sup>T</sup>                      t-am
- ART.NOM.M.SG      time.NOM.SG
- ‘the time’
- b. i rith an                      ama
- in flow ART.GEN.M.SG      time.GEN.SG

c. i rith na<sup>H</sup>                      hama  
 in flow ART.GEN.F.SG time.GEN.SG  
 ‘all the time’

This kind of variation is both lexical and dialectal. DG agreement may occur with different nouns in different dialects: e.g. *leoraí* is DG1 in Gaoth Dobhair, *aistir* ‘journey’, *méid* ‘amount’ and *eolas* ‘knowledge’ are DG1 in Cois Fhairrge, and *radharc* ‘sight’ is DG1 in Kerry (Ó Siadhail, 1984: 174; 1989: 146).

Then there are nouns which are not DG in any one dialect but have different genders in different dialects (Ó Siadhail, 1984), e.g. *coláiste* ‘college’, which is feminine in Gaoth Dobhair Irish and masculine elsewhere, and *mí* ‘month’, which is masculine in Munster but feminine in Connacht and Donegal.<sup>26</sup>

## 5.5 Scottish Gaelic

### 5.5.1 Realization and scope of agreement

The gender systems of Scottish Gaelic (ScG) and Irish are very similar. In ScG, as in Irish, gender agreement is only marked in the singular, and, as in Irish, the system has undergone some changes; case marking, in particular, has been dramatically reduced in the varieties of younger generations (MacAulay, 1992: 162f.; also cf. Gillies, 1993: 172).<sup>27</sup>

Table 5.11 shows the paradigm and mutation patterns for agreement with the article and the adjective in traditional varieties, whose system is described here. As we can see, gender agreement is marked only by initial mutation in the nominative; in the genitive it is marked both by initial mutation and morphologically (by different forms of the article and the adjective). In the dative, gender agreement is marked only morphologically (adjective inflection) if the noun phrase contains the article, in which case lenition characterizes both genders; in the absence of the article, on the other hand, lenition is used to mark feminine agreement (§5.5.1.4 below).

<sup>26</sup>Cross-dialectal oscillation can sometimes be explained as a result of the differential treatment of ancient neuters in different dialects; this is for instance the case of *ainm* ‘name’ and *dlí* ‘law’, feminine in Kerry and masculine elsewhere (Ó Siadhail, 1984: 175).

<sup>27</sup>Following the decline of case marking, contemporary varieties of ScG have generalized either the nominative or the dative forms (MacAulay, 1992: 162f.).

Table 5.11: Traditional noun-phrase inflection in ScG

|          | Masculine ‘the little dog’ |               |               | Feminine ‘the little hen’ |                |               |
|----------|----------------------------|---------------|---------------|---------------------------|----------------|---------------|
|          | Art.                       | Noun          | Adj.          | Art.                      | Noun           | Adj.          |
| Nom. sg. | an                         | cù            | beag          | a’                        | <u>ch</u> earc | <u>bh</u> eag |
| Gen. sg. | a’                         | <u>ch</u> oin | <u>bh</u> ig  | <u>na</u>                 | circe          | bige          |
| Dat. sg. | a’                         | <u>ch</u> ù   | <u>bh</u> eag | a’                        | <u>ch</u> irc  | <u>bh</u> ig  |

(After MacAulay, 1992: 162)

### 5.5.1.1 Lenition

As in Irish, the initial mutation of *l*-, *r*- and *n*- is not noted in spelling and has ceased to be distinctive in many varieties of ScG. In general, lenition has become less frequent in contemporary varieties (MacAulay, 1992: 241).

Lenition may be blocked between dentals, but this seems to be the case almost only in set phrases.<sup>28</sup> Table 5.12 illustrates the effects of lenition in ScG.

### 5.5.1.2 *h*-prefixing

As in Irish, the prefixing of *h*- to vowel-initial words as a mark of hiatus can be used to mark semantic distinctions, as in the third-person singular possessive (*a h-athair* ‘her father’ vs. *a athair* ‘his father’).

### 5.5.1.3 Article

Table 5.13 shows how gender agreement between the article and the noun is marked. The following should be noted:

1. As in Irish, masculine nouns beginning with a vowel have a *t*-prefixed in the nominative singular masculine: *each* (m.) ‘horse’ : *an t-each* ‘the horse’.
2. The notation  $a(n)^{L+T}$  represents the leniting form of the article, which is normally reduced to /ə/ (e.g. f. *a’ chailleach* ‘the old woman’).  
Dentals *t*, *d*, *l*, *n*, *r* do not undergo lenition after the article (e.g. *an*

<sup>28</sup>Cf. *Clann Dòmhnail* ‘Clan Donald’ (traditional denomination of a family) vs. *clann Dhòmhnail* ‘Donald’s children’ (Gillies, 1993: 167f., 177).

Table 5.12: Lenition in ScG

|                   | Basic consonant |                                  | Mutated consonant |                     |
|-------------------|-----------------|----------------------------------|-------------------|---------------------|
| <b>Plosives</b>   | ⟨c⟩             | /k <sup>h</sup> k <sup>ʰ</sup> / | ⟨ch⟩              | /x x <sup>ʰ</sup> / |
|                   | ⟨g⟩             | /k k <sup>ʰ</sup> /              | ⟨gh⟩              | /ɣ ɣ <sup>ʰ</sup> / |
|                   | ⟨t⟩             | /t <sup>h</sup> t <sup>ʰ</sup> / | ⟨th⟩              | /h h <sup>ʰ</sup> / |
|                   | ⟨d⟩             | /t t <sup>ʰ</sup> /              | ⟨dh⟩              | /ɣ ɣ <sup>ʰ</sup> / |
|                   | ⟨p⟩             | /p <sup>h</sup> p <sup>ʰ</sup> / | ⟨ph⟩              | /f f <sup>ʰ</sup> / |
|                   | ⟨b⟩             | /p p <sup>ʰ</sup> /              | ⟨bh⟩              | /v v <sup>ʰ</sup> / |
| <b>Fricatives</b> | ⟨s⟩             | /s s <sup>ʰ</sup> /              | ⟨sh⟩              | /h h <sup>ʰ</sup> / |
|                   | ⟨f⟩             | /f f <sup>ʰ</sup> /              | ⟨fh⟩              | /∅/                 |
| <b>Nasals</b>     | ⟨n⟩             | /N N <sup>ʰ</sup> /              | ⟨n⟩               | /n/                 |
|                   | ⟨m⟩             | /m m <sup>ʰ</sup> /              | ⟨mh⟩              | /ṽ ṽ <sup>ʰ</sup> / |
| <b>Laterals</b>   | ⟨l⟩             | /L L <sup>ʰ</sup> /              | ⟨l⟩               | /L l <sup>ʰ</sup> / |
| <b>Vibrants</b>   | ⟨r⟩             | /R R <sup>ʰ</sup> /              | ⟨r⟩               | /r/                 |

(Source: MacAulay, 1992: 241)

*tàir* ‘the reproach’, Calder, 1990: 103). Lenitable *s-* is treated as in Irish (cf. §5.4.1.4), e.g. *an t-súil* /ənˈtu:lʰ/ (nom. f. sg.) ‘the eye’ (Gillies, 1993: 177).

3. The article does not mark gender agreement in the dative.

#### 5.5.1.4 Adjectives

Adjectives agree when used attributively in post-nominal position, where they are normally found. Adjective inflection traditionally marked gender, number and case, but there has been considerable morphological simplification (Gillies, 1993: 178). Table 5.14 exemplifies the traditional inflection paradigm of adjectives: in the dative singular, lenition of the adjective after a masculine noun is normal if the noun is preceded by the article, omitted otherwise.

Not all adjectives have the same agreement-marking potential: Mark (2003: Appendix 3) distinguishes four different types of ScG adjectives in this regard:

Table 5.13: Article paradigm in ScG: singular forms

|      | Before vowels   |                 | Before consonants   |                     |
|------|-----------------|-----------------|---------------------|---------------------|
|      | Masc.           | Fem.            | Masc.               | Fem.                |
| Nom. | an <sup>T</sup> | an              | an (am)             | a(n) <sup>L+T</sup> |
| Gen. | an              | na <sup>H</sup> | a(n) <sup>L+T</sup> | na                  |
| Dat. | an              | an              | a(n) <sup>L+T</sup> | a(n) <sup>L+T</sup> |

(After Gillies, 1993: 177f.)

Table 5.14: Adjective inflection in ScG (*mór* ‘big’)

|          | Masculine     | Feminine      |
|----------|---------------|---------------|
| Nom. sg. | mór           | <u>mh</u> ór  |
| Gen. sg. | <u>mh</u> óir | móire         |
| Dat. sg. | m(h)ór        | <u>mh</u> óir |
| Voc. sg. | (= gen. sg.)  | (= nom. sg.)  |

(Examples from Mulally, 1994: 12)

1. fully inflecting monosyllabic adjectives endings in a broad consonant, e.g. *mór* ‘big’, *geal* ‘white’. These inflect as shown in Table 5.14, i.e. by suffixation and palatalization.
2. Polysyllabic adjectives ending in a broad consonant. They differ from type-1 adjectives in that they they tend to avoid suffixation.
3. Adjectives ending in a consonant and inflecting by means of suffixation (e.g. *glíc* ‘wise’).
4. Indeclinable adjectives, e.g. *ceàrr* ‘wrong’.

### 5.5.1.5 Attributive nouns

Nouns that are used as modifiers may show initial lenition depending on a number of criteria, including semantics, syntax, and grammatical gender. An indefinite noun following a feminine singular noun is treated like an adjective and lenited accordingly.

(39) ScG (Mark, 2003: s.vv. *glainne*, *cupan*)

a. *glainne*                    *b**hainne*  
 glass(F).NOM.SG milk.GEN.SG  
 ‘a glass of milk’

b. *cupan*                        *bainne*  
 cup(M).NOM.SG milk.GEN.SG  
 ‘a cup of milk’

If the attributive noun is an person’s given name, it is lenited if masculine and left unmutated if feminine, regardless of the gender of the preceding noun, as in (40). If it is a place name, it is lenited always, irrespective of grammatical gender ((41) and (42)).

(40) ScG (Plank, 1994: 5)

a. *cat*                            *S**heumais*  
 cat(M).NOM.SG Seumas(M).GEN.SG  
 ‘Seumas’s cat’

b. *cat*                            *S**ine*  
 cat(M).NOM.SG *S**ine*(F).GEN.SG  
 ‘Sine’s cat’

(41) ScG (Gillies, 1993: 200)

a. *muinntir*                    *B**harraidh*  
 people(F).NOM.SG Barra(F)  
 ‘the people of Barra’

b. *drochaid*                    [*B**haile*                    a’ Chaolais]  
 bridge(F).NOM.SG Ballachullish(M)  
 ‘the bridge of Ballachullish’

(42) ScG (Mulally, 1994: 8)

*Baile*                            [*D**hùn*                    *Èideann*]  
 town(M).NOM.SG Edinburgh(M)

‘the town of Edinburgh’

### 5.5.1.6 Possessives

Possessive clitics are used, at least traditionally, to express inalienable possession (typically, body parts and kinship); alternatively, prepositional pronouns are used (Dorian, 1973: 419; MacAulay, 1992: 200). With clitics, as shown in Table 5.15, (anaphoric) gender agreement between the pronoun and its antecedent is marked by the presence or absence of initial mutation on the following word (cf. Calder, 1990: 164, Mark, 2003: 661).

Table 5.15: Possessives in ScG

| Person  | Possessive                     | Examples          | Gloss     |
|---------|--------------------------------|-------------------|-----------|
| 3sg. m. | a <sup>L</sup> /∅ <sup>L</sup> | a <u>m</u> hac    | his son   |
|         |                                | (a) each          | his horse |
|         |                                | (a) <u>f</u> halt | his hair  |
| 3sg. f. | a <sup>H</sup>                 | a mac             | her son   |
|         |                                | a <u>h</u> -each  | her horse |
|         |                                | a falt            | her hair  |

(Data from Mark, 2003: Appendix 3, pp. 660f.)

### 5.5.1.7 Personal pronouns

Personal pronouns show gender agreement only in the third person singular (e.g. m. *e* vs. f. *i*).

## 5.5.2 Semantic assignment

With nouns denoting humans, grammatical gender correlates with the referent's sex (Gillies, 1993: 172). The few exceptions, like m. *boirionnach* 'woman', take syntactic agreement within the noun phrase and semantic agreement anaphorically (mixed agreement).

(43) ScG (Calder, 1990: 74)

Is deas am boirionnach i.  
 COP nice ART.NOM.M.SG woman(M).NOM.SG 3SG.F

'She is a handsome woman.'

Nouns denoting animals often denote individuals of either sex (e.g. m. *cat* ‘cat’). Others show a mismatch between grammatical gender and sex (e.g. m. *capull* ‘mare’, *mart* ‘cow’), and may take mixed agreement (Calder, 1990: 74; Holmer, 1957: 102; 1962: 66).

Nouns for inanimates may be feminine or masculine and take corresponding agreement forms within the noun phrase; anaphorically, they tend to be referred to by masculine pronouns irrespective of their grammatical gender in a number of dialects, or there is oscillation (cf. Dorian, 1978a: 81–84; 1981: 124–29); the speaker of (44), for instance, uses feminine agreement forms with *jugag* ‘jug’ within the noun phrase (*bheag /veg/*), but hesitates over the form of the anaphoric pronoun, which suggests that she is aware of an alternative and perhaps preferable agreement form.

(44) East Sutherland Gaelic (Dorian, 1981: 125)<sup>29</sup>

huč<sup>h</sup> ə jugag veg ə vã:n s vriš a nə  
 fall.PST ART.SG jug(F).SG little.F.SG [down?] and break.PST 3SG.M or  
 vriš i nə vriš a  
 break.PST 3SG.F or break.PST 3SG.M

‘The little jug fell down and it(M) broke. Or it(F) broke. Or it(M) broke’

Secondary semantic regularities, whose predictive power appears to be fairly limited in scope, are found in the traditional gender assignment of inanimate nouns, as shown in (45) and (46).

(45) Semantic criteria for masculine inanimates (Calder, 1990: 76)

- Natural elements: *teine* ‘fire’;
- Seasons of the year: *earrach* ‘spring’;
- Days of the week: *Di-luain* ‘Monday’;
- Metals: *iarunn* ‘iron’;
- Colours: *corcur* ‘purple’;
- Grains: *cruinneachd* ‘wheat’;
- Vegetables: *càl* ‘kail’;
- Alcoholic drinks: *leann* ‘ale’;

<sup>29</sup>Dorian’s (1981) phonological transcription is reproduced here.

- Timber: *giubhas* ‘fir’.

(46) Semantic criteria for feminine inanimates (Calder, 1990: *ibid.*)

- Countries: *Alba* ‘Scotland’;
- Musical instruments: *piòb* ‘bagpipe’;
- Heavenly bodies: *grian* ‘sun’;
- Diseases: *breac* ‘smallpox’;
- Copses: *giùsach* ‘fir-copse’.

Counterexamples are not difficult to find: e.g. f. *beòir* ‘beer’, f. *leiteas* ‘lettuce’, m. *giotàr* ‘guitar’ (Calder, 1990: 76).

### 5.5.3 Formal assignment

As in Irish, the gender of a noun may be predicted to a certain extent based on inflectional and derivational morphology,<sup>30</sup> and there is a tendency for palatalized endings to be associated with the feminine gender; non-palatalized endings, on the other hand, can be of either gender. In addition, nouns ending in /ə/ tend to be masculine if preceded by a non-palatalized consonant (e.g. *balla* /baLə/ ‘wall’), feminine if preceded by a palatalized consonant (e.g. *doille* /døiL’ə/ ‘blindness’).

## 5.6 Manx

Manx had no native speakers left by 1980; in this sense it is regarded here as extinct. It has since been revived and is now used by a number of people for social and literary purposes (Thomson, 1992: 101f.), but the revived variety is ignored here.

Thomson (1992) and Broderick (1993) distinguish between Late Manx (the language as was spoken by the last native speakers), and Classical Manx (exemplified by the 18th-century translation of the Bible). Thomson describes Late Manx as a variety in which interference from English is pervasive, involving lexicon, syntax and morphology (Thomson, 1992: 101f.).

<sup>30</sup>For a detailed account based on the more conservative dialects of the modern spoken language see Gillies (1993: 173–75) and Calder (1990: 76).

### 5.6.1 Realization and scope of agreement

The initial mutation system, which in Classical Manx (as in Irish and ScG) was an important exponent of gender agreement, had almost completely disappeared in Late Manx (Thomson, 1992: 118, 132; Broderick, 1993: 239).

The mutational distinctions [...] which sustain gender have long been falling into disorder or disuse, and the gender of inanimates is poorly reflected in the third person singular pronouns; *effectively some nouns are feminine, the marked class, and the remainder, the unmarked majority, are masculine.* [Thomson, 1992: 118, emphasis mine.]

Broderick (1993: 239) adds that gender agreement is now only marked anaphorically and observes that reference to inanimates is expressed “almost exclusively” by masculine pronouns.

*[N]ouns can essentially be regarded as masculine unless there is evidence to suggest they are not.* Any gender distinction appears in the third-person singular personal pronoun, but even here discrepancies are frequent, as the notion of ‘it’ is almost exclusively expressed by *eh* /e/ ‘he, him’. [Broderick, 1993: 239, emphasis mine.]

#### 5.6.1.1 Lenition

Lenition in Classical Manx (Table 5.16) was substantially similar to that of Irish and ScG; it was generally blocked between two dentals and was not very frequent with initial *f*- (Thomson, 1992: 132f.).

#### 5.6.1.2 *h*-prefixing

In Classical Manx, *h*- was prefixed to vowel-initial nouns after the article in the genitive feminine singular, as in (47); it was also used to mark anaphoric agreement between a possessive clitic and its feminine antecedent, as in (48).

(47) Manx (Thomson, 1992: 133)

ny<sup>H</sup>                      hooirey  
ART.GEN.F.SG earth  
‘of the earth’

Table 5.16: Lenition in Manx

|                 | Basic consonant   |         | Mutated consonant |         |
|-----------------|-------------------|---------|-------------------|---------|
| <b>Plosives</b> | ⟨c⟩               | /k k' / | ⟨ch, chi⟩         | /x x' / |
|                 | ⟨g⟩               | /g g' / | ⟨gh, y/ghi⟩       | /ɣ ɣ' / |
|                 | ⟨t⟩               | /t t' / | ⟨h, hi⟩           | /h h' / |
|                 | ⟨d⟩               | /d d' / | ⟨gh, y⟩           | /ɣ ɣ' / |
|                 | ⟨p⟩               | /p /    | ⟨ph⟩              | /f /    |
|                 | ⟨b⟩               | /b /    | ⟨v, w⟩            | /v, w / |
|                 | <b>Fricatives</b> | ⟨s⟩     | /s s' /           | ⟨h, hi⟩ |
| ⟨f⟩             |                   | /f /    | ∅                 | ∅       |
| <b>Nasals</b>   | ⟨m⟩               | /m /    | ⟨v, w⟩            | /v, w / |

(Thomson, 1992: 132)

(48) Manx (Thomson, 1992: *ibid.*)

e<sup>H</sup>                    haigney  
 POSS:3SG.F mind

‘her mind’

### 5.6.1.3 Article

There are two basic article morphemes, *yn* /ən/ (or /in/) and *ny* /nə/. In Classical Manx the pattern of initial mutations after the article was as in Irish and ScG: lenition for feminine nouns vs. no lenition for masculines in the nominative singular (49), *h*-prefixing for feminine nouns vs. lenition for masculines in the genitive singular (50).

(49) Manx (Broderick, 1993: 237, 246)

a. yn                                    fer  
 ART.NOM.M.SG man(M).NOM.SG  
 ‘the man’

b. yn                                    yen                                    (cf. *ben*)  
 ART.NOM.F.SG woman(F).NOM.SG  
 ‘the woman’



In Late Manx, however, this distinction was lost as  $e^L$  started to be used irrespective of the antecedent's grammatical gender (Broderick, 1984: 10).

#### 5.6.1.6 Personal pronouns

Personal pronouns could mark gender agreement in the third person singular: masculine *eh* (emphatic *eshyn*)<sup>32</sup> vs. feminine *ee* (emphatic *ish*).<sup>33</sup> As already mentioned above, in Late Manx anaphoric reference to inanimates was made almost exclusively using masculine pronouns, meaning that feminine pronouns were only used for reference to female animates. However, according to Phillips (2004) a third form was available which allowed speakers to make a finer distinction: the form *yh* /ə/, a less emphatic realization of the masculine pronoun *eh*; *yh* could be used generally for non-female reference, while *eh* would be used optionally and specifically for reference to males (Phillips, 2004: 18).

#### 5.6.2 Gender assignment

While gender assignment in Late Manx was essentially pronominal and semantic, in earlier stages of the language inanimates could be either masculine or feminine and their gender could be predicted, at least to some extent, based on certain formal and semantic features of the noun. In particular, certain derivational suffixes and palatalized final consonants were associated with the feminine in more or less the same way as in Irish and ScG (Kneen, 1973: 49f.).

### 5.7 Welsh

There is in Welsh a diglossic divide between a high variety (literary Welsh, henceforth LW) and a low variety (colloquial Welsh, CW). Each has its own domains of application, in accordance with Ferguson's (1959) canonical definition of diglossia (Ball & Müller, 1992: 8–10). The situation is complicated further by the fact that there are distinct regional varieties of CW (Ball, 1988).

<sup>32</sup>Variously realized as /e/, /a/, /ɑ/ or /i/; emphatic /eʒən/, /efən/ (Broderick, 1984: 53).

<sup>33</sup>Variously realized as /i/, /ei/, /ai/; emphatic /if/ (Broderick, 1984: 53).

LW and CW are distinct at all levels: phonological, morphological, and syntactic (Coupland & Ball, 1989; Fife, 1986; Jones, 1988; Borsley et al., 2007: 5–7), but there is no major structural difference between LW and CW, or between different dialects of CW, as far as grammatical gender is concerned.<sup>34</sup>

In numerical terms, it has been estimated (SurrIDGE, 1989) that the masculine gender is the larger: feminine nouns only represent 27.5% of all nouns (and 32.2% of all inanimates).

### 5.7.1 Realization and scope of agreement

#### 5.7.1.1 Soft mutation (SM)

From a structural and functional point of view, SM (Table 5.17) corresponds to lenition in Goidelic and is employed by both LW and CW in the same environments (Thomas, 2001: §2.7.1). It affects voiceless and voiced plosives, the voiceless approximants /ɫ/ and /r<sup>h</sup>/<sup>35</sup> and /m/. Under SM, voiceless consonants are replaced by their voiced counterparts, voiced consonants by homorganic fricatives and /g/ simply deleted (Ball & Müller, 1992: 12f.).

#### 5.7.1.2 Aspirate mutation (AM)

AM only affects the three phonemes /p, t, k/, which are replaced by the homorganic fricatives /f, θ, x/, respectively. Its position is particularly unstable in CW, where it is now mostly avoided (particularly in southern dialects); however, where its presence marks gender agreement (i.e. with the feminine possessive *ei*),<sup>36</sup> its use appears to be still fairly regular (Awbery, 1986; Ball, 1993; Ball & Müller, 1992: 252f.; Thomas, 2001: §2.8.1; Watkins, 1993: 307).

<sup>34</sup>Minor differences concern the assignment of individual nouns and the realization of agreement and will be considered in §5.7.4.

<sup>35</sup>/r<sup>h</sup>/ is the notation adopted by Ball & Müller (1992); other authors use /r̥/ (e.g. Jones, 1998a). Ball & Müller's notation is adopted here because of the structural facts related to the strong aspiration of this phoneme: the dialects of the south-east, which typically lack /h/, also lack /r<sup>h</sup>/ (see §5.7.4.1 below).

<sup>36</sup>See §5.7.1.9 below.

Table 5.17: SM in Welsh

|                     | Basic consonant |                   | Mutated consonant |     |
|---------------------|-----------------|-------------------|-------------------|-----|
| <b>Plosives</b>     | ⟨c⟩             | /k/               | ⟨g⟩               | /g/ |
|                     | ⟨g⟩             | /g/               | ∅                 | ∅   |
|                     | ⟨t⟩             | /t/               | ⟨d⟩               | /d/ |
|                     | ⟨d⟩             | /d/               | ⟨dd⟩              | /ð/ |
|                     | ⟨p⟩             | /p/               | ⟨b⟩               | /b/ |
|                     | ⟨b⟩             | /b/               | ⟨f⟩               | /v/ |
| <b>Approximants</b> | ⟨ll⟩            | /ʎ/               | ⟨l⟩               | /l/ |
|                     | ⟨rh⟩            | /r <sup>h</sup> / | ⟨r⟩               | /r/ |
| <b>Nasals</b>       | ⟨m⟩             | /m/               | ⟨f⟩               | /v/ |

### 5.7.1.3 *h*-prefixing (PVA)

As far as gender agreement is concerned, *h*-prefixing or PVA (“pre-vocalic aspiration”) is triggered by the feminine possessive clitic (§5.7.1.9).

### 5.7.1.4 Article

The article has three singular forms (*yr* /əɾ/, *y* /ə/ and *'r* /r/). Their distribution depends on morphosyntactic and phonological factors which are not relevant at present and will be discussed later in §6.3. Feminine nouns following the article undergo SM, except those beginning in /ʎ/ and /r<sup>h</sup>/; Ball & Müller (1992) use the abbreviation SMR to represent this reduced pattern of mutation.

(52) Welsh (Thomas, 2001: §2.11.1)

- a. *y*                    *ci*  
ART.M.SG dog(M).SG  
‘the dog’
- b. *y*<sup>SMR</sup>            *gath*            (cf. *cath*)  
ART.F.SG cat(F).SG  
‘the cat’
- c. *y*<sup>SMR</sup>            *llygoden*  
ART.F.SG mouse(F).SG

SM of /g/ is consistently inhibited in the case of loanwords and non-Welsh place-names, e.g. f. *gêm* ‘game’ : *y gêm* ‘the game’ (< English *game*).

### 5.7.1.5 Adjectives

Attributive adjectives mark gender agreement by undergoing SM when modifying a feminine noun; in LW, a number of adjectives can also mark gender agreement by internal vowel shift. In either case, gender agreement is only marked in the singular (Watkins, 1993: 312).

(53) Welsh (Thomas, 2001: §2.9.1)

a. ci            mawr    du  
 dog(M).SG big.M.SG black.M.SG  
 ‘a big black dog’

b. cath        fawr    ddu  
 cat(F).SG big.F.SG black.F.SG  
 ‘a big black cat’

As can be seen from example (53), SM following a feminine noun is found on all mutable adjectives in the phrase.<sup>37</sup>

Adjective phrases mutate as wholes: if an adverbial modifies and precedes an adjective, it is the adverbial, not the adjective, that undergoes SM (Borsley et al., 2007: 177), as shown in (54).<sup>38</sup>

(54) Welsh (Borsley et al., 2007: 177)

gorchest                    [dra<sup>AM</sup> chymhleth]    (cf. *tra*, *cymhleth*)  
 achievement(F).SG [quite    complex].F.SG  
 ‘quite a complex achievement’

<sup>37</sup>However, every new intervening trigger will take the precedence. In the following example, *a* ‘and’ triggers aspirate mutation.

i. Welsh (Ball & Müller, 1992: 162)

merch    dal        a<sup>AM</sup> chref            (cf. *tal*, *cref*)  
 girl(F).SG tall.F.SG and strong.F.SG  
 ‘a tall and strong girl’

<sup>38</sup>The aspirate mutation of *cymhleth* is triggered by *tra* and is not a gender mutation.

Adjectives modifying a feminine head noun undergo SM even if they are actually adjacent to an intervening constituent which depends on that noun, as shown in (55).

(55) Welsh (Borsley et al., 2007: 178)

[cynffon mochn]                      gyrliog      (cf. *cyrliog*)  
 [tail(F).SG pig(M).SG](F).SG gyrliog.F.SG

‘a curly pig’s tail’

Agreement marking by internal vowel shift is restricted to a limited set of adjectives and is no longer productive (Watkins, 1993: 312). *Gwyn* /gwin/ (f. *gwen* /gwen/ ‘white’) is an example.

(56) Welsh (Thomas, 1992b: 298)

|                                 |                            |
|---------------------------------|----------------------------|
| a. ci                      gwyn | b. cath <sup>SM</sup> _wen |
| dog(M).SG white.M.SG            | cat(F).SG white.F.SG       |
| ‘a white dog’                   | ‘a white cat’              |

However, it has been observed that even the use of SM as a mark of adjective agreement is often avoided in CW, while its distinctive function seems to be blurred by its use being now extended after masculine head nouns (Thomas, 2001: §2.11; also cf. Jones, 1998a).<sup>39</sup> There is also evidence of inconsistently marked noun phrases, with SM applying to the noun but not the adjective (57a), or vice versa (57b).

(57) Welsh (Thomas, 2001: §2.11.3)

a. *y gath du* ‘the black cat’ (expected: *y gath ddu*)  
 b. *y pêl goch* ‘the red ball’ (expected: *y bêl goch*)

Variation in the use of mutation appears to be found “both between and within speakers” (Thomas & Gathercole, 2005: 2238).

<sup>39</sup>Furthermore, when the adjective is used in traditional, fossilized constructions (like epithets), it is normal to find SM after masculine nouns, as in *Arthur Fawr* ‘Arthur the Great’ (cf. *mawr* ‘great’). Here, SM would appear is typical of the formulaic expression and is found with nominal appositions as well: e.g. *Arthur Frenin* ‘Arthur the King’ (cf. *brenin* ‘king’) (Ball & Müller, 1992: 148; Thorne, 1993: 25).

### 5.7.1.6 Attributive nouns

Nouns which are used attributively may undergo SM if the head noun is feminine and singular:

(58) Welsh (Thomas, 1992b: 298)

a. wal        frics (cf. *brics*)  
 wall(F).SG bricks  
 ‘a brick wall’

b. ffon        daflu (cf. *taflu*)  
 stick(F).SG throw.VN  
 ‘a throwing stick’

However, SM in this context is now rather uncommon even in LW (Ball, 1993: 191).

It has already been observed apropos of the argument–modifier distinction in Irish (§5.4.1.6 above) that dependent nouns tend not to mutate if they express an argument of the head noun. In Welsh, the distinction is traditionally drawn between broadly-defined “attributive” noun phrases and those which denote possession (Borsley et al., 2007: 184f.; Williams, 1980: 38): gender mutations typically do not apply to the latter. Cf. example (59), where there is no mutation on *mab* despite *siop* being feminine, and (60), where feminine *siop* triggers soft mutation on attributive *llyfrau*.

(59) Welsh (Borsley et al., 2007: 184)

siop        [mab chwaer y    meddyg]  
 shop(F).SG [son    sister    ART doctor]  
 ‘the shop of the doctor’s sister’s son’

(60) Welsh (Borsley et al., 2007: 185)

siop    lyfrau (cf. *llyfrau*)  
 shop(F) book.PL  
 ‘a book shop’

While, in CW at least, SM of personal names is generally avoided (Thorne, 1993: 74), the case of proper names in general is particularly unpredictable, as they might or might not mutate, even when they express

the possessor, and being proper names they tend to be repeated in the established form: after feminine *ynys* ‘island’, for instance, both the form *Ynysforgan* (*ynys* + *Morgan*), with SM, and the form *Ynys Môn*, without SM, are attested (Williams, 1980: 38).

One example found in CC-W is (61).

(61) (jonsi.cha 2258)

yn eisteddfod Môn ym Mrynsiencyn  
in eisteddfod(F) Anglesey in Brynsiencyn

‘at the Anglesey *eisteddfod* in Brynsiencyn’

Here, *Môn* ‘Anglesey’ does not mutate after feminine *eisteddfod*.<sup>40</sup> Elsewhere in the same text, however, *Môn* is found in its mutated form after the feminine noun *sir* ‘county’ (62).

(62) Welsh (jonsi.cha 1561)

mae o t(u)a Sir Fôn (y)na yng Nghaergybi  
be.PRS.3 3SG.M towards county(F) Anglesey there in Holyhead

‘he is over there in Anglesey, in Holyhead’

There are therefore two ways (one language-specific, the other language-independent) in which the presence of proper names can affect gender agreement:

1. if the attributive noun is a proper name, it may display idiosyncratic mutation patterns, as just explained (language-specific);
2. if the noun phrase as a whole is a proper name, and contains a gender-marking initial mutation, this mutation may be “protected” against change by being part of a proper name (our language-independent hypothesis).

Both will be considered in our analysis.

### 5.7.1.7 Demonstratives

In LW, demonstrative adjectives and pronouns have gendered forms in the singular (Table 5.18 and example (63)).

<sup>40</sup>An *eisteddfod* is a traditional Welsh festival.

Table 5.18: Inflection of demonstratives in LW

|                  | ‘this’ | ‘that’ |
|------------------|--------|--------|
| <i>Masculine</i> | hwn    | hwmmw  |
| <i>Feminine</i>  | hon    | honno  |

(Thomas, 1992b: 292)

(63) Welsh (Thomas, 1992b: 292)

- a. y            dyn            hwn            /    hwmmw  
 ART.M.SG man(M).SG this.M.SG / that.M.SG  
 ‘this / that man’
- b. y<sup>SMR</sup>        ferch        hon            /    honno  
 ART.F.SG girl(F).SG this.F.SG / that.F.SG  
 ‘this / that girl’

The same forms can be used pronominally.

(64) Welsh (Thorne, 1993: 191)

Yf    hwn  
 drink this.M.SG  
 ‘Drink this’

There are two neuter (non-gendered) forms, *hyn* ‘this’ and *hynny* ‘that’, used in LW as discourse anaphors (65).

(65) Welsh (Thomas, 1992b: 294)

[Oedd gwerthu’r car yn beth ffôl i’w wneud.]<sub>i</sub> Mae hynny<sub>i</sub>’n amlwg  
 rŵan.  
 ‘[Selling the car was a foolish thing to do.]<sub>i</sub> That<sub>i</sub> is obvious now.’

However, in at least some varieties of CW, they may replace masculine or feminine demonstrative in attributive use: *y dyn hyn* ‘this man’ (instead of *y dyn hwn*), *y wraig hyn* ‘this woman’ (instead of *y wraig honno*) (Thorne, 1993: 194). However, both in CW and in less formal writing styles, demonstrative adjectives tend to be used rarely, replaced by the deictic adverbs *yma* ‘here’ and *yna* ‘there’, which are not inflected.

(66) Welsh (Thomas, 1992b: 292)

- a. y            dyn            yma / yna  
 ART.M.SG man(M).SG here / there  
 ‘this / that man’
- b. y<sup>SMR</sup>        ferch        yma / yna  
 ART.F.SG girl(F).SG here / there  
 ‘this / that girl’

### 5.7.1.8 Numerals

The numeral *un* ‘one’ behaves like the article, triggering SMR on a following feminine noun: *un<sup>SMR</sup> ferch* ‘one girl’ (f. *merch*), *un<sup>SMR</sup> llaw* ‘one hand’ (f. *llaw*), *un<sup>SMR</sup> rhwyd* ‘one net’ (f. *rhwyd*) (Thorne, 1993: 30). As a pronoun replacing a feminine noun, it triggers full SM on a following adjective.<sup>41</sup>

(67) Welsh (Thorne, 1993: 30)

- a. mae    hon        yn    un<sup>SM</sup>    ddiddorol  
 be.PRS this.F.SG PTC one.F.SG interesting.F.SG  
 ‘she is an interesting one’
- b. un<sup>SM</sup>    rŷfedd        yw    hi  
 one.F.SG strange.F.SG be.PRS 3SG.F  
 ‘she is a strange one’

The numerals for ‘2’, ‘3’ and ‘4’ mark gender agreement as shown in Table 5.19 by means of internal vowel shift.

(68) Welsh (Thorne, 1993: 31)

- a. dau    frawd  
 two.M brother(M).SG  
 ‘two brothers’

---

<sup>41</sup>At least in literary usage, however, *un* mutates both masculine and feminine nouns when its meaning is ‘similar’.

i. Welsh (Thorne, 1993: 30)

- un<sup>SM</sup> ben        â'i            dad    [cf. *pen*]  
 one    head(M).SG as=POSS:3SG.M father

‘(he has) a head like his father’s, he is as clever as his father’

Table 5.19: Gendered numerals in Welsh

|   | Masculine | Feminine |
|---|-----------|----------|
| 2 | dau       | dwy      |
| 3 | tri       | tair     |
| 4 | pedwar    | pedair   |

(Thomas, 1992b: 294)

- b. *dwy ferch*  
 two.F girl(F).SG  
 ‘two girls’

A tendency to avoid the gender-marked forms of these numerals has been observed in younger speakers’ varieties (Jones, 1998a: 68, 173f.).

### 5.7.1.9 Possessives

Gender agreement between a third-person possessive clitic and its antecedent is marked by the different types of initial mutation that appear on the following word, as shown in Table 5.20 and in examples (69–72). The allomorph *’w* is selected after the preposition *i* ‘to’ (*i’w*).<sup>42</sup>

Table 5.20: Possessives in Welsh

|            | Forms         |                             | Allomorphs |                             |
|------------|---------------|-----------------------------|------------|-----------------------------|
| POSS:3SG.M | <i>ei, ’i</i> | <i>/i<sup>SM</sup>/</i>     | <i>’w</i>  | <i>/u<sup>SM</sup>/</i>     |
| POSS:3SG.F | <i>ei, ’i</i> | <i>/i<sup>AM+PVA</sup>/</i> | <i>’w</i>  | <i>/u<sup>AM+PVA</sup>/</i> |

(69) Welsh (Thomas & Gathercole, 2005: 2236)

- ei<sup>SM</sup> bêt*  
 POSS:3SG.M ball  
 ‘his ball’

<sup>42</sup>In some southern dialects the form *iddi* is found instead of *i’w*.

(70) Welsh (Thorne, 1993: 162)

i'w<sup>SM</sup>            gar  
to=POSS:3SG.M car  
  
'to his car'

(71) Welsh (Thomas & Gathercole, 2005: *ibid.*)

ei<sup>AM</sup>            phêl  
POSS:3SG.F ball  
  
'her ball'

(72) Welsh (Thorne, 1993: *ibid.*)

i'w<sup>AM</sup>            bwthyn  
to=POSS:3SG.F cottage  
  
'to her cottage'

The gender of the antecedent in the possessive construction may additionally be marked by means of a suffixed clitic, as in example (73).

(73) Welsh (Thomas, 1992b: 336)

ei<sup>PVA</sup>            hafal hi  
POSS:3SG.F apple 3SG.F  
  
'her apple'

It has been observed that less conservative varieties of CW may recast the opposition in terms of mutation (SM) versus no mutation at all (cf. Awbery, 1986), or apply SM after possessive *ei* regardless of the antecedent's grammatical gender, thus neutralizing the distinction (Thomas, 2001: 360).

#### 5.7.1.10 Personal pronouns

LW has two sets of personal pronouns, independent and dependent. The details regarding their distribution and functions are not important here: the reader is referred to Thomas (1992b: 303–306) and Thorne (1993: 154–169) for further information. Gender agreement is consistently marked in the third person singular (Table 5.21).

In LW, anaphoric pronouns commonly agree with the grammatical gender of their antecedent, as in example (74), while in contemporary varieties

Table 5.21: Personal pronouns in LW

| Type        | Subtype             | 3sg. m.      | 3sg. f. |
|-------------|---------------------|--------------|---------|
| Independent | simple              | ef/fe/fo     | hi      |
|             | reduplicated        | (e)fe/efô/fo | hyhi    |
|             | conjunctive         | yntau        | hithau  |
| Dependent   | simple affixed      | ef/fe/fo     | hi      |
|             | conjunctive affixed | yntau        | hithau  |

(Based on Thorne, 1993: 154–169)

of CW it is common to witness the overgeneralization of the masculine form for inanimate antecedents, as in example (75). (Also cf. Jones, 1998a: 66, 171.)

(74) Welsh (LW) (Thomas, 2001: §2.10)

a. Syrthiodd y<sup>SMR</sup>      gadair<sub>i</sub>      ar ei<sup>AM</sup><sub>i</sub>      chefn  
 fall.PST    ART.F.SG chair(F).SG on POSS:3SG.F back  
 ‘The chair fell on its back’

b. Syrthiodd y              bwrdd<sub>i</sub>      ar ei<sup>SM</sup><sub>i</sub>      gefn  
 fall.PST    ART.M.SG table(M).SG on POSS:3SG.M back  
 ‘The table fell on its back’

(75) Welsh (CW) (Thomas, 2001: §2.10)

Lle    mae'r<sup>SMR</sup>              gadair<sub>i</sub>?    Dyma fo<sub>i</sub>.  
 where be.PRS=ART.F.SG chair(F)? here    3SG.M

‘Where is the chair? Here it is.’

In other words,

gender distinctions as marked in distant constructs are dying when in reference to inanimate nouns and slowly moving towards a more semantic-based system. [Thomas, 2001: 351]

Feminine pronouns are also traditionally used to anticipate an extra-posed clause and for statements about the weather (Thomas, 1992b: 303).

(76) Welsh (Thomas, 1992b: 303)

- a. Mae hi'n amlwg [bod Mair wedi anghofio]<sub>i</sub>.  
 be.PRS 3SG.F=PTC obvious [be.VN Mair after forget.VN]  
 'It<sub>i</sub> is obvious [that Mair has forgotten]<sub>i</sub>.'
- b. Mae hi'n bwrw.  
 be.PRS 3SG.F=PTC rain.VN  
 'It is raining.'

### 5.7.2 Semantic assignment

Nouns specifically denoting male and female human referents are normally masculine and feminine, respectively (m. *brawd* 'brother', f. *merch* 'girl' etc.). In the case of non-human animates both masculine and feminine nouns may refer to individuals of either sex (epicene nouns).

Double-gender nouns are infrequent, the only widely-attested one being the English loanword *ffrind* 'friend' (Watkins, 1993: 310).

(77) Welsh (Watkins, 1993: 310)

- a. Ffrind da yw e.  
 friend(M).SG good.M.SG be.PRS 3SG.M  
 'He is a good friend'
- b. Ffrind dda yw hi.  
 friend(F).SG good.F.SG be.PRS 3SG.F  
 'She is a good friend.'

Other regular or quasi-regular correspondences between semantics and grammatical gender are commonly found in descriptions of the language. According to Surrige (1989), Thorne (1993: 120–22) and Thomas (2001: §2.9.2), nouns for (a) seasons, months and days of the week, (b) points of the compass, (c) substances and materials, as well as (d) verbal nouns tend to be masculine; while nouns for (e) countries and regions, (f) academic disciplines, (g) languages and dialects, (h) rivers and streams, (i) mountains, (j) trees, and (k) collective nouns tend to be feminine. Exceptions are commonly found in most subgroups.

### 5.7.3 Formal assignment

A number of derivational suffixes can be associated quite reliably with either grammatical gender (Surrige, 1989: 207). Masculine *-yn* and feminine *-en*, in particular, are used to form singulative nouns out of collective or inherently plural stems.

(78) Welsh

a. m. *coed* ‘wood’ + *-en* → f. *coeden* ‘tree’

b. pl. *plant* ‘children’ + *-yn* → m. *plentyyn* ‘child’

Other suffixes are statistically associated with a particular gender, as shown in Tables 5.22 and 5.23.

Table 5.22: Formal gender assignment in Welsh (masculine)

| Type                         | Ending    | Example       | Gloss         |
|------------------------------|-----------|---------------|---------------|
| Abstracts from adjectives    | -aint     | hen-aint      | old age       |
|                              | -deb      | union-deb     | rightness     |
|                              | -der/-ter | poeth-der     | heat          |
|                              | -did/-tid | glen-did      | cleanliness   |
|                              | -dra/-tra | cyfleus-tra   | convenience   |
|                              | -had/-âd  | eglur-had     | clarification |
|                              | -i        | tlod-i        | poverty       |
|                              | -id       | rhydd-id      | freedom       |
|                              | -ineb     | ffol-ineb     | folly         |
|                              | -ioni     | hael-ioni     | generosity    |
|                              | -awd/-od  | un-awd        | solo          |
|                              | -rwydd    | caredig-rwydd | kindness      |
|                              | -wch      | tywyll-wch    | darkness      |
| -yd                          | segur-yd  | idleness      |               |
| Abstracts from nominal bases | -ad       | enw-ad        | denomination  |
|                              | -iad      | car-iad       | love          |
|                              | -iant     | ffyn-iant     | prosperity    |

(After Thorne, 1993: 123f.)

For monosyllables there is a correlation between the vowel and grammatical gender (Watkins, 1993: 309): *w* /u/ and *y* /i/ (e.g. *rhwd* ‘rust’, *hyd*

Table 5.23: Formal gender assignment in Welsh (feminine)

| Type                         | Ending | Example     | Gloss             |
|------------------------------|--------|-------------|-------------------|
| Abstracts from nominal bases | -aeth  | gwybod-aeth | knowledge         |
|                              | -as    | prïod-as    | wedding           |
| Concretes from nominal bases | -ell   | llin-ell    | line              |
|                              | -fa    | nodd-fa     | (place of) refuge |

(After Thorne, 1993: 123f.)

‘length’) are associated with the masculine gender, *o* /ɔ/ and *e* /ɛ/ (e.g. *ton* ‘wave’, *ffêr* ‘ankle’) with the feminine.<sup>43</sup>

## 5.7.4 Variation in agreement marking

### 5.7.4.1 Geographical variation

The dialects of south-eastern Wales typically lack any phonemic opposition involving aspiration (/h/) (Jones, 1998a: 361; Watkins, 1993: 301), including that between /r<sup>h</sup>/ and /r/, which eliminates one possible environment for gender agreement through SM; the opposition between possessive *ei* (m.) and *ei*<sup>PVA</sup> (f.) before a vowel is also neutralized.

The opposite situation is found in the dialects of northern Wales, where aspiration has phonemic status and the domain of application of PVA includes non-syllabic /j/ and /w/ and to the sonorants /l, m, n, r/, where it is realized as devoicing (Thomas, 1992b: 336; Thorne, 1993: 76), as shown in (79).<sup>44</sup> This extension of PVA is not found in all environments where PVA is found, but only with the possessive clitics *ei* (3sg. f.) and *eu* (3pl.) (cf. Ball & Müller, 1992: 26f., Thorne, 1993: 76, Jones, 1998a: 367).

<sup>43</sup>Exceptions are not infrequent: cf. feminine *llwy* ‘spoon’, *hwch* ‘sow’, *dysgl* ‘dish’ and masculine *corff* ‘body’, *pres* ‘money’ (Thomas, 2001: 2.9.2). Earlier monosyllabic loanwords comply with this general tendency: e.g. English *shawl* > Welsh *siôl* (f.), English *mud* > Welsh *mwd* (m.); more recent loanwords tend to be assigned to the masculine gender by default (Watkins, 1993: 310).

<sup>44</sup>The association between PVA and the devoicing of these glides and sonorants is suggested by Gwenllian Awbery (cited in Ball & Müller, 1992: 25).

(79) Welsh (from Jones, 1998a: 367)

|         |          |   |               |                    |
|---------|----------|---|---------------|--------------------|
| mam     | ‘mother’ | : | ei /m̥/am     | ‘her/their mother’ |
| nodwydd | ‘needle’ | : | ei /n̥/odwydd | ‘her/their needle’ |
| lamp    | ‘lamp’   | : | ei /l̥/amp    | ‘her/their lamp’   |
| ruban   | ‘ribbon’ | : | ei /r̥/uban   | ‘her/their ribbon’ |
| wall    | ‘wall’   | : | ei /w̥/all    | ‘her/their wall’   |
| /j/âr   | ‘hen’    | : | ei /j̥/âr     | ‘her/their jar’    |

As regards gender agreement with the adjective, *bach* ‘small’ is unique in systematically resisting mutation in a number of northern dialects (Thomas & Thomas, 1989: 100, Williams, 1980: 37).

(80) Welsh (Thomas, 2001: §2.11.2 and cf. *jonsi.cha* 985)

- a. hogyn      bach      da  
 boy(M).SG small.SG good.M.SG  
 ‘a good little boy’
- b. hogan bach      dda  
 girl(F) small.SG good.F.SG  
 ‘a good little girl’

#### 5.7.4.2 Age-related variation

A number of studies on diachronic variation in gender agreement in Welsh have been carried out using an apparent-time approach. Among these are Jones (1998a) and Thomas & Gathercole (2005), both treating gender agreement variation as an indicator of language obsolescence.

Jones (1998a) is a naturalistic study based on data collected in the course of a series of interviews with the informants rather than elicited in an experimental setting. Jones interviewed both adults and children, but for the purpose of comparison with the present study I will focus on the the adult data only. The adults were divided into four age groups: 20–39 (A), 40–59 (B), 60–74 (C), and over 75 (D). Jones considered two different communities (Rhymney in southern Wales, where Welsh was spoken by just 6.7% of the population, and Rhosllannerchrugog in northern Wales, where it was spoken by 38.1% of the population) and seven agreement contexts:

1. SM after masculine *ei* ([+male] antecedent);

2. AM after feminine *ei* ([+female] antecedent);
3. pronominal agreement (non-possessive pronouns);<sup>45</sup>
4. SM of attributive adjectives after the article;
5. SMR of feminine nouns after the article;
6. SMR of feminine nouns after the numeral *un* ‘one’;
7. agreement of the numerals “2”, “3” and “4”.<sup>46</sup>

She found that the expected mutation after the possessive pronouns (with animate reference) was universally found in all age groups in both Rhymney and Rhosllannerchrugog (the only exception being the A group in the latter, where SM after masculine *ei* was found only 78% of the time).

Non-possessive pronouns were found to always agree grammatically with masculine antecedents in both communities and in all age groups; with feminine antecedents, on the other hand, grammatical agreement was never produced by Rhymney speakers in age groups A and B, but it was observed with the other two age groups (30% of the time in C and 71% of the time in D). In Rhosllannerchrugog, grammatical agreement in this context was observed 10% of the time in the youngest group (A), and between 50% and 60% of the time in the other three groups (54% in B, 59% in C and 55% in D).<sup>47</sup>

Agreement with feminine nouns within the noun phrase was found to occur at comparable levels in both Rhymney and Rhosllannerchrugog, except with the youngest speakers. In Rhymney, SMR of feminine nouns after the article was observed to occur between 84% and 92% of the time in B, C and D; in the same age groups, SM of the attributive adjective after a feminine

<sup>45</sup>It is not explicitly stated whether the data pertain to all agreement contexts or just to non-semantic agreement, but it makes sense to assume that only non-semantic agreement is considered: not only based on the examples Jones provides when explaining this variable (1998a: 55) and in Appendix IV (1998a: 382–405), but also on her figures (certain groups use only masculine pronouns, which would be unlikely to happen if female antecedents were being counted).

<sup>46</sup>As is clear from the examples in Appendix IV (Jones, 1998a: 382–405), only their use as adjectives was studied.

<sup>47</sup>As observed by Jones (1993, cited in Thomas & Gathercole, 2005: 2245), “[i]t is not uncommon for an inanimate noun to be marked as feminine in one context, but to occur with a masculine pronoun”.

noun was observed between 85% and 91% of the time. Group A mutated feminine nouns after the article only 29% of the time and attributive adjectives after feminine nouns 69% of the time.

In Rhosllannerchrugog, SMR of feminine nouns after the article was produced between 91% and 96% of the time in the three older age groups (B–D) and 62% of the time in the youngest one (A). SM of adjectives after feminine nouns was similarly observed between 92% and 96% of the time in the older groups (B–D) and 67% of the time in A. SMR of feminine nouns after the numeral *un* was also rather consistent in Rhymney (79% mutation rate in group B and 100% in groups C and D, but no opportunities in group A); insufficient data for this variable was available for Rhosllannerchrugog.

Finally, the use of the appropriate gender form for numerals in the “2”–“4” range was considered. Lower counts are available for this category. In Rhymney, the use of feminine numerals with masculine nouns is only recorded in group B (three times out of seven); use of masculine numerals with feminine nouns is recorded three times out of six (50%) in group A, two out two (100%) in group C, and one out of six (17%) in group D. In Rhosllannerchrugog, the use of feminine numerals with masculine nouns is never observed. The use of masculine numerals with feminine nouns is found two times out of nine (22%) in group A and two out of 17 (12%) in group B. In the other two groups the appropriate feminine form is always used.

A follow-up study was undertaken by Thomas & Gathercole (2005), who set out to gather data about gender agreement comparable to Jones’s (1998a) but from Gwynedd and Anglesey, described as “the counties with the highest proportion of Welsh speakers in all of Wales” (Thomas & Gathercole, 2005: 2245),<sup>48</sup> to specifically answer the question whether “the nature of the change in the system may be different [in situations where the majority of speakers are bilingual] from what is found when only a minority of the speakers are bilingual” (ibid.). Unlike Jones’s study, Thomas & Gathercole’s is semi-naturalistic (the subjects were administered a story-telling task). Data were collected from both adults and children but again only the findings related to the adults will be reviewed here. The adults were divided into three age groups: 16–30 (I), 31–50 (II) and 51–60 (III).

<sup>48</sup>Thomas & Gathercole (Thomas & Gathercole, 2005) base their estimate on the 1991 Census figures reported by Aitchison & Carter (1994: 89–93), according to which more than 60% of the population in this area spoke Welsh.

As regards article agreement, a repeated measures ANOVA was carried out in which animacy, grammatical gender and age were treated as independent variables. The only main effect was gender and there were no interaction effects. Masculine nouns were never mutated after the article while feminine nouns were mutated 88.3% of the time; however there was one speaker who never mutated feminine nouns, and if his data are ignored the percentage of mutation for feminine nouns rises to 90%. There was no statistically significant difference in performance across age groups and no significant effect of animacy either. Not enough examples of adjectives were produced by the adults to allow for a statistical analysis of the data. Nevertheless it was observed that adjectives were never mutated after masculine nouns, but quite frequently after feminine nouns: excluding occurrences of *bach* (see §5.7.4.1), SM of the attributive adjective after a feminine noun occurred 86% of the time in group I, 71% in group II and 93% in group III. It was not possible to collect a large enough amount of data about pronominal agreement or agreement with numerals.

Thomas & Gathercole's conclusion about gender agreement marking within the noun phrase, in the speech community under observation, is that it "remains grammatical in nature: performance does not reflect a sudden change to represent semantic or natural gender in 'local' gender-marked constructs" (2005: 2250).<sup>49</sup>

Thomas & Gathercole (2005: 2253) also argue that their findings present a "different picture of the processes of change currently underway in Welsh" from those of Jones (1998a), the difference being that in their study both younger and older adults perform at the same level, which "clearly suggests that gender distinctions are stable in their language" (*ibid.*). This—they claim—is due to the fact that bilingualism is dominant in this area: "Due to the stronger element of bilingualism in Gwynedd and Anglesey, the frequency of exposure to the gender may be adequate to maintain the system" (Thomas & Gathercole, 2005: 2254).<sup>50</sup>

---

<sup>49</sup>The findings of the present study also support the conclusion that gender agreement marking within the noun phrase is not influenced by semantic or natural gender.

<sup>50</sup>The role played by frequency of exposure is another important point to which I will return in chapters 7 and 8.

### 5.7.4.3 Variation in gender assignment

Certain nouns may be masculine or feminine depending on the dialect,<sup>51</sup> while others oscillate between masculine and feminine assignment within the same variety; some of the latter are common, high-frequency nouns like *braich* ‘arm’, *troed* ‘foot’ and *penelin* ‘elbow’ (SurrIDGE, 1989: 202).<sup>52</sup>

A few frequent nouns take inconsistent agreement forms: *munud* ‘minute’, for instance, is masculine in northern dialects except in the common phrase *y<sup>SMR</sup> funud hon* ‘this.F minute’; similarly, m. *man* ‘place’ is mutated in the fixed expression *yn y<sup>SMR</sup> fan* (‘immediately’, lit. ‘in the place’); *math* ‘kind’ is treated as masculine in the phrase *y math hwn* ‘this.M kind’, but as feminine in the phrase *y<sup>SMR</sup> fath beth* ‘this kind of thing, such a thing’ (Thomas, 2001: 83; Thorne, 1993: 116).<sup>53</sup>

## 5.8 Breton

Different dialects of Breton are spoken in the rural areas of Brittany and the few remaining native speakers typically speak only their own local variety; mutual intelligibility is often problematic (Ternes, 1992). The revivalists’ efforts to introduce a common standard have not been successful as native speakers tend to refuse the revivalists’ “colourless” variety (Kuter, 1989), and the language is widely regarded as being “severely endangered” (Moseley, 2010).

The process of standardization itself has been long and troublesome: the hundred-year-old quest for a unified orthography, started in 1907, is perhaps one of its most emblematic aspects (Jones, 1994: 304–10).<sup>54</sup> Breton has two standards, based on the Leon and Gwenedeg dialects, respectively; the former is the more conservative and is sometimes called “standard Breton”

<sup>51</sup>A few examples are listed by Thomas (2001: 82),

<sup>52</sup>Uncertain assignment is also found with recent borrowings: e.g. *blows* ‘blouse’, *record* ‘record’, *coler* ‘collar’ etc. (Thorne, 1993: 116).

<sup>53</sup>According to Borsley et al. (2007: 174), *fath* in expressions such as *y fath beth*, although historically derived from the noun *math*, is now “probably an independent item, bearing fixed soft mutation”.

<sup>54</sup>There are today some four orthographical systems for standard Breton, and two for standard Gwenedeg (Ternes, 1992: 382). The main sources for this section, Stephens (1993) and Ternes (1992), both use the same orthographic system, namely *zedachek*, which is therefore adopted here.

tout court, while the latter is “so divergent from all other dialects that it might be called a language of its own on purely linguistic grounds” (Ternes, 1992: 380f.).

Breton has two genders, masculine and feminine. As Hemon (1975: 26f.) observes, gender in Breton is highly covert, and its assignment is sometimes very inconsistent across the various dialects.

## 5.8.1 Realization and scope of agreement

### 5.8.1.1 Soft mutation (SM)

The effects of SM (or “lenition”) in Breton can be represented as in Table 5.24.<sup>55</sup> Voiceless plosives and fricatives become voiced, while voiced plosives become homorganic fricatives.

Table 5.24: SM in Breton

|                   | Basic consonant |     | Mutated consonant |     |
|-------------------|-----------------|-----|-------------------|-----|
| <b>Plosives</b>   | ⟨k⟩             | /k/ | ⟨g⟩               | /g/ |
|                   | ⟨g⟩             | /g/ | ⟨c’h⟩             | /h/ |
|                   | ⟨t⟩             | /t/ | ⟨d⟩               | /d/ |
|                   | ⟨d⟩             | /d/ | ⟨z⟩               | /z/ |
|                   | ⟨p⟩             | /p/ | ⟨b⟩               | /b/ |
|                   | ⟨b⟩             | /b/ | ⟨v⟩               | /v/ |
| <b>Fricatives</b> | ⟨f⟩             | /f/ | ⟨f⟩               | /v/ |
|                   | ⟨s⟩             | /s/ | ⟨s⟩               | /z/ |
|                   | ⟨ch⟩            | /ʃ/ | ⟨ch⟩              | /ʒ/ |
|                   | ⟨c’h⟩           | /x/ | ⟨c’h⟩             | /ɣ/ |
| <b>Nasals</b>     | ⟨m⟩             | /m/ | ⟨v⟩               | /v/ |

(After Stephens, 1993: 359–61 and Ternes, 1992: 443)

### 5.8.1.2 Aspirate mutation (“spirantization”)

AM (or “spirantization”) involves the voiceless plosives /p, t, k/, replaced by the homorganic fricative /f, z, x/, respectively (Ternes, 1992: 428, 443).

<sup>55</sup>Note that *c’h* is a trigraph which represents /h/ in the context of initial mutation and /x/ elsewhere, and that SM of fricatives is not represented orthographically.

AM is receding but tends to be preserved where it marks gender agreement (Stephens, 1993: 363).

### 5.8.1.3 Interfering sandhi phenomena

A synchronically active sandhi rule operating across the word boundary constrains gender agreement by neutralizing the effect of lenition in certain phonological contexts: any two homorganic adjacent plosives or fricatives across a word boundary will be realized as voiceless: so for instance /s # t/, /s # d/ and /z # d/ will all be realized as [st] (Ternes, 1992: 436; Stephens, 1993: 358).

(81) Breton (Stephens, 1993: 358)

*ed du* ‘buckwheat’ /ed dy/ → [ety]

### 5.8.1.4 Articles

Breton has both a definite article (*ar*) and an indefinite one (*ur*). Both have allomorphs in phonologically conditioned distribution, as shown in Table 5.25.

Table 5.25: Forms of the articles in Breton

| Form | Example      | Gloss       | Conditions         |
|------|--------------|-------------|--------------------|
| an   | an tan       | the fire    | before dental or V |
| un   | un anv       | a name      |                    |
| al   | al loa       | the spoon   | before /l/         |
| ul   | ul logodenn  | a mouse     |                    |
| ar   | ar beisanted | the farmers | elsewhere          |
| ur   | ur pesketer  | a fisherman |                    |

(Cf. Stephens, 1993: 385)

Gender agreement is marked as shown in Table 5.26: feminine nouns in the singular, and a subset of masculines (only those denoting human referents) in the plural, undergo SM after the article (Stephens, 1993: 361–363). Things are further complicated by the fact that initial /k/ always

Table 5.26: Initial mutations after the article in Breton

|          | [+human]    |             | [-human]    |             |
|----------|-------------|-------------|-------------|-------------|
|          | masc.       | fem.        | masc.       | fem.        |
| singular | no mutation | SM          | no mutation | SM          |
| plural   | SM          | no mutation | no mutation | no mutation |

undergoes AM in this context, regardless of gender and semantics, while initial /d/ is not mutated.

(82) Breton (Stephens, 1993: 361): masculine singular

- a. *pont* ‘bridge’ → *ar pont* ‘the bridge’
- b. *ki* ‘dog’ → *ar c’hi* ‘the dog’

(83) Breton (Stephens, 1993: ibid.): feminine singular

- a. *pluenn* ‘feather’ → *ar bluenn* ‘the feather’
- b. *seizhvet* /sɛjzvət/ ‘seventh’ → *ar seizhvet* /zɛjzvət/ ‘the seventh.F’

(84) Breton (Stephens, 1993: ibid.): masculine plural [+human]

- paotred* ‘boys’ → *ar baotred* ‘the boys’

### 5.8.1.5 Adjectives

Attributive adjectives also mark gender agreement via SM. The triggering features for SM are the same as for the article, namely feminine gender in the singular and masculine gender in the plural for nouns denoting humans. Spirantization of initial /k/ does not apply to the adjective, and the actual realization of agreement is constrained by the phonological context in a more restrictive way than it is for nouns:

1. if the noun ends in a vowel, or in one the sonorants /m, n, l, r/, SM is always triggered;
2. otherwise, SM only applies to adjectives beginning with /b, g, gw, m/.

(85) Breton (Stephens, 1993: 361f.)

- a. *ur*<sup>SM</sup> *\_wezenn* *d**ev* (cf. *gwezenn, tev*)  
 ART.F.SG tree(F).SG fat.F.SG  
 ‘a thick tree’

Table 5.27: Gendered numerals in Breton

|   | Masculine | Feminine |
|---|-----------|----------|
| 2 | daou      | div      |
| 3 | tri       | teir     |
| 4 | pevar     | peder    |

- b. ur<sup>SM</sup>      gazeg      \_wenn      (cf. *kazeg*, *gwenn*)  
 ART.F.SG mare(F).SG white.F.SG  
 ‘a white mare’

### 5.8.1.6 Attributive nouns

In compounds (Ternes, 1992: 426), attributive nouns behave like adjectives, undergoing SM after a feminine singular head noun.

(86) Breton (Ternes, 1992: 426)

- |              |                     |                     |
|--------------|---------------------|---------------------|
| a. kroc’henn | b. askell-groc’henn | c. eskell-kroc’henn |
|              | wing(F).SG-skin     | wing(F).PL-skin     |
| ‘skin’       | ‘bat’               | ‘bats’              |

### 5.8.1.7 Numerals

The cardinal numerals for “2”, “3” and “4” mark gender agreement, as shown in Table 5.27 and in the following examples.

(87) Breton (Ternes, 1992: 406)

- |                |                |
|----------------|----------------|
| a. daou di     | b. div daol    |
| two.M house(M) | two.F table(F) |
| ‘two houses’   | ‘two tables’   |

### 5.8.1.8 Personal pronouns

Pronouns only mark gender agreement in the third person singular (Table 5.28). The two proclitic pronouns, *e* (m.) and *he* (f.), both /e/, are only distinguished by the different initial mutations they trigger, as illustrated by (88).

Table 5.28: Personal pronouns in Breton

|         | Stressed | Unstressed                          |          |
|---------|----------|-------------------------------------|----------|
|         |          | proclitic                           | enclitic |
| 3sg. m. | en       | e /e <sup>SM</sup> /                | -en      |
| 3sg. f. | hi       | he /e <sup>AM</sup> /<br>hec'h /ex/ | -hi      |

(88) Breton (Ternes, 1992: 403)

- |                       |                        |
|-----------------------|------------------------|
| a. e <sup>SM</sup> di | b. he <sup>AM</sup> zi |
| 3SG.M house           | 3SG.F house            |
| ‘his house’           | ‘her house’            |

The form *hec'h /ex/* is used before a vowel for agreement with a feminine antecedent.

(89) Breton (Ternes, 1992: 442)

- hec'h anv  
3SG.F name  
‘her name’

An optional enclitic pronoun can avoid ambiguity about the gender of the antecedent if the appropriate initial mutation cannot apply.

(90) Breton (Stephens, 1993: 393)

- |                |                |
|----------------|----------------|
| a. e levr-en   | b. he levr-hi  |
| 3SG book-3SG.M | 3SG book-3SG.F |
| ‘his book’     | ‘her book’     |

## 5.8.2 Gender assignment

### 5.8.2.1 Semantic assignment

For sex-differentiated animates, grammatical gender usually coincides with biological gender (e.g. m. *tarv* ‘bull’, f. *buoc'h* ‘cow’) (Stephens, 1993: 364; Ternes, 1992: 414). Some secondary semantic criteria are illustrated by Stephens (1993: 364): metals, names of mountains, time divisions (excepts *eur* ‘hour’, *noz* ‘night’, *sizhun* ‘week’) are associated with the masculine gender; names of countries, cities and rivers tend to be feminine.

### 5.8.2.2 Formal assignment

Association with grammatical gender is observed for certain derivational suffixes (Stephens, 1993: 364); singulative nouns in particular (morphologically derived from collective nouns) are consistently assigned to the feminine gender (91).

(91) Singulatives in Breton (Ternes, 1992: 416)

| Collective    |         | Singulative (feminine) |        |
|---------------|---------|------------------------|--------|
| <i>merien</i> | ‘ants’  | <i>merien-enn</i>      | ‘ant’  |
| <i>gwez</i>   | ‘trees’ | <i>gwez-enn</i>        | ‘tree’ |

However, formal criteria are not very reliable gender predictors and a number of nouns are assigned to different genders in different dialects (Stephens, 1993: 364).

## 5.9 Cornish

There was no native speakers of Cornish left by the end of the eighteenth century but Cornish had already ceased to be a community language some time before; its last speakers were a geographically scattered group of individuals (George, 1993: 414; Thomas, 1992a: 346). The present account is based primarily on George (1993) and Thomas (1992a), both of whom are concerned with Traditional Cornish (i.e. the language as spoken before its tradition was interrupted) and ignore the revivalists’ variety.<sup>56</sup>

The extreme dearth of literary texts and the lack of a literary tradition of any prestige (the Bible was never entirely translated in Cornish) resulted in a remarkably inconsistent orthography, which—coupled with the absence of any acoustic record—makes the reconstruction of the phonology and morphology of Traditional Cornish an arduous task (George, 1993: 417–21; Thomas, 1992a: 347).

### 5.9.1 Realization and scope of agreement

Traditional Cornish had two genders, masculine and feminine (Jenner, 1904: 78).

<sup>56</sup>The revival movement was not started before the 20th century (George & Broderick, 1993: 645).

### 5.9.1.1 Soft mutation

Cornish SM (Thomas, 1992a), or lenition (George, 1993), was substantially similar to Welsh SM (Table 5.29).

Table 5.29: SM in Cornish

|                 | Basic consonant | Mutated consonant |
|-----------------|-----------------|-------------------|
| <b>Plosives</b> | /k/             | /g/               |
|                 | /g/             | /∅/               |
|                 | /t/             | /d/               |
|                 | /d/             | /ð/               |
|                 | /p/             | /b/               |
|                 | /b/             | /v/               |
| <b>Nasals</b>   | /m/             | /v/               |

(After George, 1993: 437)

### 5.9.1.2 Spirant mutation

Cornish spirantization (George, 1993), or spirant mutation (Thomas, 1992a), repeats the pattern of Welsh AM (Table 5.30).

Table 5.30: Spirant mutation in Cornish

|                 | Basic consonant | Mutated consonant |
|-----------------|-----------------|-------------------|
| <b>Plosives</b> | /k/             | /h/               |
|                 | /t/             | /θ/               |
|                 | /p/             | /f/               |

(After George, 1993: 437)

### 5.9.1.3 Article(s)

According to George (1993: 439), Traditional Cornish only possessed a definite article *an*; the numeral *un* ‘one’ was used to express indefiniteness (‘a certain’), and Thomas (1992a: 353) considers it an indefinite article. Until

at least the Middle Cornish period (approximately 1200–1575), lenition after the definite article was occurred with feminine nouns in the singular and with [+human] masculines in the plural (as in Breton: see §5.8.1.4 above).

(92) Cornish (Jenner, 1904: 73)

a. an<sup>SM</sup>        yenen        (cf. *benen*)  
 ART.F.SG woman(F).SG  
 ‘the woman’

b. an<sup>SM</sup>        dassow        (cf. *tassow*)  
 ART.M.PL father(M).PL  
 ‘the fathers’

Mutation after the indefinite article occurred with feminine nouns in the singular (George, 1993: 438).

#### 5.9.1.4 Adjectives

According to Jenner (1904: 71), SM of adjectives was consistently found as a gender agreement marker only after feminine singular nouns; however, its correlation with gender was somewhat weakened by the fact that all nouns ending in *-a /ə/* in the singular also mutated the following adjective, e.g. *sîra* ‘father’ : *sîra* *widn* ‘grandfather’ (cf. unmutated *gwidn*; Jenner, 1904: *ibid.*).<sup>57</sup>

(93) Cornish (Jenner, 1904: 71)

a. tâs            mas  
 father(M).SG good  
 ‘a good father’

b. mergh        yas  
 daughter(F).SG good.F.SG  
 ‘a good daughter’

(94) Cornish (Jenner, 1904: *ibid.*)

a. tassow        mas  
 father(M).PL good  
 ‘good fathers’

b. merhes        mas  
 daughter(F).PL good  
 ‘good daughters’

#### 5.9.1.5 Demonstrative pronouns

Demonstrative pronouns in Middle Cornish had gender-specific forms in the singular, but by the end of the Late Cornish period (1575–1800) the proximal

<sup>57</sup>According to George (1993: 437), SM was also found after masculine plural nouns with human reference.

Table 5.31: Demonstrative pronouns in Middle and Late Cornish

|        |    | Middle Cornish | Late Cornish |
|--------|----|----------------|--------------|
| ‘this’ | m. | hemma/helma    | ebah         |
|        | f. | homma/holma    |              |
| ‘that’ | m. | henna          | hedda        |
|        | f. | honna          | hodda        |

(After George, 1993: 440)

form no longer marked gender agreement (Table 5.31).

### 5.9.1.6 Numerals

In Middle Cornish, the numerals corresponding to ‘2’, ‘3’ and ‘4’ had distinct morphemes to mark gender agreement: *dev/dew* (m.) : *dyv/dyw* (f.) ‘two’; *try* (m.) : *tyr* (f.) ‘three’; *peswar* (m.) : *peder* (f.) ‘four’ (Jenner, 1904: 94; George, 1993: 444).

### 5.9.1.7 Personal pronouns

Middle Cornish had a number of pronominal series: independent, suffixed (enclitic), possessive (proclitic), and infix; in Late Cornish, independent and enclitic pronouns eventually fell together (George, 1993: 440–2). Gender was distinguished in the third person singular. In the possessive series, masculine *y* lenited the following word, while feminine *hy* triggered spirant mutation (George, 1993: 437f.), as in the following examples.

(95) Cornish (Jenner, 1904: 71)

|                    |                               |                    |                                |
|--------------------|-------------------------------|--------------------|--------------------------------|
| a. e <sup>SM</sup> | <u>b</u> en (cf. <i>pen</i> ) | b. e <sup>SM</sup> | <u>dh</u> ên (cf. <i>dên</i> ) |
|                    | POSS:3SG.M head               |                    | POSS:3SG.M man                 |
|                    | ‘his head’                    |                    | ‘his man’                      |

(96) Cornish (Thomas, 1992a: 368)

|                     |                            |                     |                              |
|---------------------|----------------------------|---------------------|------------------------------|
| a. /i <sup>AM</sup> | <u>θ</u> ir/ (cf. /tir/) ) | b. /i <sup>AM</sup> | <u>x</u> olon/ (cf. /kolon/) |
|                     | POSS:3SG.F land            |                     | POSS:3SG.F heart             |
|                     | ‘her land’                 |                     | ‘her heart’                  |

### 5.9.2 Gender assignment

In Late Cornish, gender assignment was exclusively semantic:

[t]here is no rule whereby to tell the gender of a word, except in the case of animate objects, where the gender simply follows the sex. [Jenner, 1904: 78]

Specific formal features that were once associated with either grammatical gender had become been irrelevant in Late Cornish, and only fossilized remnants of SM in phrases headed by feminine singular nouns survived: cf. /bro/ ‘area’ vs. /an vro/ ‘the area’ (Thomas, 1992a: 368).

## 5.10 Summary

In this chapter some facts about the history of the Insular Celtic languages were discussed, and their gender systems were examined in some detail; a comparison between conservative and less conservative varieties was provided where meaningful, in order to illustrate the complexity of the former vis-à-vis the simplification observed in the latter. Particular attention was paid to theoretical issues surrounding the so-called initial mutations, which are—among other things—used as gender agreement markers (§5.3); the analysis of various theoretical approaches led to the proposal of a morphological model of initial mutation in FDG (§5.3.3).

In the four languages that are still spoken natively (Irish, ScG, Welsh and Breton), gender agreement displays some uncertainties concerning the assignment of inanimates, whereas in Manx and Breton masculine forms had been almost completely overgeneralized except in the case of pronominal reference to females. Grammatical gender in Insular Celtic is a fairly covert category: very few formal features of the noun univocally correlates with either gender; semantic criteria, other than biological gender, are of limited validity.

The initial mutation system is also quite complex and variable: for example, lenition and soft mutation, which in the noun phrase mark gender agreement with a feminine noun, mark agreement with a masculine antecedent in anaphoric possessive use. In Breton and Cornish, soft mutation may also be found in masculine noun phrases. Perhaps more importantly,

the mutation system is limited in scope: agreement targets are necessarily only those with a mutable initial.<sup>58</sup>

Together, chapters 5–7 form the analytic core of this study. The survey just conducted is preparatory to the work done in chapters 6 and 7, where the variation described qualitatively in this chapter is analyzed quantitatively and a theoretical model of the system is provided.

---

<sup>58</sup>The available literature does not provide any numerical data as to what percentage of the nominal lexicon of each language have mutable initials. If available, such figures would reveal how effective initial mutations are as an agreement-marking system. However, in order to get a representative picture, it would be necessary to rank all the nouns in the lexicon according to their relative frequency, which would require very large corpora.



## Chapter 6

# Gender agreement in the noun phrase

### 6.1 Overview

Having presented the grammatical gender system of Insular-Celtic in the previous chapter, we are now in a position to consider the variation observed in the data.

This chapter is structured as follows: in §6.2 I present the findings of the statistical analysis carried out on the Irish and Welsh data sets described in chapter 3. The typological and theoretical problems that these findings pose is also discussed.

The model of agreement proposed in §6.3 takes into account the significance of the data and is able to accommodate both the traditional agreement system of Irish and Welsh and the new system which characterizes less conservative varieties; it is in line with what we know about agreement patterns in a typological perspective and accounts for the language-internal aspect of the observed linguistic changes.

### 6.2 Variation in gender agreement within the noun phrase: the data

The traditional agreement systems of Irish and Welsh were described in §§5.4 and 5.7 respectively, where it was remarked that they have undergone some changes, including, in Irish, through the loss of certain phonetic

oppositions (§5.4.1.1) and the reported reduction in usage of attributive adjectives (§5.4.1.5); in spoken Welsh, particularly significant from this point of view was the reduced number of agreement targets (e.g. as uninflected deictics replace inflected ones, §5.7.1.7) and the tendency to leave attributive adjectives and especially attributive nouns unmutated (§§5.7.1.5 and 5.7.1.6). Furthermore, as is also apparent from the previous chapter, and in particular from the description of gender agreement of Scottish Gaelic (§5.5), Manx (§5.6) and Cornish (§5.9), it seems possible to trace a path of linguistic change leading to a gender system in which consistent gender marking is found only with anaphoric pronouns and is almost exclusively semantic in nature (I return to anaphoric agreement in chapter 7).

The traditional system of gender agreement within the noun phrase is well reflected in OC-I and OC-W. As regards CC-I and CC-W, the picture emerging from the data is at variance with the traditional system, especially as regards the treatment of prenominal agreement targets as opposed to postnominal ones.

## 6.2.1 Irish

### 6.2.1.1 Article agreement

As illustrated in Figure 6.1, grammatical gender agreement between the article and the noun is regularly observed with both masculine and feminine controllers in both OC-I and CC-I, although in the latter it is significantly less frequent with feminines than with masculines.<sup>1</sup>

<sup>1</sup>As we have seen in §5.4.1.4, agreement with the article is marked by the initial mutation of the following noun and, only in the genitive singular, also by the form of the article itself. In the latter case, it was decided that if the expected form of the article were found, then a noun phrase would be regarded as showing article agreement even when the expected mutation was not produced. As we have also seen, certain otherwise lenitable consonants, *t-* and *d-*, are traditionally left unmutated after the article regardless of gender and case. Initially, nouns beginning with these consonants were included in the analysis, in order to establish whether initial *t-* and *d-* might sometimes by analogy be lenited after the article in the nominative or genitive singular. This, however, was not found to be the case: in fact, lenition of initial /d'/ is only found once in OC-I. Therefore, these nouns were not considered.

Nouns which follow the article in a prepositional phrase only mark gender agreement when they begin in lenitable *s-* (cf. §5.4.1.1). Even though we have distinguished two prepositional cases based on the type of mutation that follows the article, the figures for these two cases are conflated in what follows because for *s*-initial nouns gender marking

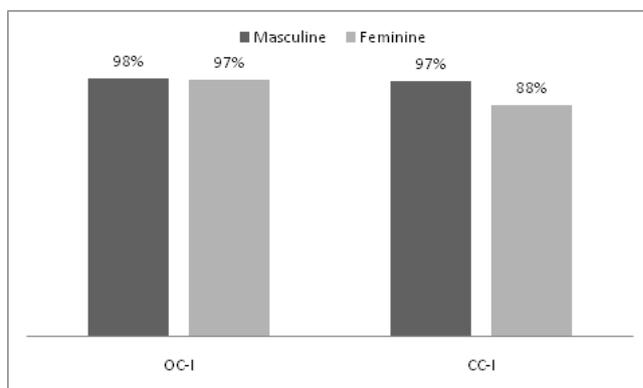


Figure 6.1: Article agreement in Irish, OC-I vs. CC-I

In OC-I, agreement rate with the article is almost the same for both masculine (98%) and feminine controllers (97%), with no statistically significant difference, while in CC-I it appears to be significantly more frequent with masculines (97%) than with feminines (88%), as shown in Table 6.1.

Table 6.1: Agreement with the article in Irish

|            | OC-I         |      | CC-I         |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 227          | 107  | 176          | 136  |
| Total      | 232          | 110  | 182          | 155  |
| % agreeing | 98           | 97   | 97           | 88   |
|            | $p = 0.7154$ |      | $p = 0.0028$ |      |

As explained in §3.2.2.2, CC-I is comprised of two sub-components, one including material from the national Gaeltacht-based broadcaster (RnaG), the other material from broadcasters outside the official Irish-speaking districts. If results for the two sub-components of CC-I are considered separately it can be seen that the difference between the two genders appears to be significant in both the RnaG and the other set (Table 6.2).<sup>2</sup>

Finally, it was observed that the agreement rate with feminine controllers

---

is identical in both cases.

<sup>2</sup>However, in the case of the RnaG sub-component the confidence interval for the odds ratio (cf. §3.3.1) straddles 1, if only by very little (0.96 to 14.5), therefore we cannot be sure that there is a real difference in the RnaG sub-component.

Table 6.2: Agreement with the article in CC-I, RnaG vs. others

|            | RnaG         |      | Others       |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 114          | 103  | 62           | 33   |
| Total      | 118          | 115  | 64           | 40   |
| % agreeing | 97           | 90   | 97           | 82.5 |
|            | $p = 0.0395$ |      | $p = 0.0256$ |      |

is higher in OC-I than in CC-I (97% vs. 88%, respectively), which is a statistically significant difference ( $p = 0.0059$ ) not observed in the case of masculine controllers ( $p = 0.546$ ). The presence of semantic agreement (Appendix A.1.3) and proper names (Appendix A.1.4) was found to be negligible and to have no significant impact on article agreement rates.

### 6.2.1.2 Adjective agreement

As regards the attributive adjective (Figure 6.2 and Table 6.3), no statistical evidence of any between-gender difference in agreement rates in OC-I was found: grammatical agreement was observed 94% of the time with masculines and 92% with feminines, a negligible difference ( $p = 0.6853$ ). In CC-I, on the other hand, adjectives grammatically agree with masculines significantly more often than they do with feminines ( $p < 0.0001$ ): the rate of grammatical agreement is 91% with masculines and only at 45%, i.e. less than chance level, with feminines.<sup>3</sup>

As can be seen in Table 6.4, the between-gender difference in agreement rate is particularly evident in the non-RnaG sub-component of CC-I, where only one adjective out of six shows the expected form; in the RnaG sub-component, the agreement rate with feminines is just above chance level at

<sup>3</sup>Like articles, adjectives traditionally rely on two different strategies of agreement marking: initial mutations and modification of the endings. The latter strategy is only possible in the genitive singular, of which just 11 tokens were found (six in OC-I and five in CC-I). (Morphological gender marking was available to only six of them, all of which show morphological gender agreement as expected. However, five of these are proper names. As regards their distribution, one of these six occurrences was found in CC-I and the other five in OC-I.) Therefore, no meaningful generalization can be made about the regularity of morphologic agreement and what follows only refers to agreement via initial mutations.

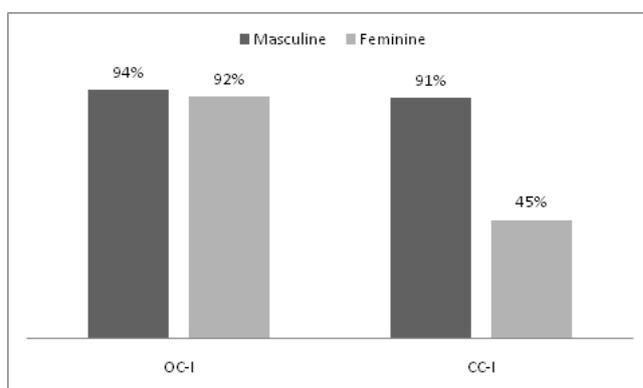


Figure 6.2: Adjective agreement in Irish, OC-I vs. CC-I

Table 6.3: Agreement with the adjective in Irish

|            | OC-I         |      | CC-I         |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 49           | 33   | 51           | 9    |
| Total      | 52           | 36   | 56           | 20   |
| % agreeing | 94           | 92   | 91           | 45   |
|            | $p = 0.6853$ |      | $p < 0.0001$ |      |

57%. In both components the between-gender difference was found to be significant (RnaG:  $p = 0.0042$ ; others:  $p = 0.0037$ ).

All in all, we can observe that attributive adjectives do not occur too frequently in either component: there are less than 100 such occurrences in each. Although it has to be borne in mind that only adjectives that could mark gender agreement were counted, this observation is consistent with Mac Eoin's (1993: 116) aforementioned remark (see §5.4.1.5 above) about the decline of the attributive position in spoken Irish.

As already observed in relation to the article, the agreement rate of attributive adjectives with feminine controllers is significantly higher in OC-I than in CC-I (92% and 45%, respectively,  $p = 0.0002$ ), a difference that is not observed in the case of masculine controllers ( $p = 0.7175$ ).

Only two occurrences of semantic agreement were found, both in OC-I, one with a masculine and one with a feminine controller. The impact of the semantic factor can therefore be deemed negligible, as the impact of adjective agreement forms within proper names (§A.2.1).

Table 6.4: Agreement with the adjective, RnaG vs. others

|            | RnaG         |      | Others       |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 33           | 8    | 18           | 1    |
| Total      | 35           | 14   | 21           | 6    |
| % agreeing | 94           | 57   | 86           | 17   |
|            | $p = 0.0042$ |      | $p = 0.0037$ |      |

### 6.2.1.3 Agreement with attributive nouns

As explained in §5.4.1.6, attributive nouns are expected to undergo lenition when they depend on a feminine noun standing in the nominative; in Connacht Irish, one variety of which is represented by OC-I, the attributive noun is also expected to lenite after a masculine head noun in the genitive singular, in line with adjective lenition norms (de Bhaldraithe, 1953: 262); in Munster Irish (cf. Ó Sé, 2000: 61 for the Corca Dhuibhne variety), and in Standard Irish (cf. Caighdeán, 1979: 83f.), on the other hand, lenition of the attributive noun cannot be said to productively apply in the genitive masculine. However, the above is true of indefinite dependents only, as definite ones are supposed to be lenited regardless of the grammatical gender of the head noun—although this was not always found to be the case in our corpus. Furthermore, attributive nouns can be considered modifiers within their noun phrase, thus standing in opposition with dependent nouns and noun phrases whose relationship with the head noun is that, or closer to that, of an argument: e.g., object or subject of a verbal noun, partitive, possessive, etc. This distinction is important, since nouns which function as arguments of the head do not normally undergo lenition. Some examples of dependent nouns and noun phrases from the Irish corpus are given in Table 6.5.

In OC-I, only ten out of 34 dependent nouns (i.e. 29%) are lenited after a feminine noun. However, in two cases lenition is blocked by the interidental context (see §5.4.1.1) as would be expected (e.g. *beirt deartháir* ‘two brothers’), which means that lenition is found in ten out of 32 occurrences where it could be expected based on phonological considerations only (or 31% of the time). All in all, dependents’ lenition in OC-I is in line with

Table 6.5: Types of dependent noun phrases in Irish

| Category        | Examples   |
|-----------------|--|
| Attributive     | <i>seomra cúirte</i> [room court] ‘courtroom’ (RnaG “Cumarsáid” 06/11/1997)<br><i>soitheach seoil</i> [vessel sail] ‘sailing vessel’ (Wigger, 2000 1-01-14)                |
| Matter          | <i>maológ mhóna</i> [heap turf] ‘a heap of turf’ (Wigger, 2000 6-01-14)<br><i>cóta teara</i> [coat tar] ‘a coat of tar’ (Wigger, 2000 4-03-03)                             |
| Object (of VN)  | <i>a(g) lorg duine</i> [at seeking person] ‘looking for someone’ (RnaG “Barrscéalta” 23/10/1997)<br><i>cnuchairt mhóin</i> [stacking turf] (RnaG “Seal Aneas” 21/10/1997)  |
| Part-whole      | <i>pointe boise</i> [tip palm] ‘tip of the hand’ (Wigger, 2000 1-01-27)<br><i>deireadh seachtaine</i> [end week] ‘weekend’ (Newstalk “Splanc” 19/01/2007)                  |
| Partitive       | <i>beirt fhear</i> [pair men(GEN.PL)] ‘two men’ (Wigger, 2000 4-03-05)<br><i>neart báistí</i> [plenty rain] ‘plenty of rain’ (Wigger, 2000 4-03-07)                        |
| Possessive      | <i>teanga sionnach</i> [tongue fox] ‘fox’s tongue’ (Wigger, 2000 6-01-08)<br><i>teach duine uasail</i> [house gentleman] ‘the house of a gentleman’ (Wigger, 2000 1-01-21) |
| Subject (of VN) | <i>briseadh croí</i> [breaking heart] ‘heartbreak’ (Wigger, 2000 1-01-28)<br><i>beannacht Dé</i> [blessing God] ‘God’s blessing’ (RnaG “I Measc na nDaoine” 22/03/2007)    |

what grammatical descriptions tell us: lenition is not found, in particular, following compound prepositions, verbal nouns, and when a part-whole relationship or partitive is instantiated (with the consistent exception of partitives introduced by the feminine noun *beirt* ‘pair’, which never fails to trigger lenition except in interdental phonological contexts; cf. §5.4.1.6). If all such instances, in which the grammatical descriptions tell us that lenition is not to be expected, are not counted, then the proportions of dependents lenited after feminine nouns goes up to ten out of 16 (62.5%).

In comparison, only two out of 55 dependent nouns (or 4%) are lenited after a masculine noun, although neither, contrary to what might be expected, is in the genitive singular.

In CC-I, dependents’ lenition appears to be quite infrequent following both masculine nouns (4/108 or about 4%) and feminine ones (2/24 or 8%). Contrary to expectation, lenition after masculine nouns involves not only the genitive case (one occurrence out of one) but also the nominative (three occurrences), where it is not easily explained in terms of semantic relation, lexical collocation or otherwise.

Lenition after feminine nouns, however, also seems to be rather exceptional. There are only four occurrences in which lenition would traditionally be blocked by the interdental context, and if these are excluded the proportion of lenited nouns goes up to 2/20 or 10%. A closer scrutiny of the only two lenited observations reveals that both are found in the same interview and uttered by the same speaker, a radio presenter from RnaG.

## 6.2.2 Welsh

### 6.2.2.1 Article agreement

No evidence of any significant difference in agreement rates was found in the Welsh corpus in relation to the article: the expected agreement forms appear more often than 95% of the time with controllers of either gender, in both OC-W and CC-W (Figure 6.3 and Table 6.6).<sup>4</sup>

<sup>4</sup>As seen in §5.7, there is no case marking in Welsh, and soft mutation after the article only marks feminine agreement. Normally (cf. §5.7.1.4), the article does not trigger soft mutation on initial *ll* (/l̪/) or *rh* (/r<sup>h</sup>/). In certain geographical areas, however, soft mutation is triggered by the article on these two segments as well (cf. Thomas & Thomas, 1989: 138). However, no examples were found in the relevant texts in which mutation of these two segments was to be expected.

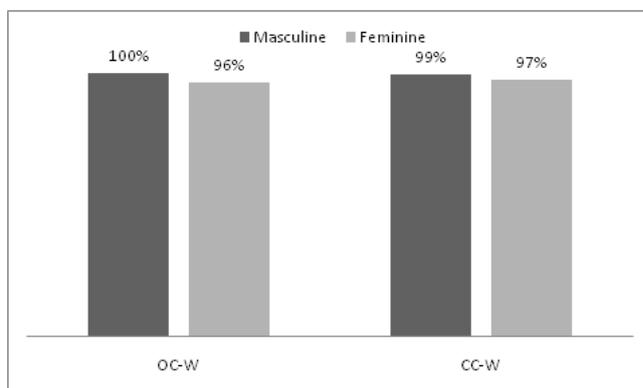


Figure 6.3: Article agreement in Welsh, OC-W vs. CC-W

Table 6.6: Agreement with the article in Welsh

|            | OC-W         |      | CC-W         |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 174          | 51   | 261          | 72   |
| Total      | 174          | 53   | 263          | 74   |
| % agreeing | 100          | 96   | 99           | 97   |
|            | $p = 0.0537$ |      | $p = 0.2108$ |      |

A number of observations in both components were flagged as uncertain and excluded from the analysis when the possibility of gender oscillation was indicated by the two major Welsh dictionaries (Thomas, 2002; Meurig Evans & Thomas, 2007). For instance, *tôn* ‘tune’ is recorded as feminine by Meurig Evans & Thomas (2007) and as oscillating by Thomas (2002).<sup>5</sup> In OC-W, *tôn* only occurs once after the article, in its non-mutated form. In this and similar cases, it was impossible to determine whether (a) the speaker would normally mutate the noun (but failed to do so on this particular occasion) or (b) the speaker would never mutate the noun. Similar observations were therefore excluded from the analysis.

Article agreement with nouns of both genders also appears to be stable across the two components. With masculines, the difference between the

<sup>5</sup>Thomas (2002) flags oscillating nouns as either “eg.b.” (= *enw gwry-waidd/benywaidd* “masculine/feminine noun”) or “eb.g.” (= *enw benywaidd/gwrywaidd* “feminine/masculine noun”), in what seems to imply that the noun is more commonly masculine in one case and more commonly feminine in the other.

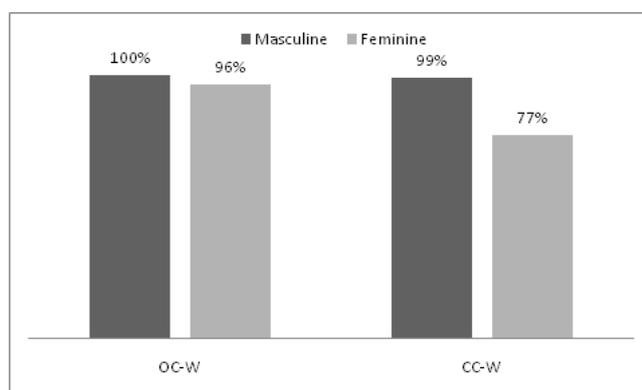


Figure 6.4: Adjective agreement in Welsh, OC-W vs. CC-W

two components is just one percentage point (100% in OC-W as opposed to 99% in CC-W,  $p = 0.5196$ ); with feminines, the difference is also one percentage point (96% in OC-W as opposed to 97% in CC-W) and is also statistically not significant ( $p = 1$ ).

The impact of semantic agreement (Appendix B.1.1) and of that of proper names (Appendix B.1.2) were both found to be negligible.

### 6.2.2.2 Adjective agreement

As regards the attributive adjective, syntactic agreement in OC-W is slightly more frequent with masculine (100%) than with feminine controllers (96%), but the difference did not test as statistically significant ( $p = 0.4286$ ). In CC-W, on the other hand, syntactic agreement with feminine controllers was found to be statistically less likely than with masculine ones (77% vs. 99%,  $p < 0.0001$ ).<sup>6</sup> This is represented in Figure 6.4 and in Table 6.7.

As was already the case for the Irish data set, syntactic agreement with feminines turned out to be significantly more frequent in OC-W than in CC-W ( $p < 0.0314$ ), while no significant between-component difference was observed in the case of masculine controllers ( $p = 1$ ).

When gender oscillation was recorded by the dictionaries for a particular noun, the observation was counted as agreeing if the adjective was mutated,

<sup>6</sup>Adjectives may agree with the head noun via initial mutation and/or by means of an internal vowel shift. As already mentioned in §5.7.1.5, the latter type of agreement is lexically constrained and no longer productive in the spoken language: in our corpus, no instance of agreement marking via vowel shift was recorded.

Table 6.7: Agreement with the adjective in Welsh

|            | OC-W         |      | CC-W         |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 36           | 26   | 82           | 47   |
| Total      | 36           | 27   | 83           | 61   |
| % agreeing | 100          | 96   | 99           | 77   |
|            | $p = 0.4286$ |      | $p < 0.0001$ |      |

and as uncertain (and thus excluded from the count) if not; this is because a non-mutated attributive adjective with a noun of oscillating gender could be either due to innovation (overgeneralization of masculine agreement with a formerly feminine noun) or adherence to a local norm; in order to provide a conservative picture of language change, I preferred to ignore such instances than to count them as innovative.

The adjective *bach* ‘small, little’ is known to systematically resist mutation in northern dialects (cf. §5.7.4.1). In OC-W, the adjective occurs after feminine nouns only three times and only when the speaker is from northern Wales; the mutated form never appears. In CC-W, the unmutated form *bach* after a feminine noun appears four times, three with northern speakers and once with a speakers of uncertain provenance. The mutated form *fach* after a feminine noun, on the other hand, appears seven times, four with a speaker from the south and three with speakers from the north. One of the latter (JON, cf. Table 3.2), in particular, uses both forms in the same phrase, as in example (1); note that in both utterances the adjective *mawr* is mutated (*fawr*).

## (1) Welsh

## a. (jonsi.cha 1189)

olwyn        fawr    yn y    front ac    olwyn        bach  
wheel(F).SG big.F.SG in ART front and wheel(F).SG small.M(?).SG  
yn y                cefn  
in ART.M.SG back(M).SG

‘a big wheel at the front and a small wheel at the back’

b. (jonsi.cha 2994)

olwyn      fawr    yn y    front ac    olwyn      fach      yn  
 wheel(F).SG big.F.SG in ART front and wheel(F).SG small.F.SG in  
 y            cefn  
 ART.M.SG back(M).SG

‘a big wheel at the front and a small wheel at the back’

Once again, to provide a conservative picture of the ongoing change, I decided to regard mutated forms of *bach* after feminine nouns as marked for feminine agreement and unmutated forms of the same adjective after feminine nouns as uncertain and to exclude the latter from subsequent analyses.

As already observed apropos of the article, semantic agreement (Appendix B.2.1) and the presence of proper names (Appendix B.2.2) were found to have a negligible impact on the observed agreement rates.

### 6.2.2.3 Agreement with attributive nouns

As seen in §5.7.1.6, gender-triggered SM applies to dependent nouns which modify the head noun in the same way as attributive adjectives, but does not normally apply to dependent nouns which express an argument of the head noun. However, this distinction is not always clear-cut. An example found in CC-W is given in (2).

(2) Welsh (sianthomas.cha 296)

pwyl drodd    i fyny ar y    stepen drws  
 who turn.PST up    on ART step(F) door

‘who turned up on the door step?’

In this example, it is unclear why the dependent noun *drws* ‘door’ is not mutated after feminine *stepen*, whether because the two stand in a part-whole relationship and *drws* is an argument of *stepen* (‘the step of the door’) or just because the speaker in question does not apply SM in this context (we cannot be sure, since this is the only opportunity for the speaker in question of producing a mutated attributive noun).<sup>7</sup> In similar cases, the occurrence was considered uncertain and not counted.

<sup>7</sup>As already observed in §5.7.1.6, SM in this context is rare even in literary forms of the language

Other occurrences which were considered uncertain and discarded are those in which the gender of the head noun is itself uncertain. Example (3) is found in CC-W.

(3) Welsh (**beti.cha** 240)

yr uned dysgu gydol oes  
ART unit learning entire lifetime

‘the life-long learning unit’

Here, *uned* ‘unit’ might be masculine or feminine, and is in fact recorded as feminine by one dictionary (Meurig Evans & Thomas, 2007) and as oscillating between feminine and masculine by the other one (Thomas, 2002). *Dysgu* ‘learning’, which heads the noun phrase *dysgu gydol oes* ‘life-long learning’, would be expected to mutate to *ddysgu* if *uned* were feminine, given that the noun phrase *dysgu gydol oes* is used attributively.

Mutation of the attributive noun after a feminine noun is found nine times out of 12 (or 75% of the time) in OC-W; conversely, it is never found after a masculine noun (45 occurrences).

In CC-W, mutation of the attributive noun is observed seven out of 12 times (or 58% of the time) after feminine nouns and two out of 39 times (or 5% of the time) after masculine nouns. The rate of SM after a feminine head noun is therefore 75% in OC-W and 58% in CC-W. This difference does not appear to be statistically significant ( $p = 0.6668$ ); however, one must bear in mind that we are dealing with rather low counts (only 12 occurrences in each component).

As only one instance of semantic agreement with a feminine controller was recorded—see (4), in CC-W—the impact of the semantic factor was deemed negligible.

(4) Welsh (**beti.cha** 413)

athrawes gerdd (cf. unmutated *cerdd*)  
teacher(F).SG music(F).SG

‘a music teacher’

It should be recalled here that the marking of gender agreement may be affected by the presence of proper names in two ways, one language-

specific and the other language-independent, as observed in §5.7.1.6. The two relevant observations are repeated here for convenience:

1. if the attributive noun is a proper name, it may display idiosyncratic mutation patterns, as just explained (language-specific);
2. if the noun phrase as a whole is a proper name, and contains a gender-marking initial mutation, this mutation may be “protected” against change by being part of a proper name (our language-independent hypothesis).

As regards the first point, proper names are found to mutate after feminine nouns four times out of four in OC-W—a small number of observations, three of which contain the feminine noun *nos* ‘night’: e.g. *nos F<sub>erchar</sub>* ‘Wednesday night’, *nos W<sub>ener</sub>* ‘Friday night’. In CC-W, mutation of the attributive noun after a feminine head is found in three cases out of eight: in *nos W<sub>ener</sub>* ‘Friday night’, once, and twice after *f. Sir* ‘county’ in the place name *Sir F<sub>ôn</sub>* ‘Anglesey’ (twice)—hence only in fairly lexicalized cases.

As regards the second point, i.e. gender-marking mutations within proper names, the only relevant observations were in CC-W, namely the two aforementioned occurrences of *Sir F<sub>ôn</sub>*.

However, once these occurrences are removed, we are left with a small number of observations, on which a meaningful generalization could hardly be based (Table 6.8): in other words, our data does not permit to assess whether gender-marking mutation with attributive nouns is still really productive.

Table 6.8: Attributive noun agreement with feminine controllers, OC-W vs. CC-W, excluding semantic agreement and proper names)

|            | Fem. (OC) | Fem. (CC) |
|------------|-----------|-----------|
| Agreeing   | 5         | 3         |
| Total      | 8         | 5         |
| % agreeing | 62.5      | 60        |

#### 6.2.2.4 Agreement with demonstrative adjectives

The use of inflected demonstrative adjectives in the Welsh corpus turned out to be quite infrequent in both components and limited to a few adverbial phrases (mostly temporal, with controllers like *adeg* ‘time’) which appear to be of relatively high frequency. The most frequent collocation in the Welsh corpus involves the feminine noun *adeg* (*adag*) ‘(period of) time’ and the distal demonstrative adjective, a combination which occurs six times altogether, three in each component. All other noun–demonstrative combinations appear only once each and all involve the neuter form of the demonstrative.<sup>8</sup>

The feminine proximal form (*h*)*onno* is found only once in OC-W and twice in CC-W; elsewhere, the historically neuter form *hynny* is used (cf. §5.7.1.7), as in example (5).

(5) Welsh (Thomas & Thomas, 1989: 92 l. 33)

O’           (hyn)ny’n    lot o bres   adag   hynny  
be.IMP.F.3 that.NT=PTC lot of money time(F) that.NT

‘That was a lot of money at the time’

Other relatively frequent collocations (occurring three times each in CC-W) involve masculine *pryd* ‘(period of) time’ (with the distal demonstrative) and feminine *blwyddyn* ‘year’ (with the proximal demonstrative), but these appear with a neuter form of the demonstrative adjective rather than the expected gendered form.

The use of inflected demonstrative adjectives with a masculine controller is recorded only once in OC-W, but the form in which it appears is the neuter rather than the masculine (6).

(6) Welsh (Thomas & Thomas, 1989: 118 l. 40)

(y)r    pren       ’yunny  
ART.M wood(M) that.NT

‘that wood’

As regards feminine controllers, six occurrences are found: four with a neuter agreement form, one with a masculine form (7), and one with the

<sup>8</sup>There were only two exceptions: (9c) below, which appears with the expected feminine form, and (7), in which a masculine demonstrative is found with a feminine noun.

expected feminine demonstrative (8)—which, remarkably, is not a case of semantic agreement.

(7) Welsh (Thomas & Thomas, 1989: 108 l. 8)

odd y platform yn packed y noswith  
 be.IMPF.3 ART.M.SG platform(M).SG PTC packed ART evening(F).SG  
 'nw  
 that.M.SG

'the platform was packed that evening'

(8) Welsh (Thomas & Thomas, 1989: 97 l. 10)

adeg 'onno  
 time(F) that.F

'that time'

In CC-W there are six occurrences of demonstrative adjectives with masculine controllers, all of which take the neuter form, and eight with feminine controllers, only three of which appear with the expected feminine form; none of these three are instances of semantic agreement, as shown in (9).

(9) Welsh

a. (beti.cha 183)

yr adeg honno  
 ART time(F) that.F

'that time'

b. (beti.cha 561)

(y)r adeg honno  
 ART time(F).SG that.F.SG

'that time'

c. (jonsi.cha 1092)

mi oedd o (y)n canu yna noson honno  
 PTC be.IMPF.3 3SG.M PTC sing.VN there night(F).SG that.F.SG

'he sang there that night'

The remaining five all take neuter forms of the demonstrative adjective.

No instances of semantic agreement between a demonstrative adjective and either a masculine or feminine noun were observed, nor was any proper name found which included a demonstrative adjective.

### 6.2.2.5 Agreement with numeral adjectives

Like the demonstrative adjectives, numeral adjectives were found to occur rather infrequently in the Welsh corpus. It should be recalled from §5.7.1.8 that agreement is marked by means of initial mutation after *un* ‘one’ and by internal vowel shift with numerals from ‘2’ to ‘4’.

In OC-W five instances of *un* modifying a masculine noun were recorded, all of which behave as expected in triggering no mutation, and only one occurrence with a feminine noun, which triggers the expected mutation (*un \_waith* ‘one time, once’, a high-frequency collocation which is also found as one word in dictionaries).<sup>9</sup> As regards the numerals from ‘2’ to ‘4’, there were two instances with masculine controllers, both masculine as expected, and 11 with feminine controllers, ten of which (90.91%) showed the expected feminine form and one the masculine form.

In CC-W, ten instances of *un* with masculine nouns were counted, with the following word unmutated as expected, and only one with a feminine noun, mutating the following word as expected. 11 occurrences of inflected numerals (‘2’–‘4’) with masculine controllers were recorded, all of which showed the expected masculine form, and 17 with feminine controllers, of which 13 (or 76%) showed the expected feminine form.

To recap, *un* occurs with feminine nouns only once in each component; there were more observations for the numeral series ‘2’–‘4’, and a decrease in agreement rate with feminine controllers is observed in CC-W (Table 6.9); however, given the sample size at hand, the difference in agreement rate between the two components tested as statistically not significant ( $p = 0.6195$ ).

The impact of semantic agreement with feminine controllers, represented by two occurrences in CC-W, was deemed negligible.

---

<sup>9</sup>Cf. Meurig Evans & Thomas (2007: s.v. *unwaith*).

Table 6.9: Agreement of numeral adjectives ('2'–'4') with feminine controllers, OC-W vs. CC-W

|              | Fem. (OC) | Fem. (CC) |
|--------------|-----------|-----------|
| Agreeing     | 10        | 13        |
| Total        | 11        | 17        |
| %            | 91        | 76        |
| $p = 0.6195$ |           |           |

### 6.2.3 Consequences for a model of gender assignment

The psycholinguistic evidence reviewed in §4.4 suggests that grammatical gender information is stored lexically (lexical hypothesis), possibly as the form of a closely associated modifier, like the article, rather than as an abstract feature; it was also pointed out that salient multi-word units appear to be lexically stored just as single-word units, and that for Insular Celtic in particular this might be favoured by the presence of initial mutation phenomena, which might cement further the article–noun sequence and lend it salience as a unit.

This is supported by the finding that, within the noun phrase, the article is the only target with which grammatical gender agreement is regularly found in contemporary Irish and Welsh. Grammatical agreement with attributive adjectives and nouns is not consistently observed and the use of non-mutated agreement forms with feminine controllers is frequent. Feminine agreement forms of the attributive adjective with grammatically feminine nouns occur at chance level in Irish; they appear more regularly in Welsh, but still significantly less often than with the article.<sup>10</sup>

As observed in §4.4, a weaker version of the lexical hypothesis would be that gender information is recorded in the lexicon not for all nouns but only for relatively more frequent nouns (cf. Butterworth, 2004; Stemberger & MacWhinney, 2004); under the weaker hypothesis, the grammatical gender of nouns that the speaker has only rarely if ever heard may still be determined based on a set of observed regularities (assignment rules). In order to

<sup>10</sup>It is important to bear in mind that the influence of semantic agreement on the above findings is negligible; statistical analyses conducted after excluding all nouns with biologically-gendered referents as controllers of agreement showed that their incidence was of no statistical consequence.

choose empirically between the stronger and the weaker version of the lexical hypothesis, one could consider a representative sample of nouns, ranked for frequency, and observe whether the rate of agreement is consistent across the board or correlates with frequency. To do so, however, a much larger corpus would be necessary.<sup>11</sup>

#### 6.2.4 Consequences for a model of gender agreement

The first observation to be made is that the context-sensitivity of grammatical gender agreement is a well-known typological feature of gender systems: a decrease in grammatical agreement rates as one moves from the attributive to the pronominal domain, in particular, is predicted by Corbett's Agreement Hierarchy (see §4.5.3): when a language allows semantic agreement alongside syntactic agreement, the former is least likely to be found within the noun phrase (attributive position) and most likely to be found with pronouns; other target types, such as predicative elements and the relative

---

<sup>11</sup>Jeffrey Kallen (p.c.) further suggests that the lexical and the processual hypothesis enable different predictions:

if gender is entered in the lexicon, then a speaker cannot change the gender of a noun and any error reflects either a lack of knowledge or a performance error. But if gender can be generated by the speaker, then what looks like an error could be an idiosyncratic change, and the accumulation of idiosyncratic changes over time or within social groups could account for variation and linguistic change. [J. Kallen, p.c.]

(I assume that a performance error under the lexical hypothesis would be one caused by problematic retrieval and that "idiosyncratic changes" under the processual hypothesis refers to the idiosyncratic modifications of one or more rules in the gender-assigning algorithm.) Therefore, choosing one hypothesis over the other would have crucial implications for a model of language change.

In response to this suggestion, I would argue that errors due to lack of knowledge or performance errors may also occur in the processual model and idiosyncratic changes may also occur in the lexical model. In the processual model, errors due to lack of knowledge could be observed if a speaker had imperfect knowledge of the relevant rules; performance errors are also possible, precisely because a process is involved and something might go wrong in the application of the gender-assigning algorithm. As regards the idiosyncratic changes, they should be equally possible under either hypothesis: lexically-encoded gender is language-specific and learned under the lexical hypothesis, and so are assignment rules under the processual hypothesis; therefore, if the idiosyncratic modification of an assignment rule is possible, so is the idiosyncratic modification of the lexical gender of a noun.

pronoun, occupy intermediate positions.

The findings relative to anaphoric agreement, and why contemporary Irish and Welsh can be said to have developed a semantic system, will be discussed in the next chapter; here it should be observed that the Agreement Hierarchy does not predict that there may be a further position to the left of the attribute's for determiners, and Dikker (2004: 39) explicitly includes the latter "under the attributive agreement targets" in her FDG-based model of agreement. In other words, a mismatch in agreement rates between the article and the adjective is a typologically uncommon occurrence. From a theoretical point of view, Dikker's model (*ibid.*) interprets the typological possibilities summarized in the Agreement Hierarchy in terms of syntactic structure: the more phrasal borders intervene between the agreement controller and a particular target, the more likely it is "for semantic information to take the lead over syntactic information" in the selection of the appropriate agreement form; to add a further position to the hierarchy just to the left of the attribute would entail to posit a further type of phrase which contained just the noun and the determiner, but not the attribute.

The problem posed by our data can therefore be summarized as follows: if feature copying is all there is to gender agreement in Irish and Welsh, and if the syntactic distance between the noun and the article is the same as between the noun and the attribute, then we would not expect to observe any difference in agreement rate between the article and the attribute. To account for the observed difference, one should hypothesize either that there is more than feature copying to agreement within the noun phrase, or that the article and the noun are syntactically closer to each other than the noun and the attribute.

Under the latter hypothesis, the article and the noun are separated by the rest of the noun phrase by a syntactic boundary, as shown in (10).



In principle, the intermediate constituent could in principle be headed by the article and thus be a Determiner Phrase (Dp), or be a projection (N')

of the noun as in X-bar theory (cf. e.g. Carnie, 2008: ch. 7). The former option would imply an extension to the current FDG model, which does not include Determiner Phrases, and would be problematic semantically in that the determiner in a Determiner Phrase is understood to have scope over the entire Noun Phrase, i.e. over the Adjective Phrase as well. In theories which do include a Determiner Phrase (e.g. Abney, 1987), the latter takes a Noun Phrase as a complement and includes it, as in (11).



Syntactically, we would have to justify the fact that Np is headed by a Dp, by positing a non-headed rule like  $\text{Np} \rightarrow \text{Dp} (\text{Ap})$ .<sup>12</sup>

The other possibility is for [Art N] to be a projection of N, i.e. an N' in X-bar-theory terms. Semantically this is still counterintuitive, and syntactically we would have to posit a restriction on the occurrence of Adjective Phrases in N'. This of course would be problematic, since one of the reasons why N's were postulated in X-bar theory is precisely to account for *one*-substitution examples like (12), where *one* can stand for an N' comprised of a noun and an Adjective Phrase (cf. Carnie, 2008: 114).

(12) I want the big [plastic bag]<sub>N'</sub>, not the small [one].

The positing within the Noun Phrase of an intermediate [Art N] constituent which cannot contain Adjective Phrases is therefore problematic for both semantic and syntactic theory. The alternative hypothesis that agreement in the Irish and Welsh noun phrase involves more than just feature copying is considered in what follows.

<sup>12</sup>Presumably, in order to do so it would be necessary to posit another rule (like  $\text{Dp} \rightarrow \text{Determiner N}$ ) to prevent the generation of noun-less Determiner Phrases; without that, the model could generate Noun Phrases comprised of a determiner alone.

### 6.3 Modelling noun-phrase agreement in Irish and Welsh

In the light of the facts summarized in §6.2, a model of gender agreement in Celtic should be able to explain the fact that agreement, in contemporary speakers, is not marked consistently on all types of target within the noun phrase.

In FDG, agreement is described as a process whereby “information properly pertaining to a single element of the Clause (or of the Phrase) is copied to one or more other elements” (Hengeveld & Mackenzie, 2008: 350, 394).<sup>13</sup> This is referred to as the feature-copying model of agreement. The terminology employed by Hengeveld & Mackenzie is broad enough to cover both syntactic agreement (which depends on grammatical information) and semantic agreement (which depends on semantic information), as discussed in §§4.4–4.5. The characterization of agreement as the copying of information or features captures its asymmetric nature, based on the displacement of some information, which appears to be a fairly consensual notion in current models (Corbett, 2006: 114–116). In FDG, agreement rules are applied by the Morphosyntactic Encoder, after all the slots in the clausal template have been filled (either with lexical material or with placeholders): this ensures that all agreement controllers and all agreement targets are already in place and available (Hengeveld & Mackenzie, 2008: 350f.). The final PL representation of the agreement target (initially an underspecified abstract form) depends on features that are inherent to the agreement controller; agreement rules copy these features to the agreement target and the Phonological Encoder then selects the appropriate form of the latter based on the copied features (ibid., p. 352).

Agreement between the noun and the attributive adjective, as in example (13), may be represented as in (14). Copied features in the ML representation are shown in angled brackets, e.g. <f> for feminine (Hengeveld &

<sup>13</sup>A further distinction is made at Clause level between *argument agreement* and *operator agreement* (Hengeveld & Mackenzie, 2008: 350–352), but this is not relevant at the phrasal level and will not therefore be considered here.

Mackenzie, 2008: 351).

(13) Irish (Cois Fhairrge dialect, Ó Siadhail, 1989: 113)

móin                    dhubh  
peat(F).NOM.SG    black.NOM.F.SG

‘black peat’

(14) IL: (-id R) (assuming the noun phrase is used referentially)

RL: ( $^m x_i$ : [( $f_i$ : móin<sub>Nf</sub> ( $f_i$ )) ( $x_i$ ): [( $f_j$ : dubh<sub>A</sub> ( $f_j$ )) ( $x_i$ )] ( $x_i$ ))

ML: (Np: [(Nw: /mo:N'/ (Nw)) (Ap: (Aw: dubh (Aw)) (Ap))<f>]  
(Np))

PL: (PP<sub>i</sub>: [(PW<sub>i</sub>: /mo:N'/ (PW<sub>i</sub>)) (PW<sub>j</sub>: /yu:/ (PW<sub>j</sub>))] (PP<sub>i</sub>))

The representations in (14) reads as follows: at the Interpersonal Level, the element is represented as non identifiable (-id) and used referentially (R). At the Representational Level, an entity characterized as a mass ( $^m x_i$ ) is represented, of which a Property ( $f_i$ ) lexically specified by *móin* is predicated. Another Property ( $f_j$ ) is predicated of  $x_i$ ; this second Property is lexically specified by an adjective (*dubh<sub>A</sub>*). The RL representation further indicates that *móin* is a Lexeme belonging to the class of feminine Nouns (Nf). At the Morphosyntactic Level, *móin dhubh* is represented in terms of syntactic constituency, i.e. as a Noun Phrase (Np) made up of a Noun Word (Nw: *móin*) and an Adjective Phrase (Ap), the latter containing an Adjective Word (Aw: *dubh*). The Adjective Phrase is represented as having the feature <f> (feminine), which pertains to the Lexeme *móin*, copied on it. The feature must be copied onto the Adjective Phrase as a whole, since (in Irish as in Welsh) the initial mutation, where applicable, applies to the leftmost word in the phrase, which may or may not coincide with its head (cf. §5.7.1.5).

As a depth-first model, FDG stipulates that the phonological form of an element is finalized as soon as all the required information is available. The phonological form of the Nw can already be determined at this stage and is therefore introduced; the form of the Aw, on the other hand, cannot, as there is another encoding operation to be carried out, namely the realization of the appropriate agreement form required by the syntactic feature feminine (<f>). After this further operation is carried out, the final representation

is the PL one: the unit is represented as a Phonological Phrase ( $PP_i$ ) made up of two Phonological Words ( $PW_i$  and  $PW_j$ ). The final phonological representation of the adjective is now available (/yʉ:/).

To model agreement as feature copying can handle the initial mutation of the adjective, but some clarification are needed as regards the attributive noun. Nouns have values of their own for gender and number. What does it mean then for the features of the controller to be copied to an attributive noun?

In example (15), the head of the attributive noun phrase [*llyfrau Cymraeg*] is clearly marked as plural. Moreover, its inherent lexical gender is masculine, although gender agreement is not marked in plural noun phrases.

(15) Welsh (Borsley et al., 2007: 192)

siop            [lyfr-au Cymraeg]  
shop(F).SG book-PL Welsh

‘Welsh-language book shop’

Since the initial mutation of *llyfrau* > *lyfrau* is due to the gender and number features of *siop*, then *llyfrau* would appear to possess at ML double gender and plural values and to be marked for number twice: as a plural noun by the plural ending *-au* and as agreeing with a singular (feminine) noun by the initial mutation. But what are the features <f.sg>, pertaining to the noun *siop*, copied to? As discussed in §5.7.1.5, Adjective phrases (Aps) in Welsh mutate as wholes, so any relevant feature must be copied onto the Ap; the first element of an Ap marked as <f.sg> undergoes Soft Mutation. The same is seen here with the first element of the Noun Phrase undergoing mutation.

There appear to be two possibilities: either the features in question are copied to the modifier Np directly or to a specially posited Ap in which the Np is embedded. Under the first hypothesis a dependent Np may attract and host features pertaining to its head just like an Ap; once these features are copied to the Np, its first element undergoes Soft Mutation. The problem is of course that the same does not happen to all dependent Noun Phrases: in examples like (16), as already discussed in §5.7.1.6, the embedded Np (*mab chwaer y meddyg*) neither attracts nor hosts the features <f.sg> pertaining to its syntactic head, *siop* (if it did, its first element would mutate:

*mab* > *fab*). Argument Noun Phrases, unlike attributive ones, do not show agreement with their head.

(16) Welsh (Borsley et al., 2007: 184)

[siop [mab chwaer y meddyg]<sub>Np</sub>]<sub>Np</sub>  
shop(F).SG son sister ART doctor

‘the shop of the doctor’s sister’s son’

Structurally, we could explain this difference by proposing that the head’s features can only be attracted and hosted by an Ap, and that an attributive Np is embedded into an Ap. At first blush, it could be thought that this Ap is headed by an Adjectival Word (Aw) resulting from a process of derivation, described in FDG as an operation that “adapt[s] the form of a lexeme that has been inserted into an underlying semantic slot it was not designed to occupy, and produces the appropriate word form” (Hengeveld & Mackenzie, 2008: 229). The specific type of derivation involved here would be conversion, as no morphological change to the base form is required. This can be represented as in (17), where the resulting structure is an Aw.

(17) (1x: llyfr<sub>Nm</sub> (x)) → (**Aw**: llyfr (**Aw**))

But there is a problem with this solution, since *llyfrau* in example (15) is clearly an Nw bearing plural inflection. Whereas conversion as just described should be considered an interface process adapting a Lexeme (an RL primitive) to a specific morphosyntactic slot, the conversion of *llyfrau* involves an Nw-to-Aw conversion and would thus seem to take place entirely within the Morphosyntactic Level, as formally represented in (18).

(18) (Nw: llyfr-pl (Nw)) → (**Aw**: (Nw: llyfr-pl (Nw)) (**Aw**))

Furthermore, in (15) the Ap embeds a complex Np (*llyfrau Cymraeg*) rather than just an Nw. This would require a process of Np-to-Aw conversion which might be formally represented as in (19).

(19) (Np<sub>j</sub>: [(Nw<sub>j</sub>: llyfr-pl (Nw<sub>j</sub>)) (Ap<sub>j</sub>: (Aw<sub>i</sub>: /kəmra:ig/ (Aw<sub>i</sub>)) (Ap<sub>j</sub>))] (Np<sub>j</sub>))  
→ (**Aw**<sub>j</sub>: (Np<sub>j</sub>: [(Nw<sub>j</sub>: llyfr-pl (Nw<sub>j</sub>)) (Ap<sub>j</sub>: (Aw<sub>i</sub>: /kəmra:ig/ (Aw<sub>i</sub>)) (Ap<sub>j</sub>))] (Np<sub>j</sub>)) (**Aw**<sub>j</sub>))



of the two Properties  $f_k$  and  $f_l$ , expressed by the Lexemes *llyfr* (a noun) and *Cymraeg* (an adjective), respectively. In the ML representation, after morphosyntactic encoding,  $f_j$  is expressed by  $Np_j$ , headed by  $Nw_j$  (*llyfr-pl*), where *pl* represents a Morphosyntactic Operator introduced to mark plurality (expressed by  $mx_j$  in the RL representation). The Aw *Cymraeg* requires no further encoding and is therefore found in its final phonological representation (/kəmra:ig/) at this stage.

The final representation of  $Nw_j$ , *lyfrau* /livrai/, is produced by the Phonological Encoder and depends on the presence of the Morphosyntactic Operator *pl* and the features <f.sg>. The former adds the ending *-au* /ai/, and the latter are responsible for the initial mutation (*ll- /ɫ/ > l- /l/*).  $Np_j$  (*lyfrau Cymraeg*) therefore modifies *siop* like an attributive adjective and agrees with the modified noun like an attributive adjective; as just explained, this is accounted for by hypothesizing that it is incorporated into an Adjectival Word ( $Aw_j$ ) heading  $Ap_i$ .

This might explain the difference between Irish and Welsh in terms of attributive-Np mutation. As shown in §5.4.1.6, attributive nouns in traditional varieties of Irish are lenited to agree with their head, but more complex attributive noun phrases, e.g. a noun modified by an adjective, are not; this was illustrated with the following example, where the Np *gaoithe móire* is not mutated to *ghaoithe móire*.

(23) Irish (Ó Siadhail, 1989: 121)

|                 |                 |                 |
|-----------------|-----------------|-----------------|
| oíche           | [gaoithe        | móire]          |
| night(F).NOM.SG | [wind(F).GEN.SG | great.GEN.F.SG] |

‘a night with a high wind’

The difference might be described in structural terms by saying that while Welsh allows both Nw-to-Aw and Np-to-Aw incorporation, Irish allows the former but not the latter.

It is also necessary to clarify what is involved in what we have referred to in the preceding chapters as article–noun agreement. As we have seen in chapter 5, agreement between the article and the noun is marked on both the dependent and the head in Goidelic (where the article has two forms, /ən/ and /nə/, and the noun is mutated), only on the head in Brythonic (where only noun mutation marks gender agreement). Furthermore, the

form of the article may depend on the initial segment of the following word. In Welsh, for instance, and simplifying somewhat, 'r occurs after a vowel when the article cliticizes to the preceding word; otherwise, *yr* occurs before a vowel and *y* before a consonant. This is illustrated by examples (24a), (24b) and (24c), respectively.

(24) Welsh

|                |     |             |       |           |     |
|----------------|-----|-------------|-------|-----------|-----|
| a. gyda'r      | ci  | b. yr       | afon  | c. y      | ci  |
| with=ART       | dog | ART         | river | ART       | dog |
| 'with the dog' |     | 'the river' |       | 'the dog' |     |

If *y* is the bare lexical form of the article, as is commonly assumed (Watkins, 1993: 313), then its final phonological representation will be available at the Morphosyntactic Level in the case of (24c), but only at the Phonological Level in the case of both (24a) and (24b). If on the other hand as the bare lexical form is *yr*, then the article will be finalized at the Morphosyntactic Level in (24b) and only at the Phonological Level in the other two examples.

In Irish, the situation is more nuanced in that stylistical factors and formality, as well as the dialect, may affect the realization of the article, specifically as regards the realization of the final /n/ (cf. Ó Curnáin, 2007b: 1330–1332). What seems to be regular, however, is the deletion of /n/ when the article occurs between two consonants within a larger constituent (cf. Ó Curnáin, 2007b: 1330 for a Connemara variety; and Ó Sé, 2000: 166f., 176–197 for a Munster variety): cf. *an sagart* (25a) vs. *bean a(n) tí* (25b).

(25) Irish (Ó Sé, 2000: 184, 191)

|              |        |                 |         |       |
|--------------|--------|-----------------|---------|-------|
| a. an        | sagart | b. bean         | a'      | tí    |
| ART          | priest | woman           | ART.GEN | house |
| 'the priest' |        | 'the housewife' |         |       |

Assuming that the bare lexical form of the article is *an*, this is identical to its final phonetic representation in (25a), but not in (25b); so the article would be finalized at the Morphosyntactic Level in the former case, but only at the Phonological Level in the latter. This state of affairs can be represented as in (26); as usual, words presented orthographically symbolize phonologically non-finalized representations.

(26) Irish *an fear cliste* 'the clever man'

IL: (+id R) (assuming the noun phrase is used referentially)

- RL: ( $1x_i$ : [( $f_i$ : fear<sub>Nm</sub> ( $f_i$ )) ( $x_i$ ): [( $f_j$ : cliste<sub>A</sub> ( $f_j$ )) ( $x_i$ )] ( $x_i$ ))
- ML: (Np: [(Gw: an<nom.m.sg> (Gw)) (Nw: /f'ar/ (Nw)) (Ap: (Aw: cliste (Aw)) (Ap))<nom.m.sg>] (Np))
- PL: (PP<sub>i</sub>: [(PW<sub>i</sub>: /ənf'ar/ (PW<sub>i</sub>)) (PW<sub>j</sub> /k'l'is't'ə/ (PW<sub>j</sub>))] (PP<sub>i</sub>)) [<sup>15</sup>]

The copying of morphosyntactic features onto the article, indicated in (26) as a Grammatical Word (Gw), can only account for the selection of its own form (*an* or *na*), but not for the (syntactic) mutation of the noun, since to model article–noun agreement uniquely in terms of feature copying, we would have to assume that once the relevant features have been copied to the article’s placeholder, a morphophonemic trigger is attached to the article at the Phonological Level, as in (27).

- (27) ML: an<nom.f.sg>  
 PL: /əN<sup>L</sup>/

But as discussed in §5.3, there are insurmountable problems with morphophonemic accounts of initial mutation like this.

On the other hand, the noun only mutates when preceded by the article, i.e. its mutation is not triggered just by a specific bundle of morphosyntactic features (e.g. <nom.f.sg>) interacting with definiteness: in the absence of the article, the combination of all these features does not trigger mutation, as shown in examples (28a) and (28b), where both *Cáit* and *bean* are nominative feminine singular nouns heading definite noun phrases yet neither is lenited.<sup>16</sup>

(28) Irish

- a. Tá Cáit / \*Cháit in-a múinteoir anois.  
 be.PRS Cáit(F).NOM.SG in-POSS:3SG.F teacher now  
 ‘Kate is a teacher now.’

<sup>15</sup>Unless stressed for emphasis, the article is a clitic in both Irish and Welsh (Mac Eoin, 1993: 119; Sadler, 1997), and as such here it forms a phonological word with the following noun (Dixon & Aikhenvald, 2002: 13; Hengeveld & Mackenzie, 2008: 446).

<sup>16</sup>In both Irish and Welsh an embedded definite noun phrase causes the matrix noun phrase to be interpreted as definite, hence the head of the latter need not, and cannot, be preceded by an article: cf. in examples (28b) and (29b), respectively, [bean [an tÍ]] vs. \*[an bhean [an tÍ]], [merch [y brenin]] vs. \*[y ferch [y brenin]].

- b. Bhí bean / \*bhean an tí deas.  
 be.PST woman(F).NOM.SG ART.GEN.M.SG house(M).GEN.SG nice  
 ‘The housewife was nice.’

The same applies to Welsh, as shown in (29a) and (29b).

(29) Welsh

- a. Mae Meri / \*Feri'n gweithio.  
 be.PRS Meri(F).SG=PTC work.VN  
 ‘Mary is working.’
- b. mae merch / \*ferch y brenin yn cael  
 be.PRS daughter(F).SG ART.M.SG king(M).SG PTC get.VN  
 ei galw'n Infanta.  
 POSS:3SG.F call.VN=PTC Infanta  
 ‘The daughter of the king is called Infanta.’

In both Irish and Welsh the effect of initial mutation can be the deletion of an initial consonant (e.g. /f/ in Irish under lenition, /g/ in Welsh under soft mutation, as seen in §§5.4 and 5.7; this means that a noun beginning in a consonant may surface as beginning in a vowel, which will in turn affect the form of the article. In other words, the mutated form of the noun must be available before the form of the article is finalized. Hannahs & Tallerman (2006: 781) therefore suggest that content words, like nouns, are inserted in the syntactic derivation before grammatical words, like the article. For instance, the derivation of Welsh *yr ardd* ‘the garden’ is modelled as in (30).

(30) Welsh *yr ardd* ‘the garden’ (Hannahs & Tallerman, 2006: 808)

|       | Syntactic structure:      | [Article | Noun] |
|-------|---------------------------|----------|-------|
| (i)   | insertion of content word | ---      | gardd |
| (ii)  | triggering of mutation    | ---      | ardd  |
| (iii) | insertion of article      | yr       | ardd  |

In (30), a gap (---) represents the placeholder for the article, whose phonological form is finalized only in the last step; in the FDG model, this corresponds to the Phonological Level.

However, we also need to assume that both steps (ii) and (iii) in (30) (triggering of mutation and insertion of article) take place at the Phonological Level. The alternative would be to say that the mutated form of the

noun is already available at the Morphosyntactic Level, a situation which could be represented as in (31).

(31) Welsh *yr ardd* ‘the garden’

IL: (+id R) (assuming the noun phrase is used referentially)

RL: ( $1x_i$ : [( $f_i$ : gardd<sub>Nf</sub> ( $f_i$ )) ( $x_i$ )] ( $x_i$ ))

ML: (Np: [(Gw: y<f.sg> (Gw)) (Nw: /arð/ (Nw))] (Np))

PL: (PP: [(PW: /ər'arð/ (PW))] (PP))

The problem with (31) lies in the implication that /arð/ is the basic form of the noun, which is untenable for morphosyntactic reasons (the mutated form would not appear if the word was cited in isolation) as well as historical reasons (we know that the mutated form is secondary as it derives from the grammaticalization of a sandhi phenomenon).

Nevertheless, all the morphosyntactic information needed to finalize the form of the noun (gender, number, case and whether there will be an article in the final utterance) is already available at the Morphosyntactic Level and can therefore be sent down to the Phonological Level, where the selection of the appropriate form of the article is possible because the form of the following noun can already be determined. Given an ML representation such as that in (32), the Phonological Encoder delays the finalization of the form of the article, “looks ahead” to ascertain the final representation of the noun, and then finally encodes the article.

(32) ML: (Np: [(Gw: y<f.sg> (Gw)) (Nw: gardd (Nw)) (Ap: (Aw: mawr (Aw)) (Ap))<f.sg>] (Np))

PL: (PP: [(PW<sub>i</sub>: /ər'arð/ (PW<sub>i</sub>)) (PW<sub>j</sub>: /vaʊr/ (PW<sub>j</sub>))] (PP))

It is important to notice that article–noun agreement (insofar as it is head-marked) crucially differs from noun–attribute agreement (dependent-marked): in terms of mutation type, as discussed in §5.3, both are instances of incorporated, or syntactic, mutations, but while the latter depends on feature copying, the former does not; instead, it depends (a) on a bundle of morphosyntactic features (i.e. gender, number and case) and (b) on the syntactic configuration of the final utterance (i.e. whether the article will be present or not).

Because agreement thus depends on two separate mechanisms, one for head marking and one for dependent marking, we can make sense of the differential agreement patterns revealed by the data and described above. If feature copying ceases to operate, it follows that attributive elements will no longer mutate in response to the gender of the noun; that is, they will cease to be agreement targets, in line with the observations already made apropos in chapter 5. Nouns can still be mutated after the article because their mutation does not depend on feature copying, whereas the selection of the article morpheme in the genitive singular in Irish, which still does, should cease: in fact, as has already been observed, genitive marking is increasingly rare in spoken Irish (cf. discussion in §5.4.3).<sup>17</sup>

The process whereby feature copying within the noun phrase ceases to operate is gradual. This means that certain speakers may retain feature copying in controlled speech, that is, as a feature of a higher register, and ignore it in lower registers; in this sense, adjective agreement may be, for those speakers, a stylistic choice. Other speakers may have ceased to apply feature copying altogether, but may retain some fixed mutations in particular phrases (idiomatic or otherwise high-frequency phrases). The following example (from CC-I) is a case in point.

(33) Irish (RnaG “Barrscéalta” 23/10/1997)

a. tá dainséar ann ... go gcaillfimis  
 be.PRS danger(M).NOM.SG there ... COMP loose.COND.1PL  
 cuid mhaith den rud  
 share(F).NOM.SG good.NOM.F.SG of.ART.SG rud(M).SG  
 so  
 DEM.PROX

‘we risk to lose a considerable amount of this thing’

b. ba mhaith linn ... go mbeadh  
 COP good with.1PL ... COMP beCOND  
 tuiscint maith sa  
 understanding(F).NOM.SG good.NOM.M.SG in.ART.PREP.SG  
 Ghaelainn acu  
 Irish(F).SG at.3PL

‘we would like for them to have a good comprehension of Irish’

<sup>17</sup>As shown in §A.1.2, the occurrence of gen. sg. *na* in CC-I seems to be linked to high-frequency and/or formulaic terms, which makes it difficult to argue in favour of its productivity as a marker of gender agreement.

*Cuid mhaith* ('a good amount', 'a great deal') occurs twice with the speaker of examples (33a) and (33b), *tuiscint maith* ('a good comprehension') just once. Both *cuid* and *tuiscint* are feminine, but *cuid mhaith* is a frequent collocation and as such is also found in standard dictionaries (cf. Ó Dónaill, 1977, s.v. *cuid*), whereas *tuiscint m(h)aith* is not. The presence of initial mutation on certain adjectives in speakers who would otherwise omit it might be explained as an application of the principle of lexical priority (§2.4): that is, the form *cuid mhaith* with the meaning 'a great deal' would be stored in the lexicon as a whole and this would prevent the occurrence of the productively combinatorial phrase *cuid maith*.

## 6.4 Summary

In this chapter I presented the findings of a corpus-based investigation into gender agreement marking in the Irish and Welsh noun phrase and demonstrated that an incremental model of speech production such as FDG, capable of handling processes that are not strictly linear, can adequately describe the mechanisms of agreement in this context and account for the differential agreement patterns that are observed in contemporary spoken varieties. The problem this differential pattern creates in the traditional model of agreement as the copying of morphosyntactic features is solved by distinguishing dependent-marked from head-marked agreement, with only the former depending on feature copying.

By doing so, we can explain the difference in agreement patterns without adding a further position for the determiner to the Agreement Hierarchy, an addition which is typologically unwarranted and theoretically difficult to justify: the lower agreement rates that characterize attributive elements within the noun phrase are interpreted as indicative of a gradual loss of the attributive elements as agreement targets. The consequences of this for pronominal agreement will be discussed in the next chapter.



## Chapter 7

# Anaphoric agreement and structure of the lexicon

### 7.1 Overview

In this chapter I will first outline the findings of my analysis of anaphoric agreement in gender with lexical antecedents<sup>1</sup> (§7.2); then, in §7.3, I will argue that the pattern of variation emerging from the analysis of the data and from the literature is not uncommon from a typological point view and, unlike the differential pattern of agreement found within the noun phrase (chapter 6), does not pose any particular problem for the existing FDG model of agreement. In the same section I will also address the question how the different types of information that might influence pronominal agreement are stored and how they might be represented in this model.

Having explained the variation observed in anaphoric agreement as a case of resemantization of pronominal agreement, I address the questions why this resemantization might have taken place and why in the direction that is observed, i.e. resulting in the overgeneralization of masculine rather than feminine forms: the relevant connections between the decline of gender agreement within the noun phrase and the resemantization of pronominal gender agreement are made in §7.4, where parallels with the historical evolution documented in other languages are also drawn, while in §7.5 I consider

---

<sup>1</sup>Anaphoric pronouns with non-lexical antecedents (e.g. declarative clauses) were not considered since they lack inherent grammatical gender. Also cf. §3.3.5 for further considerations on discourse anaphora.

the direction of the change.

A few considerations regarding the debate on the function of gender, based on the available evidence and the typological parallels that were identified, are offered in §7.6, followed in §7.7 by a final summary of the discussion on the lexical vs. discourse-contextual sources of gender-related information from the point of view of FDG.

## 7.2 Anaphoric gender agreement: the data

### 7.2.1 Irish

#### 7.2.1.1 The older component

As shown in Table 7.1, while almost all masculine pronouns in OC-I have masculine antecedents (98%), only 43% of the feminine pronouns have feminine antecedents; more than half of them (57 pronouns) have masculine antecedents. There are two reasons for this at first unexpected distribution: semantic agreement on the one hand and historically well-established idiosyncrasies on the other.

Table 7.1: Gender of pronouns and of their antecedents in OC-I

|                | Total | With masc. antecedent | With fem. antecedent |
|----------------|-------|-----------------------|----------------------|
| Masc. pronouns | 359   | 351 (98%)             | 8 (2%)               |
| Fem. pronouns  | 100   | 57 (57%)              | 43 (43%)             |

Semantic agreement accounts for ten of the 57 feminine pronouns in question. As we have seen in §5.4.2 above, masculine nouns sometimes refer to female referents, in which case agreement is normally syntactic within the noun phrase and semantic with anaphoric pronouns. Of the remaining 47 pronouns, 42 refer to a specific lexico-semantic category which includes masculine *bád* ‘boat’, *soitheach* ‘vessel’ and *púcán* (a type of fishing boat). According to Ó Siadhail (1984), this lexical category also includes a great variety of machines and modes of conveyance and the fact that feminine pronouns are used with grammatically masculine members of this category represents an idiosyncratic trait which all major dialects of Irish share.

In order to calculate agreement rates for this target, the 52 feminine pronouns just discussed were excluded from the count because their use with masculine antecedents is either due to semantic agreement or in line with traditional norms and cannot therefore be regarded as unexpected.

As Table 7.2 shows, pronominal agreement with masculine antecedents is significantly more frequent than with feminine ones: 99% vs. 84%, respectively ( $p < 0.0001$ ). If we exclude semantic agreement, as also shown in Table 7.2, the picture remains essentially unchanged and the difference still significant ( $p = 0.0016$ ). As can be observed, semantic agreement with feminine pronouns is only marginally represented in this component.

Table 7.2: Pronominal agreement in OC-I

|            | Overall      |      | Excl. sem. agr. |      |
|------------|--------------|------|-----------------|------|
|            | Masc.        | Fem. | Masc.           | Fem. |
| Agreeing   | 351          | 43   | 124             | 40   |
| Total      | 354          | 51   | 127             | 48   |
| % agreeing | 99           | 84   | 98              | 83   |
|            | $p < 0.0001$ |      | $p = 0.0016$    |      |

If we look at agreement rates by pronoun type (Table 7.3 and 7.4) we get similar proportions and we can observe that the between-gender difference in agreement rate is significant for both simple and prepositional pronouns. As regards possessive pronouns, a meaningful comparison is precluded by the very small number of observations available in relation to feminine antecedents.

Table 7.3: Pronominal agreement in OC-I by pronoun type

|            | Simple       |      | Prep.        |      | Poss.        |      |
|------------|--------------|------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 226          | 21   | 95           | 21   | 30           | 1    |
| Total      | 229          | 25   | 95           | 24   | 30           | 2    |
| % agreeing | 99           | 84   | 100          | 87.5 | 100          | 50   |
|            | $p = 0.0021$ |      | $p = 0.0074$ |      | $p = 0.0625$ |      |

Table 7.4: Pronominal agreement in OC-I by pronoun type, excluding semantic agreement

|            | Simple       |      | Prep.        |      | Poss. |      |
|------------|--------------|------|--------------|------|-------|------|
|            | Masc.        | Fem. | Masc.        | Fem. | Masc. | Fem. |
| Agreeing   | 82           | 20   | 32           | 20   | 10    | 0    |
| Total      | 85           | 24   | 32           | 23   | 10    | 1    |
| % agreeing | 96           | 83   | 100          | 87   | 100   | 0    |
|            | $p = 0.0407$ |      | $p = 0.0675$ |      |       |      |

### 7.2.1.2 The contemporary component

Whereas in OC-I a rather high proportion of feminine pronouns have grammatically masculine antecedents, no similar situation is found in CC-I. Here, as shown in Table 7.5, only four feminine pronouns (17%) have masculine antecedents. One is a straightforward case of semantic agreement with the masculine epicene *cara* ‘friend’ (the referent is a woman). Of the other three, one is a possessive with a masculine antecedent (*bus*).

Table 7.5: Gender of pronouns and of their antecedents in CC-I

|                | Total | With masc. antecedent | With fem. antecedent |
|----------------|-------|-----------------------|----------------------|
| Masc. pronouns | 229   | 212 (93%)             | 17 (7%)              |
| Fem. pronouns  | 23    | 4 (17%)               | 83% (19)             |

(1) Irish (RnaL “Seó an Phobail” 28/03/2007)

R: [...] tá mé ag breathnú ar an bus folamh seo. Céard atá cliste faoi sin?

S: 'Sea, bhuel b'fhéidir go bhfuil sé folamh ehm áit a bhfuil tú ag breathnú air ach tá/ tá seans ann nuair a théann sé ar aghaidh ar a turas go mbeidh líon mór daoine ag tógáil an bus sin nuair atá sé níos gaire don chathair

*Translation:*

R: [...] I'm looking at this empty bus. What's smart about that?

S: Yes, well perhaps it is empty where you see it but there is a chance as it goes on on its journey that a lot of people will take that bus

when it is closer to town

In example (1), speaker R first refers to the bus lexically, using the noun phrase *an bus folamh seo* ‘this empty bus’, characterized by masculine agreement (no lenition on either the noun or the adjective). Speaker S then refers to the same bus by means of three masculine pronouns (two simple and one prepositional: *sé . . . air . . . sé*) before referring to the same bus by the possessive *a*, followed by the unlenited noun *turas* ‘journey’ (masculine agreement would have been marked as *ar a thuras*).<sup>2</sup> Subsequent references to the same bus are made by S using a full noun phrase and a simple masculine pronoun. As regards the other two feminine pronouns with grammatically masculine antecedents, they are *scairdphaca* (a neologism calqued on the English *jet-pack*, which is attested in academic terminological dictionaries such as *focal.ie* as a masculine)<sup>3</sup> and masculine *stáisiún* ‘(TV) station’. In both cases the feminine pronoun is the simple form *í*. Little can be said about these examples; although we could speculate that the use of a feminine pronoun with *scairdphaca* is in line with the aforementioned norm (§7.2.1.1 above) that modes of conveyance are referred to by feminine pronouns regardless of their grammatical gender, in this text both speakers use masculine pronouns to refer to other vehicles (e.g. buses).

As regards masculine pronouns, 7% have grammatically feminine antecedents (Table 7.5). Among the antecedents we find

1. feminine *áit* ‘place’ (simple and prepositional pronouns, both RnaG and others);
2. feminine *argóint* ‘argument’ (simple pronouns, non-RnaG);
3. feminine *áilleacht* ‘beauty’ and *timpeallacht* ‘environment’ (simple pronouns, RnaG);
4. feminine *Gaelainn* ‘Irish language’ (simple pronouns, RnaG);

<sup>2</sup>It is also possible that *a* in the noun phrase *ar a turas* really is the article *an*, as it is normal in traditional Irish to drop the final /n/ when the article is both preceded and followed by a consonant (cf. §6.3). If this were the case, *ar a’ turas* would be in line with expectations. However, many contemporary speakers do not observe the “/n/-drop” norm. In particular, S does not: cf., in what follows, *ag tógáil an bus sin*.

<sup>3</sup><http://www.focal.ie>, accessed 1 December 2010.

5. feminine *cuid* ‘portion’ (possessive pronouns, non-Rnag) and *obair* ‘work’ (possessive pronouns, RnaG).

A few considerations may be made. As regards *áit*, masculine anaphora is very common in most Irish dialects and does not appear to be a recent innovation (see §5.4.2).<sup>4</sup> *Áilleacht* ‘beauty’ and *timpeallacht* ‘environment’ may be grouped together according to their termination: the derivative suffix *-acht* (/əxt/) normally yields feminine nouns, but non-palatalized consonant endings such as this are otherwise associated with the masculine (furthermore, there are masculine monosyllables in *-acht*, like *reacht* ‘right’, which might enter masculine compounds like *bunreacht* ‘constitution’). Moreover, *áilleacht* (like *argóint*) is abstract and non-concrete, a category which other studies have found to favour masculine anaphora (cf. Ó Curnáin, 2007b: 1259).

As illustrated in Table 7.6, the agreement rate is 98% with masculine controllers and only about 53% with feminine ones. If semantic agreement is excluded, the gap is even wider: 98% with masculine controllers as opposed to just 26% with feminine ones. In both cases, the difference is significant ( $p < 0.0001$ ).

Table 7.6: Pronominal agreement in CC-I

|            | Overall      |      | Excl. sem. agr. |      |
|------------|--------------|------|-----------------|------|
|            | Masc.        | Fem. | Masc.           | Fem. |
| Agreeing   | 212          | 19   | 145             | 6    |
| Total      | 216          | 36   | 148             | 23   |
| % agreeing | 98           | 53   | 98              | 26   |
|            | $p < 0.0001$ |      | $p < 0.0001$    |      |

There are only six feminine pronouns agreeing with inanimate antecedents, five of which come from the RnaG sub-component: four refer to the same lexical antecedent (feminine *móin* ‘turf’). Only one comes from

<sup>4</sup>Nevertheless, at least one occurrence of feminine anaphora with *áit* is recorded in the RnaG sub-component, possibly a case of hypercorrection. The speaker in question was a native Irish speaker from Corca Dhuibhne (Munster) and one of the oldest in this corpus; he had however emigrated to New Zealand in his youth after training as a school teacher and lived there for forty years before moving back to his native area.

the other sub-component, and has as its antecedent the feminine place name *Inis Oírr* (the name of an island).

Agreement rates with masculine antecedents are very high for all three pronoun types, as shown in Table 7.7: 99% for simple pronouns (98% without counting semantic agreement), 98% for prepositional pronouns (100% without counting semantic agreement)<sup>5</sup> and 92% for possessive pronouns (80% without counting semantic agreement).<sup>6</sup>

Table 7.7: Pronominal agreement with masculine controllers in CC-I by pronoun type

|            | Overall |       |       | Excl. sem. agr. |       |       |
|------------|---------|-------|-------|-----------------|-------|-------|
|            | Simple  | Prep. | Poss. | Simple          | Prep. | Poss. |
| Agreeing   | 141     | 60    | 11    | 104             | 37    | 4     |
| Total      | 143     | 61    | 12    | 106             | 37    | 5     |
| % agreeing | 99      | 98    | 92    | 98              | 100   | 80    |

Agreement rates with masculine controllers do not differ significantly between the two sub-components of CC-I (see Table 7.8): 99% in the RnaG sub-component as opposed to about 97% in the other one (or 100% vs. 96% if semantic agreement is not taken into account). As can be seen from Table 7.9, in the RnaG data only five out of 18 pronouns agree with their feminine antecedents if we consider only non-semantic agreement. In the non-RnaG component, on the other hand, only one out of five shows agreement with its antecedent (the aforementioned place name *Inis Oírr*).

### 7.2.1.3 Pronominal agreement in the Irish corpus

Feminine pronouns are not very frequent in CC-I in general, and are very rare with inanimate antecedents. Pronominal agreement with feminine antecedents is statistically significantly more frequent in OC-I than in CC-I:

<sup>5</sup>As regards prepositional pronouns, the data contain an instance of agreement clash (the noun is grammatically masculine but the referent is a female, and agreement is semantic rather than syntactic). When semantic-agreement data are discarded, all pronouns match the grammatical gender of their antecedents.

<sup>6</sup>As regards pronouns with feminine antecedents, 29 out of 36 are simple, which means that there are only seven non-simple pronouns with feminine antecedents; it was therefore impossible to meaningfully compare agreement rates across pronoun types.

Table 7.8: Pronominal agreement with masculine controllers in CC-I, RnaG vs. others

|            | Overall      |        | Excl. sem. agr. |        |
|------------|--------------|--------|-----------------|--------|
|            | RnaG         | Others | RnaG            | Others |
| Agreeing   | 127          | 85     | 70              | 75     |
| Total      | 128          | 88     | 70              | 78     |
| % agreeing | 99           | 97     | 100             | 96     |
|            | $p = 0.3066$ |        | $p = 0.2471$    |        |

Table 7.9: Pronominal agreement with feminine controllers in CC-I, RnaG vs. others

|            | Overall |        | Excl. sem. agr. |        |
|------------|---------|--------|-----------------|--------|
|            | RnaG    | Others | RnaG            | Others |
| Agreeing   | 18      | 1      | 5               | 1      |
| Total      | 31      | 5      | 18              | 5      |
| % agreeing | 58      | 20     | 28              | 20     |

including semantic agreement, agreement rate with feminine antecedents was 84% in OC-I and 53% in CC-I ( $p = 0.0018$ ); excluding semantic agreement, it was 83% in OC-I and just 26% in CC-I ( $p < 0.0001$ ).

Pronominal agreement with masculine controllers, on the other hand, is stable across the two components even when semantic agreement is excluded ( $p = 1$ ).

### 7.2.2 Welsh

Compared to Irish, Welsh has a wider range of anaphoric targets for gender agreement, including demonstrative pronouns, numeral pronouns and third-person singular personal pronouns.

Inflected demonstrative and numerals were already observed to be rather infrequent as nominal modifiers (§6.2.2), which they are also as anaphoric elements; therefore the findings for these two target types will be only briefly summarized here, before turning to the analysis of the personal pronoun data. The data relative to demonstrative and numeral anaphora are presented in full in Appendix B (see §§B.3–B.4).

In relation to demonstrative pronouns, 17 instances were recorded in OC-W and 21 in CC-W, including both pronouns with masculine antecedents and pronouns with feminine antecedents; syntactic gender agreement was consistently found in OC-W regardless of antecedent gender, whereas the CC-W data suggest that about half of feminine antecedents are resumed by masculine forms of the demonstrative. 14 instances of numeral pronouns were observed in OC-W and 10 in CC-W (one of which was considered uncertain); syntactic gender agreement was observed in both OC-W and CC-W, although anaphoric relations with a feminine antecedent were very few and in half of the cases semantically motivated.

As explained in §3.3.4, third-person singular pronouns were subdivided into three groups: simple pronouns, prepositional pronouns and possessive pronouns. However, unlike in Irish, simple pronouns can combine with both prepositional pronouns (e.g. *arno fo* ‘on.3SG.M 3SG.M’) and possessive pronouns (e.g. *ei<sup>SM</sup> ... o* ‘POSS:3SG.M ... 3SG.M’). When any such combination was found, each pronoun was counted separately since mismatching combinations may occur (e.g. *wrtho hi* ‘to.3SG.M 3SG.F’).

As regards the distinction between anaphoric and cataphoric pronouns (§3.3.4), no example of the latter was found in either component; the present discussion therefore only refers to anaphora.

### 7.2.2.1 The older component

In OC-W, pronominal agreement is virtually always found with both masculine and feminine antecedents (Table 7.10). As can be seen, the only instance in which agreement is not found, shown in (2), involves a possessive pronoun.

Table 7.10: Agreement with 3SG pronouns in OC-W, by pronoun type

|        | Simple |      | Prep. |      | Poss. |      | All types |      |
|--------|--------|------|-------|------|-------|------|-----------|------|
|        | Masc.  | Fem. | Masc. | Fem. | Masc. | Fem. | Masc.     | Fem. |
| Agr.   | 127    | 55   | 21    | 6    | 21    | 6    | 169       | 67   |
| Total  | 127    | 55   | 21    | 6    | 22    | 6    | 170       | 67   |
| % agr. | 100    | 100  | 100   | 100  | 95    | 100  | 99        | 100  |

*p* = 1

(2) Welsh (Thomas & Thomas, 1989: 87l. 13–14)

fydda          blawd          ar y          board,          a          'dda  
 be.IMPF.3SG flour(M).SG on ART.M.SG board(M).SG and be.IMPF.3SG  
 'r          iau          yn          cal          'i          roid          yn          fanno. A          'i  
 ART liver(M).SG PTC get.VN POSS:3SG.M put.VN in there and POSS  
 sglisho          wedyn, a          'i          dipio          fo          'n          y  
 slice.VN then and POSS:3SG.F dip.VN 3SG.M in ART.M.SG  
 blawd,          cyn          'i          ffrio          fo  
 flour(M).SG before POSS fry.VN 3SG.M

‘there was flour on the board, and the liver<sub>i</sub> was put there. And then it<sub>i</sub> was sliced and dipped in the flour, before it<sub>i</sub> was fried’

The antecedent, *iau* ‘liver’, is referred to using masculine pronouns such as  $(e)i^{SM}$  (after which *rhoid* is mutated to *roid*)<sup>7</sup> and *fo*, as well as two pronouns that cannot trigger any mutation on the following nominals (*sglisho* and *ffrio*, both beginning with unmutatable consonants). There is however one instance in which soft mutation could be triggered but is not, namely with *dipio*, which is the observation in question.<sup>8</sup>

Table 7.11: Agreement with 3SG pronouns in OC-W by pronoun type, excluding semantic agreement

|        | Simple |      | Prep. |      | Poss. |      | All types |              |  |
|--------|--------|------|-------|------|-------|------|-----------|--------------|--|
|        | Masc.  | Fem. | Masc. | Fem. | Masc. | Fem. | Masc.     | Fem.         |  |
| Agr.   | 56     | 21   | 14    | 1    | 14    | 4    | 84        | 26           |  |
| Total  | 56     | 21   | 14    | 1    | 15    | 4    | 85        | 26           |  |
| % agr. | 100    | 100  | 100   | 100  | 93    | 100  | 99        | 100          |  |
|        |        |      |       |      |       |      |           | <i>p</i> = 1 |  |

By comparing the figures in Table 7.11 with those in Table 7.10 it can be seen that over 60% of all feminine observations (41 out of 67) involve semantic agreement. If we exclude these, we can still observe that grammatical agreement with controllers of either gender is the norm in OC-W.

<sup>7</sup>The distinction between /r<sup>h</sup>/ and /r/ is phonemic in this speaker’s variety (North Wales).

<sup>8</sup>As was mentioned in §§5.4.1.1 and 5.7.1.4, non-native lexical elements have been observed to be somewhat more resistant to initial mutation than native roots; this could account for this particular observation, where the initial /d/ of the non-native morpheme *dip-* is not mutated.

## 7.2.2.2 The contemporary component

Table 7.12: Agreement with 3SG pronouns in CC-W, by pronoun type

|        | Simple       |      | Prep.         |      | Poss.        |      | All types    |      |
|--------|--------------|------|---------------|------|--------------|------|--------------|------|
|        | Masc.        | Fem. | Masc.         | Fem. | Masc.        | Fem. | Masc.        | Fem. |
| Agr.   | 172          | 33   | 24            | 6    | 32           | 6    | 228          | 45   |
| Total  | 174          | 36   | 24            | 8    | 35           | 8    | 233          | 52   |
| % agr. | 99           | 92   | 100           | 75   | 91           | 75   | 98           | 86.5 |
|        | $p = 0.0364$ |      | $p = 0.05645$ |      | $p = 0.2276$ |      | $p = 0.0017$ |      |

The figures in Table 7.12 show us that, on the whole, agreement with third-person singular pronouns in CC-W is regularly observed with masculine antecedents and is statistically significantly less frequent with feminine ones.

The appearance of feminine pronouns with masculine controllers was observed with both simple and possessive pronouns. In the case of the two simple pronouns, this is explained as due to semantic agreement, as shown in examples (3) and (4).

(3) Welsh (*jonsi.cha* 977–1000)

JON: mae (y)r hen gŵn fath â ni yndydyn?

LIM: ie mae (y)r hyna *fourteen and a half* dach chi (y)n gwybod?

JON: ydy o wir?

JON: (y)dy hynny (y)n hen i Labrador yndy?

LIM: *well* mae tua *hundred and thirty* i ni yndydy?

JON: ydy o wir?

LIM: yndy.

LIM: (dy)dy (ddi)m yn ryw *special* iawn.

LIM: oedd y *farrier* yn gweld hi bore (y)ma.

*Translation:*

JON: the dogs are like us, aren't they?

LIM: yeah, the oldest is fourteen and a half, you know?

JON: is he really?

JON: is that old for a Labrador, is it?

LIM: well, it's about a hundred and thirty for us, isn't it?

JON: is it really?

LIM: yes.

LIM: she's not very well.

LIM: the vet looked at her this morning.

(4) Welsh (jonsi.cha 2652–2653)

mi fydd gynnan ni feddyg yna hefyd i [/] i roi  
 PTC be.FUT.3 with.1PL 1PL doctor(M).SG there also to [/] to give.VN  
 ei<sup>SM</sup> farn o neu hi ar y sefyllfa  
 POSS:3SG.M opinion(F).SG 3SG.M or 3SG.F on ART situation(F).SG  
 yma  
 here

‘we will have a doctor there as well to give his or her opinion on this situation’

In (3), the underlined feminine pronoun *hi* refers to the noun phrase *(y)r hyna* ‘the oldest’. Per se, this noun phrase is not marked for gender, as its head is an adjectival form (*hyna* ‘oldest’) and begins with the unmutable segment /h/; however, the referent of *(y)r hyna* is understood to be a member of the set referred to in the previous turn by *cŵn* (> *gŵn*) ‘dogs’, the plural form of grammatically masculine *ci* ‘dog’. The other speaker (who ignores the sex of his interlocutor’s oldest dog) accordingly uses a masculine pronoun (*o*) in the next turn, while the feminine pronoun is used semantically by the owner of the female dog in question. In (4), the feminine pronoun is used to express the possibility that the doctor in question, referred to by the grammatically masculine *meddyg* (> *feddyg*), might be a woman.

The occurrence of feminine forms of the possessive pronouns with masculine antecedents, on the other hand, is not due to semantic agreement. In two cases, the pronoun appears to be feminine even though its antecedent is semantically masculine (examples (5) and (6)). In (5), the possessive *ei*, interpreted as feminine due to the lack of SM on the following word and thus glossed ‘POSS:3SG.F’, refers to *tadcu* (> *dadcu*), which is grammatically and semantically masculine;<sup>9</sup> while in (6) the feminine possessive could refer to either the author of the book (introduced by the noun phrase *athro o (y)r Alban o (y)r enw Bill Duncan* ‘a teacher(M) from Scotland called

<sup>9</sup> *Blacklist-* is not a native lexical item, which might explain why it is not mutated: cf. footnote 8 above (p. 208).

Bill Duncan’) or the book itself (m. *llyfr* ‘book’), so that in both cases a masculine form (which would mutate the initial of *darllenwyr*) would be expected.

(5) Welsh (*beti.cha* 299–304)

a oedd yn dadcu mae (y)n debyg yn  
 and be.PST.3 POSS:1SG grandfather(M).SG be.PRS.3 PTC probable PTC  
 sosialydd cadarn a yn ystod y *strike* mil  
 socialist(M).SG firm.M.SG and in period(F).SG ART strike thousand  
 naw dau chwech gaeth e ei *blacklist-o*  
 nine two six get.PST.3 3SG.M POSS:3SF.F blacklist-VN

‘and my grandfather seems to have been a firm socialist and during the strike of 1926 he was blacklisted’

(6) Welsh (*gaynor.cha* 513–514)

a mae  $o_i$  (y)n addo ar ôl darllen  $ei^{SM}_i$   
 and be.PRS.3 3SG.M PTC promise.VN after read.VN POSS:3SG.M  
 lyfr<sub>*j*</sub>  $o_i$  bydd  $ei_{i/j}$  darllenwyr yn teimlo (y)n  
 book(M).SG 3SG.M be.FUT.3 POSS:3SG.F reader.PL PTC feel.VN PTC  
 ddiflas  
 dull

‘and he promises that after they have read his book his/its readers will feel dull’

With feminine antecedents, masculine forms are found with all three pronoun types. In two cases (examples (7) and (8)), a masculine form is found even with semantically feminine antecedents.<sup>10</sup>

(7) Welsh (*beti.cha* 707)

gaeth Delta  $ei^{SM}$  \_eni  
 get.PST.3 Delta POSS:3SG.M bear.VN

‘Delta was born’

<sup>10</sup>In (7), Delta is a fictitious name the transcriber of the corpus uses to identify the person in question, a girl.

## (8) Welsh (ecdau.cha 384–385)

mae (we)di bod yn sgrifennu mewn er *lipstick* nodiadau  
 be.PRS.3 after be.VN PTC write.VN in FILL lipstick note.PL  
 iddo hun  
 for.3SG.M self

‘(she) has been writing notes for himself [intended: *herself*] in lipstick’

In (8) the referent of the reflexive pronoun is a named female singer, who had been the subject of the conversation between the two speakers.

If we exclude all pronouns with semantically male or female antecedents, we obtain the figures presented in Table 7.13.

Table 7.13: Agreement with 3SG pronouns in CC-W by pronoun type, excluding semantic agreement

|        | Simple       |      | Prep. |      | Poss.      |      | All types    |      |
|--------|--------------|------|-------|------|------------|------|--------------|------|
|        | Masc.        | Fem. | Masc. | Fem. | Masc.      | Fem. | Masc.        | Fem. |
| Agr.   | 33           | 2    | 7     | 0    | 12         | 2    | 52           | 4    |
| Total  | 33           | 5    | 7     | 0    | 13         | 3    | 53           | 8    |
| % agr. | 100          | 40   | 100   | –    | 92         | 67   | 98           | 50   |
|        | $p < 0.0012$ |      |       |      | $p = 0.35$ |      | $p = 0.0006$ |      |

In the case of masculine antecedents, 98% of the remaining pronouns still agree grammatically; as regards feminine antecedents, only eight pronouns were counted, and only four of those were feminine forms. Of the four masculine forms, three are simple pronouns and one a possessive, as shown in examples (9) and (10).

## (9) Welsh (beti.cha 1009–1010)

bydden i ddim heb y<sup>SM</sup> flwyddyn (yn)a [?] nawr  
 be.COND.1SG 1SG NEG without ART.F year(F) there [?] now  
 achos oedd lot o bethau da ambothtu fe  
 because be.IMP.F.3 lot of thing.PL good about 3SG.M

‘I wouldn’t want to miss that year now because there were a lot of good things about it’

(10) Welsh (gaynor.cha 65–75)

CAT: *well yeah* dyma ydy stori fawr y  
 well yeah here be.PRS.3 story(F).SG big.F.SG ART.M  
 bore (y)ma  
 morning(M) here

‘well yes, here’s this morning’s big story<sub>i</sub>’

CAT: ond er mae (y)n anodd i (y)r papurau (y)ndydy?  
 but FILL be.PRS.3 PTC difficult for ART paper.PL not.be.PRS.3

‘but it is difficult for the papers, isn’t it?’

GAY: *well* ydy mae  
 well be.PRS.3 be.PRS.3

‘well yes, it is’

GAY: o’n i yn gweld mae [/] mae o yn  
 be.IMPF.1SG 1SG PTC see.VN be.PRS.3 [/] be.PRS.3 3SG.M in  
 y papurau yn amlwg  
 ART paper.PL PTC obvious

‘I have seen it<sub>i</sub> is in the papers obviously’

CAT: yndy  
 be.PRS.3

‘yes’

GAY: mae (y)na rei (we)di (ei<sup>SM</sup>) roid o ar  
 be.PRS.3 there some after POSS:3SG.M put.VN 3SG.M on  
 y tudalen flaen  
 ART.M.SG page front

‘some have put it<sub>i</sub> on the front page’

*Wedi* ‘after’ does not trigger soft mutation per se (Williams, 1980: 185), hence the *rhoid* > *roid* mutation is interpreted as triggered by the oblique clitic, which is not phonetically realized in this context.<sup>11</sup>

<sup>11</sup> *(We)di* /ə/ *roid* > *(we)di roid*.

### 7.2.2.3 Agreement with third-person singular pronouns in the Welsh corpus

When agreement rates with third-person singular pronouns are compared across the two components, the only significant difference pertains to agreement with feminine antecedents. With masculine antecedents the agreement rate is essentially the same in both components (99% in OC-W vs. 98% in CC-W,  $p = 0.408$ ), while with feminine antecedents the agreement rate is 100% in OC-W as opposed to 86.5% in CC-W, a statistically significant difference ( $p = 0.0024$ ); excluding semantic agreement, it is 100% in OC-W and just 50% in CC-W; The agreement rate of 50% was calculated on the basis of just eight observations but is quite close to the agreement rates for the same variable which were recorded by Jones (1998a) in Rhosllannerchrugog for the three 40-plus age groups (§5.7.4.2) and which were between 54% and 59%.

### 7.2.2.4 Comparison to previous studies

I now turn to the comparison between my findings and those of earlier studies, namely Jones (1998a) and Thomas & Gathercole (2005)—cf. §5.7.4.2. This comparison can only be partial since Jones appears to only analyze non-semantic pronominal agreement,<sup>12</sup> while Thomas & Gathercole do not include a study of pronominal agreement in speech production but only in a comprehension task, the results of which cannot be compared to Jones's or my own. As a preliminary note, it should be noted that the data Jones obtained from the youngest speakers (group A, 20–39 years of age) in both communities stand out as remarkably different from those relative to the older age groups. This is evident in group A's 62% mutation rate for feminine nouns after the article in Rhosllannerchrugog, 29% in Rhymney, compared to rates higher than 90% in the former community and 80% in the latter for older cohorts; likewise for the attributive adjectives (67% in Rhosllannerchrugog's group A, 69% in Rhymney's, as opposed to mutation rates as high or higher than 90% and 80%, respectively, for the older age groups of the two communities). If we ignore Jones's group-A results, we see that SM of feminine nouns after the article is highly frequent in all samples (the two

<sup>12</sup>However, the agreement rates reported by Jones compare to those emerging from the present study when semantic agreement is excluded from the analysis.

from our corpus as well as those from Jones's and Thomas & Gathercole's studies): this means mutation rates as high or higher than 90% except in two groups (Rhymney's B and C, which are nevertheless as high as 84% and 86% respectively). SM of the attributive adjective after feminine nouns is always above the 70% level, being highest for OC-W and the Rhosllannerchrugog speakers (above 90%) and lowest for CC-W and the 31–50 year-old speakers in Thomas & Gathercole's study (77% and 71%, respectively). Finally, agreement between non-possessive pronouns and feminine antecedents shows a sharp contrast between OC-W (with its 100% agreement rate) and the other samples; as already noted above (§7.2.2.3), CC-W clusters with Rhosllannerchrugog's age groups B–D with agreement rates between 50% and 60%; in Rhymney, age groups A and B only use masculine pronouns in this context; of the other two age groups, group D (the over-75) stands out as quite conservative in respect of this variable, while the agreement rate of group C (the 60–74 year-olds) is just 30%.

All in all, the results of these two previous studies confirm the observation that pronominal agreement in more recent data differs from the traditional one; as Jones (1998a: 174) observes, “feminine pronoun replacement was found to be the least well retained gender-marking device”.

### **7.2.3 Agreement with feminine antecedents: analytic issues**

The paucity of feminine pronouns with inanimate antecedents in CC-I and CC-W is telling in and of itself; however, the question remains why there appear to be so few anaphoric relations with feminine antecedents in the contemporary components. If we consider that the older components are smaller in size than their contemporary counterparts (cf. §3.2), but yield many more instances of feminine pronouns with inanimate antecedents, it is clear that the answer to this question cannot simply be one of size: OC-W is about one third of the size of CC-W but contains 6.5 times as many feminine pronouns with inanimate antecedents; OC-I is about two thirds of the size of CC-I but contains about seven times as many feminine pronouns with inanimate antecedents. This suggests that the smaller amounts of relevant observations in the contemporary components is due to the lower frequency with which feminine forms of the anaphoric pronoun are used with inanimate antecedents in contemporary varieties. It should be noted that we are not dealing with a lack of potential antecedents, since there are hundreds of

(inanimate) feminine nouns in each component, counting just those which appear with an article or an attribute. One cause of this perception might be the difficulty in unequivocally establishing that a masculine pronoun is being used to refer to a feminine antecedent. In other words, it is usually quite straightforward, when a feminine pronoun is found, to link it back to a suitable feminine antecedent; but it can be difficult to determine whether a masculine pronoun actually co-refers with a (lexical) feminine antecedent (lack of grammatical agreement), or rather with some propositional or discursal antecedent which lacks inherent grammatical gender and for which masculine is the expected agreement form (the default agreement form, in the sense of Corbett, 1991: 205). To clarify this point, consider examples (11) and (12) from Irish and Welsh, respectively.

(11) Irish (RnaG “Barrscéalta” 22/10/1997)

ehm lasmuigh de sin abair lasmuigh de scoileanna agus cuairteoirí ehm tá leabharlann [?] againn anso chomh maith agus tá ehm tháimid ag iarraidh anois **an leabharlann** a fhorbairt ehm anois cuireadh tús maith leis **an leabharlann so** nuair a raibh an ehm an Dochtúir Canónach Ó Fiannachta [//] eh **leabharlann Mháirín agus Cearúil Uí Dhálaigh** ehm timpeall ar is dócha go bhfuil nach mór dhá mhíle leabhar againn anso, anois, míle cúig céad pé scéal é ehm leabhra a bhaineann le [/] bhuel ní hamháin le [/] leis an oileán ach leis an nGaelainn, agus stair, polaitíocht, gach haon saghas ábhar mar sin ach teastaíonn uainn i mbliana anois ehm **é seo** a fhorbairt

‘apart from that, apart from schools and visitors, we have a library here as well and we are trying to develop **the library**(F). Now **the library** was given a good start when Doctor Canon Ó Fiannachta was— eh **the library of Máirín and Cearúil Uí Dhálaigh**— ehm around— we must have about 2,000 books here, now, 1,500 whatever the story is, books having to do with, well, not only with the island and with the Irish language, but with history, politics, all sorts of subjects like that and now this year we want to develop

{ **it**(M) [i.e. the library: reading 1]  
 { **all of this**(M) [reading 2]’

(12) Welsh (**beti.cha** 207–212)

BET: mae **cerddoriaeth** yn cael lle amlwg iawn yn eich teulu chi

KIL: ydy mae **fe'n ffordd** o fyw ar bob ochr o'r teulu ers bo(d) fi'n cofio

*Translation:*

BET: **music**(F) has got a very prominent place in your family

KIL: yes, **it**(M) is a **way**(F) of life on both sides of the family since I can remember

In (11), a feminine form of the anaphor (e.g. *í seo*) would be unequivocally interpreted as co-referring with the lexical antecedent *leabharlann* (feminine) ‘library’; the masculine pronominal form *é seo*, on the other hand, may be interpreted as co-referring *leabharlann*,<sup>13</sup> or as referring (indistinctly) to the whole set of lexical and non-lexical antecedents evoked so far (school visits, range of books, the library itself, etc.). In the first case, we would be dealing with an instance of non-grammatical agreement; in the second, with an instance of default agreement with a non-lexical antecedent.<sup>14</sup> In (12), similarly, a feminine pronominal form such as *hi* would have been immediately linked back to feminine *cerddoriaeth* ‘music’ as its antecedent; but it is uncertain whether the antecedent of the masculine pronominal form *fe* is lexical (*cerddoriaeth*) or complex (propositional, e.g. ‘the fact that music is so prominent in my family’), in which case a masculine form would be grammatical.<sup>15</sup> Examples such as this were marked as uncertain and not counted, since it would have been an arbitrary choice whether to take them as evidence of either non-grammatical gender agreement or grammatical default agreement. As a result, a number of observations with potentially feminine antecedents were discarded.

What the data at hand suggest, at any rate, is a significant reduction in the use of feminine pronouns with non-female reference, which is further corroborated by the quantitative as well as non-quantitative observations found in the literature, as seen in chapters 1 and 5, especially as regards Donegal Irish (§1.2.2), Welsh (especially younger speakers, §5.7.4.2), certain

<sup>13</sup>Which may also be suggested by the repetition of *forbairt* (‘to develop’), the verb used with the first occurrence of *leabharlann*.

<sup>14</sup>Cf. §3.3.5.

<sup>15</sup>Another possibility would be forward attraction in the gender of the lexical predicate *ffordd* ‘way’, but even in this case a feminine form would be expected.

varieties of Scottish Gaelic (§5.5.2), Manx (§5.6) and Cornish (§5.9.2), where the tendency to overgeneralize masculine forms for anaphoric reference to inanimates is reported as established.

#### 7.2.4 Non-quantified variation

Variation in pronominal agreement has also been reported in the literature, where a greater incidence of masculine anaphoric forms with oblique cases of the pronoun is mentioned (see §3.3.4). However, as just shown, this was not observed in the traditional varieties, and could not be appreciated in the contemporary ones. In OC-I, agreement rates with prepositional pronouns are as high as with simple pronouns (see Table 7.3); in OC-W (Tables 7.10 and 7.11) and in the contemporary data there were too few instances of prepositional anaphora with feminine antecedents to appreciate any difference (see §7.2.1.2 for CC-I, Tables 7.12 and 7.13 for CC-W). The incidence of possessive pronouns with feminine antecedents is also very low in the data, in both the older and contemporary components: for OC-I cf. again Table 7.3, for OC-W Tables 7.10 and 7.11; for CC-I see §7.2.1.2 and for CC-W Tables 7.12 and 7.13.

### 7.3 Modelling anaphoric agreement

As discussed in §6.3, agreement in FDG is modelled as the copying of certain feature values from the controller onto the target(s). In the case of anaphoric agreement, the targets are referential elements which are normally found outside the syntactic constituent in which the controller is located, be it at phrasal or clausal level.<sup>16</sup>

<sup>16</sup>Anaphoric agreement may be found within the same noun phrase as the controller; in Turkish, for instance, the possessor is cross-referenced on the possessum by means of a possessive suffix and they both appear in the same noun phrase, as in (i).

- i. Turkish (Hengeveld & Mackenzie, 2008: 394)

Hasan-ın kitab-ı  
Hasan-GEN book-POSS:3SG

‘Hasan’s book’

As already seen in §4.5.3, it is typologically common, when there is a clash between grammatical and semantic agreement criteria, for the latter to prevail in anaphoric agreement; this is captured by the Agreement Hierarchy.

This implies that the overgeneralization of masculine agreement forms for inanimate antecedents in Insular Celtic is analyzed as a form of semantic agreement, i.e. as resulting from the resemantization of pronominal gender: in other words, the distribution of pronominal gender forms, which in traditional native varieties responded to both semantic and grammatical factors, now only responds to semantic factors, as will be made explicit below. In the resemantization of pronominal gender, the association between feminine forms and female sex becomes exclusive as the possibility for them to refer to inanimate antecedents is lost; correspondingly, the distribution of masculine pronominal forms (once associated with the grammatical feature masculine) can now be described as corresponding to a single semantic condition, [–female].

Whereas grammatical information like gender is among the inherent features of a Lexeme and therefore pertains to the Representational Level, Dikker (2004: 36) argues that semantic (or “conceptual”) information like natural gender must be drawn from the Contextual Component of the FDG model, which means that *semantic agreement* is “triggered by extra-grammatical information” (ibid.).

Dikker provides the following example, where the two types of agreement are triggered on different targets.

(13) Italian (Dikker, 2004: 37)

Sono sces-e quattro person-e, de-i  
 be.PRS.3PL alight.PTCP-F.PL four person(F)-PL of.ART-M.PL  
 qual-i tre sono andat-i.  
 which-PL three be.PRS.3PL go.PTCP-M.PL

‘Four people got off; of the four, three left.’

The epicene noun *persona* ‘person’ is grammatically feminine: the masculine forms of *de-i* and *andat-i* in example (13) might therefore sound substandard to a native speaker of Italian but might occur in spoken usage if all the people referred to are males. In the example, feminine agreement is found with the predicative participle *sces-e*, i.e. within the same clause, while masculine agreement forms are found in the subordinate clause, first on the

relative pronoun and then on the predicative participle which immediately agrees with the latter.<sup>17</sup>

In order to account for examples like (13), one must assume that the conceptual information which can influence the form of agreement markers is available and accessible throughout the whole process of sentence generation, including *after* the syntactic information has been accessed and copied onto another target (Dikker, 2004: 37). This assumption is referred to as the Maximal Input Hypothesis (see Dikker (2004: 37) and references therein) and contrasts with the Minimal Input Hypothesis, whereby “agreement is a purely syntactic operation, involving the copying of syntactic features of controllers onto targets” and “information from the conceptual structure [...] is no longer of influence on the output” once those features are retrieved (ibid.). The current FDG model (Hengeveld & Mackenzie, 2008) is in line with the Maximal Input Hypothesis: the Contextual Component also feeds into the levels of morphosyntactic and phonological encoding (cf. Figure 2.2 in §2.4 above), in accordance with Dikker’s (2004: 45) suggestion. (The model of FDG that Dikker was describing in 2004, on the other hand, followed the Minimal Input Hypothesis, with the Contextual Component only feeding into the Formulation level.)

In the following two examples, taken from the Irish data set, we see two instances of feminine pronouns being used to convey the natural gender of two referents denoted by grammatically masculine nouns. In both cases, the unlenited form of the nominative singular is consistent with masculine gender agreement, while the feminine pronominal form *sí* is consistent with natural gender.

(14) Irish

a. (Wigger, 2000 1-01-21)

Agus nuair a bhuaíl sé an doras, d’éirigh an cailín an bhfuil a fhios agat d’oscail a’ cailín is a mhac nuair a chonaic sí an ... ar ndóigh rud nach iontas ...

‘And when he knocked at the door, the girl(M) got up, you know, the girl(M) opened the door and—boy!—when she saw ... of course no wonder that ...’

<sup>17</sup>Strictly speaking, therefore, the participle *andat-i* agrees (syntactically) with the subject of the relative clause and not (semantically) with the subject of the superordinate clause.

b. (Wigger, 2000 6-01-05)

fuair an madadh boladh an tsionnach, agus thosaigh sí ag tóraíocht  
‘the dog(M) picked up the scent of the fox and (she) started pursuing’

However, there is a difference between (14a) and (14b): in the former the natural gender of the referent must be part of the lexical information (since it is central in the definition of *cailín* ‘girl’), while in the case of *madadh* ‘dog’ (epicene) the natural gender of the referent is an extra-linguistic piece of information which the speaker conveys, based on his or her knowledge of the relevant facts, through the form of the pronoun: once this happens, that piece of information becomes part of the ongoing discourse and is stored in, and accessed from, the Contextual Component (Dikker, 2004; Cornish, 2009).

Dikker (2004: 39) seems however to contradict herself as she maintains that semantic information about gender and number is drawn from the Contextual Component but then suggests the RL-representation in (15) for Italian *squadra* ‘team’, meant to capture the fact that its “conceptual” number is plural, i.e. that it is a collective noun capable of occurring with plural agreement forms, as in (16).

(15) Italian *squadra* ‘team’ (Dikker, 2004: 39)

RL: (1x: squadra<sub>NfemNpl</sub> (x))

(16) Italian (Dikker, 2004: 37)

Siamo un-a squadra che gioch-iamo.  
be.PRS.1PL one-F.SG team(F).SG REL play-PRS.1PL

‘We are a team that play.’

Clearly, if the conceptual plurality of *squadra* is (already) lexically specified as part of its meaning, then it need not be retrieved from the Contextual Component.

The question is then how the different kinds of information that can have a bearing on agreement are represented in the mental lexicon. So far, we have seen two different kinds: grammatical (e.g. the masculine gender of *cailín*) and semantic (e.g. the sex of the referent of *cailín*). We have also seen that there is semantic information which pertains to the referent but is

Table 7.14: Types of information that can affect agreement

|             | Linguistic                             | Extra-linguistic          |
|-------------|--|---------------------------|
| Grammatical | e.g. masculine gender of <i>cailín</i> | —                         |
| Semantic    | e.g. sex of <i>cailín</i>              | e.g. sex of <i>madadh</i> |

not encoded lexically, e.g. the sex of the referent of *madadh* ‘dog’. We can summarize this as in Table 7.14 (cf. Dahl, 1999).

Dikker’s proposal, represented in (15), is unsatisfactory because it appears to give the feature value <plural> equal status as the feature value <feminine> and is therefore at odds with the fact that singular number agreement is possible (and indeed normally observed) with Italian *squadra*. A better representation should distinguish between the two different kinds of lexically given information, and keep the latter distinct from extra-linguistic information.

Unlike in German, where neuters like *Mädchen* ‘girl’ may at least in principle be resumed both by neuter and by feminine anaphoric pronouns, in Irish only feminine forms of the pronoun (i.e. only semantic agreement) are admitted with *cailín*; a masculine form of the pronoun (syntactic agreement) would be considered ungrammatical, as observed in §5.4.2. This means that when a noun refers to a human and the sex of its referent is part of the lexical information, the form of the pronoun must reflect it.

In addition, as also illustrated in §5.4.2, non-syntactic agreement forms which do not reflect natural gender are common in traditional Irish dialects with certain nouns (e.g. masculine *bád* ‘boat’).

This state of affairs can be captured as a set of constraints on the selection of pronominal gender forms valid for conservative varieties, which may be stated as in (17).

- (17) a. A masculine pronoun may not refer to a female human; otherwise, it may co-refer with any masculine antecedent (with a few exceptions).  
 b. A feminine pronoun may not refer to a male human; otherwise, it may co-refer with any feminine noun (and exceptionally with a few masculine nouns).

(Speakers may, of course, deliberately violate either of these constraints for

pragmatic reasons.)

There are therefore three questions: (i) how semantic gender information is represented in the lexicon; (ii) how the idiosyncratic pronominal gender preferences of certain exceptional nouns are represented in the lexicon; (iii) how non-linguistic information which is relevant to semantic agreement is accessed.

Wechsler & Zlatić (2003) propose a model whereby two sets of agreement features are distinguished: *concord* features on the one hand and *index* features on the other. Concord features are morphosyntactic, i.e. they pertain to, or are motivated by, the morphology of the noun, and are responsible for local grammatical relations, in particular for agreement within the noun phrase (Wechsler & Zlatić, 2003: 14–16). Index features, on the other hand, have to do with the noun’s semantics and are indexical because they have to do with reference; they are involved in the type of agreement that is found between coreferential elements, e.g. a noun phrase and an anaphoric pronoun (ibid., pp. 15–17, 28).<sup>18</sup> Concord features may include gender, number and case; index features, on the other hand, gender, number and person.<sup>19</sup> The crucial point is stated by Wechsler & Zlatić as follows:

For a given noun the gender/number features of CONCORD and INDEX normally match [...] This gives the illusion (or so we claim) that a single feature bundle on the noun is responsible for all the agreeing items: determiners, adjective, verbs, and pronouns. However, with some nouns the values diverge, leading to mixed agreement patterns [Wechsler & Zlatić, 2003: 17; emphasis original.]

Wechsler & Zlatić (2003: 65) represent this situation as in (18), using the Serbian/Croatian noun *devojče* ‘girl’, which is grammatically neuter and behaves like German *Mädchen* in allowing both neuter and feminine anaphoric agreement forms.

<sup>18</sup>In the case of masculine nouns like *bád*, it is not obvious why the indexical value of gender should be feminine, and they should therefore be regarded as exceptions, unless one treats the occurrence of feminine pronouns as a case of personification, as does Ó Siadhail (1984: 175).

<sup>19</sup>Case is not an index feature since agreement in case between an antecedent and a coreferential pronoun is not required (Wechsler & Zlatić, 2003: 15).

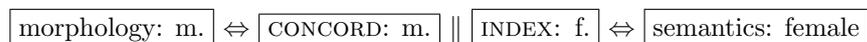
- (18) Serbian/Croatian *devojčē* ‘girl’, nt. (from Wechsler & Zlatić, 2003: 65)



In (18), the symbol  $\Leftrightarrow$  represents a match and the symbol || a mismatch; in this case, the mismatch is between index features on the one hand and semantics on the other. According to Wechsler & Zlatić (2003: 84), anaphoric pronouns may show index agreement or semantic/pragmatic agreement, as in the case of German *Mädchen* and Serbian/Croatian *devojčē*, for which both neuter and feminine pronouns are allowed, representing index agreement and semantic/pragmatic agreement, respectively.

However, unlike in the case of German *Mädchen*, and as stated in (17), only feminine agreement forms are admitted with nouns like *cailín* and *bád* in Irish. The mismatch is therefore between concord and index features, as the former agree with the morphology, while the latter agree with the semantics. This situation may be represented as in (19).<sup>20</sup>

- (19) Irish *cailín* ‘girl’, masc.



The next problem arises with purely contextual (i.e. non-linguistic) information such as the sex of epicenes like the aforementioned *madadh* ‘dog’ in Irish or masculine *ffrind* ‘friend’ in Welsh, both of which may refer to individuals of either sex. Their concord gender is masculine, as reflected by the agreement forms one observes within the noun phrase; both masculine and feminine pronominal agreement forms are possible, the latter representing semantic/pragmatic agreement with extra-linguistic, or contextual, information, if the referent is known to be a female. The mismatch, in this case, is potentially between index gender and semantics. This can be represented as in (20). As shown in the figure, the semantics of the noun is not specified, as it depends on the context; this is the kind of information that is stored in, and supplied by, the Contextual Component (Dikker, 2004).

- (20) Irish *madadh*, Welsh *ffrind*



<sup>20</sup>Cf. §5.4.3 about the gender of Irish derivatives in *-ín*.

Table 7.15: Examples of concord/index match/mismatch

| Noun                                      | Concord gender | Index gender |
|---|----------------|--------------|
| Irish <i>madadh</i> , Welsh <i>ffrind</i> | masc.          | masc.        |
| Irish <i>cailín</i> , <i>bád</i>          | masc.          | fem.         |

Table 7.16: Gender values mismatch for inanimate nouns

| Noun   | Older varieties |         | Contemp. varieties |              |
|--|-----------------|---------|--------------------|--------------|
|  | Concord         | Index   | Concord            | Index        |
| Ir. <i>bord</i> , W. <i>bwrd</i> ‘table’       | masc.           | ↔ masc. | masc.              | ↔ masc.      |
| Ir. <i>cathaoir</i> , W. <i>cadair</i> ‘chair’ | fem.            | ↔ fem.  | fem.               | <b>masc.</b> |

The situation relative to the Lexemes we have considered so far can therefore be represented as in Table 7.15. In the traditional varieties of Irish and Welsh, concord gender and index gender tend to coincide, as we have seen in §7.2 for OC-I and OC-W; there are a few exceptional cases like *cailín* and *bád* in Irish, whose index gender feature has value feminine (since masculine pronominal forms are never used). There is a further category, that of epicene nouns (like Irish *madadh* and Welsh *ffrind*): their index gender and concord gender do match, but semantic/pragmatic pronominal agreement can be observed when the Contextual Component supplies the relevant information.

The contemporary data for Irish and Welsh and the observations contained in the literature for a number of moribund or extinct Celtic languages, on the other hand, suggest that a resemantization of pronominal agreement has taken place, resulting in the overgeneralization of masculine agreement forms for inanimate antecedents, while the use of feminine agreement forms is only regular with female antecedents. This state of affairs can be described as a mismatch between concord and index gender: inanimate nouns in these varieties still have feminine concord gender, but masculine index gender. This can be summarized as in Table 7.16.

To explain how this situation came about, it should be kept in mind that, according to Wechsler & Zlatić (2003: 40), mismatches between concord and index features are the result of “pressures from the semantics”, as we have already seen. In other words, the semantic feature [−animate] came to

be felt, at some point, as being at odds with the index feature feminine. Gradually, the pronominal occurrence of masculine agreement forms with feminine inanimates increased in *frequency* and *salience*, resulting in the lexicalization of masculine pronominal agreement with feminine inanimates, i.e. in its encoding within their index features.

This analysis leaves two questions still unanswered, which will be addressed in turn in the next two sections: (a) why this happened, and (b) why it did not happen the other way around, i.e. why it was the masculine and not the feminine gender that was generalized for all inanimate referents.

## 7.4 Agreement reduction and pronominal switch

Audring (2008) observes that, cross-linguistically, gender systems whose only exponents are pronominal (as in English) tend to be organized around semantic assignment criteria; formal assignment criteria tend to enter the picture only when gender agreement is also marked within the noun phrase. In other words, the salience of gender agreement marking within the noun phrase correlates directly with the existence of a grammatical agreement system in the pronominal domain, and, diachronically, when the former is reduced (e.g. through the loss of agreement-marking exponents), pronominal agreement tends to switch from grammatical to semantic. The latter point is exemplified by Audring with the cases of English, Dutch and a number of Mainland Scandinavian languages: in all of these languages, the loss of agreement morphology within the noun phrase (due to processes of phonological erosion and coalescence) was, or is being, followed by a resemanticization of pronominal gender agreement.

In other words, the neutralization of agreement morphology within the noun phrase causes the loss of gender-marking contexts, while pronominal exponents of gender agreement are spared; but the distribution of the latter changes, as they start to mark agreement with semantic features of the referent rather than with morphosyntactic features of the antecedent.

In languages like Dutch, this process is still ongoing and can be observed. Audring (2009) describes the way in which pronominal agreement is being reorganized in Standard Dutch, a language which distinguishes two genders within the noun phrase and three genders pronominally, as shown in Table 7.17. Within the noun phrase, the two genders are common and neuter,

Table 7.17: Genders in Standard Dutch

| Agreement domain | Genders distinctions |      |     |
|------------------|----------------------|------|-----|
| Noun phrase      | common               |      | nt. |
| Pronouns         | masc.                | fem. | nt. |

while pronominally they are masculine, feminine and neuter. Common gender historically arose out of the coalescence of the masculine and feminine gender exponents in the paradigms of NP-internal agreement targets such as the articles (definite and indefinite) and the attributive adjective.

The examples in (21) show the patterns that this system produces.

(21) Standard Dutch (De Vos & De Vogelaer, forthcoming)

- a. Waar is de stoel? Hij staat buiten.  
 where is the.COMMON chair(COMMON) 3SG.M stands outside  
 ‘Where is the chair? It is outside.’
- b. Waar is de tafel? Ze staat buiten.  
 where is the.COMMON table(COMMON) 3SG.F stands outside  
 ‘Where is the table? It is outside.’
- c. Waar is het rek? Het staat buiten.  
 where is the.NT rack(NT) 3SG.NT is outside  
 ‘Where is the rack? It is outside.’

In fact, as De Vos & De Vogelaer (forthcoming) point out, this system is maintained in southern and Belgian varieties of Dutch; in northern varieties, e.g. as spoken in the Netherlands, speakers no longer produce grammatical agreement in the pronominal domain. Instead, in informal usage,

Dutch speakers choose their pronouns on the basis of semantic patterns. Masculine pronouns are used for male persons, for all animals (even for animals of female sex) as well as for countable, bounded objects and abstracts. Neuter pronouns, by contrast, appear in combination with mass nouns and uncountable, unbounded, unspecific abstracts. Feminine pronouns have the most restricted distribution: they can only refer to female persons and (occasionally) female animals. [Audring, 2009: 215]

The reason why speakers of southern varieties of Dutch hold on to the traditional grammatical system of pronominal agreement when they speak Standard Dutch is, according to De Vos & De Vogelaer, that in their own native (non-standard) varieties masculine and feminine gender are still distinguished within the noun phrase, as shown in Table 7.18 where the two varieties are compared.<sup>21</sup>

As can be seen from Table 7.18, masculine and feminine genders are clearly and always distinguished in Dendermond Dutch within indefinite noun phrases, thanks to the different forms of the indefinite article (m. *ne(n)* vs. f. *een*), and optionally within definite noun phrases, where the masculine form of the article may be distinguished from the feminine one by the ending in *-n*. The attributive adjective may also distinguish a masculine from a feminine form, while the neuter form is always recognizable by the absence of any agreement morpheme. In Standard Dutch, on the other hand, there is no formal difference between masculine and feminine in terms of either article or adjective inflection; furthermore, the neuter indefinite article is identical to the common-gender one, while the inflection of the attributive adjective makes no gender distinction at all within a definite noun phrase.

This analysis is confirmed, as De Vos & De Vogelaer observe, by the positive correlation between grammatical-gender agreement with anaphoric pronouns and dialect proficiency in the southern Dutch-speaking areas, a correlation which has been shown to exist by Hoppenbrouwers (1983): speakers who are more proficient in their local dialect are more likely to observe grammatical agreement pronominally when speaking Standard Dutch.

If this analysis is correct, then it is not surprising that something similar might have happened in Irish and Welsh, as agreement morphology within the noun phrase declines. As discussed in chapter 6, gender agreement within the noun phrase has become less salient over time, particularly as far as attributive elements are concerned; at the same time, the demise of the genitive case in spoken Irish (and Scottish Gaelic; see §§5.4.3 and 5.5.1 above) means that in spoken varieties the form of the article may no longer distinguish gender either, leaving head-marked agreement (initial mutation of the noun) as the only exponent of gender within the noun phrase—as is

---

<sup>21</sup>Dendermonde (mentioned in Table 7.18) is a Belgian city near the border between East Flanders and Brabant.

Table 7.18: Gender marking in the noun phrase, Dendermonde vs. Standard Dutch

| Gender | Dendermonde                                    |  | Standard                                     |   |
|--------|--|--|--|---|
|        | definite                                       | indefinite                                   | definite                                     | indefinite                                |
| masc.  | de(n) lang-e(n) dag<br><i>the long-AGR day</i> | ne(n) lang-e(n) dag<br><i>a long-AGR day</i> | de lang-e dag<br><i>the long-AGR day</i>     | een lang-e dag<br><i>a long-AGR day</i>   |
| fem.   | de goed-e pijp<br><i>the good-AGR pipe</i>     | een goed-e pijp<br><i>a good-AGR pipe</i>    | de goed-e pijp<br><i>the good-AGR pipe</i>   | een goed-e pijp<br><i>a good-AGR pipe</i> |
| nt.    | het braaf kind<br><i>the good child</i>        | ee(n) braaf kind<br><i>a good child</i>      | het brav-e kind<br><i>the good-AGR child</i> | een braaf kind<br><i>a good child</i>     |

(De Vos & De Vogelaer, forthcoming: Table 1)

already the case in Welsh.

The case of Insular Celtic is therefore in line with a cross-linguistically observed tendency for pronominal gender agreement to be exclusively or almost exclusively semantic in languages where only pronouns mark gender; diachronically, as is well documented at least within Indo-European, this can result from a resemantization of pronominal gender, which takes place after gender marking within the noun phrase has become less salient, or less transparent, or has disappeared.

## 7.5 Markedness, convergence and directionality

As a result of the resemantization process just discussed, the occurrence of feminine anaphoric pronouns is now restricted; they are regularly used only when the lexical antecedent has a female referent. Masculine forms, on the other hand, have taken over the remaining semantic space, which includes all (or most) non-female referents, as shown in Figure 7.1(b). In principle, the opposite could have happened, with the distribution of masculine anaphoric pronouns being restricted to male reference and feminine ones being used elsewhere, as shown in Figure 7.1(c).

I will argue in this section that the answer to this question is partly to do with language-internal factors, i.e. markedness, and partly with language contact, i.e. with the notion of convergence, which was introduced in §1.2.2 above.

Numerically, as the data presented so far show, the masculine is the unmarked gender in both Irish and Welsh.<sup>22</sup> We can represent this situation graphically as in Figure 7.1(a). In OC-I, if we take as an example the set of nouns occurring with the article, there are 110 feminines and 232 masculines, the former representing about 33% of all nouns (see Table 6.1); in OC-W, 53 feminines and 174 masculines were counted, the former representing less than a quarter of the total (see Table 6.6). Masculine nouns are more frequent than feminine nouns in both corpora: the result is that even in the OC varieties, where pronominal agreement with inanimates is still robustly grammatical, feminine pronominal forms are also remarkably less frequent than masculine ones: they represent 22% of all pronouns ( $N = 459$ ) in OC-I and 28% of all pronouns ( $N = 237$ ) in OC-W (see Tables 7.1 and 7.10,

<sup>22</sup>Also cf. Fowkes (1957, 1987).

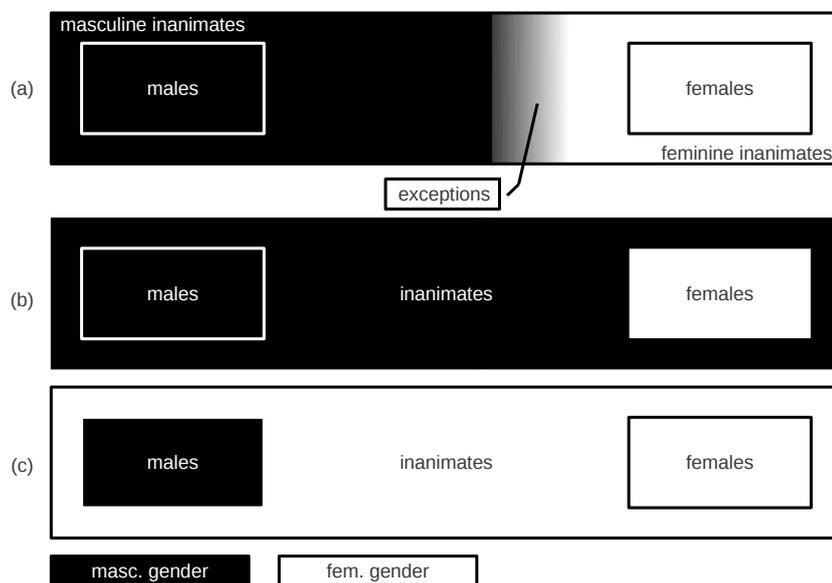


Figure 7.1: Pronominal gender distributions in Irish and Welsh  
 (a) original distribution, (b) new distribution, (c) hypothetical scenario.

respectively).

These figures were calculated on the basis of just the anaphoric occurrences of the personal pronouns, so they can be considered a conservative estimate, since they leave out all instances of masculine pronominal forms used in discourse anaphora, complex anaphora (as discussed in §3.3.5), dummy anticipation of non-finite clauses and impersonal expressions, as shown in examples (22) and (23) for Irish.

(22) Irish (Ó Siadhail, 1989: 274)

Taithníonn  $sé_i$  liom [a bheith anseo]<sub>i</sub>  
 please.PRS 3SG.M with.1SG PTC be.VN here

‘I like to be here (lit. “it pleases me being here”)’

(23) Irish (Ó Dochartaigh, 1992: 39)

Tá  $sé$  a trí a chlog  
 be.PRS 3SG.M PTC three o'clock

‘It is three o'clock’

In Welsh, the use of neuter demonstratives and feminine pronouns is attested in complex anaphora (24a) and in cataphoric relation with declarative clauses (25a),<sup>23</sup> but the use of masculine pronouns in similar contexts is also possible, as shown in (24b) and (25b).

## (24) Welsh

## a. (Thomas, 1992b: 294)

[Oedd gwerthu'r car yn beth fføl  
be.PST sell.VN=ART.M.SG car(M).SG PTC thing(M).SG foolish  
i'w wneud.]<sub>i</sub> Mae hynny<sub>i</sub>'n amlwg rŵan.  
to=POSS:3SG.M do.VN be.PRS that.NT.SG obvious now

'Selling the car was a foolish thing to do. That is obvious now.'

## b. (Thomas &amp; Thomas, 1989: 107 ll. 1–2)

Och chi'n cal [mynd un waith y  
be.PST.2PL 2PL=PTC get.VN go.VN one.F time(F).SG ART.F.SG  
flwyddyn efo'r capel i'r Rhyl]<sub>i</sub>,  
year(F).SG with=ART.M.SG chapel(M).SG to=ART.M.SG Rhyl  
a'dd o<sub>i</sub>'n *treat* mawr  
and=be.PST 3SG.M=PTC treat(M).SG great.M.SG

'You got to go to Rhyl with the chapel once a year, and it was a great treat.'

## (25) Welsh

## a. (Thomas, 1992b: 303)

Mae hi<sub>i</sub>'n amlwg [bod Mair wedi anghofio]<sub>i</sub>.  
be.PRS 3SG.F=PTC obvious be.VN Mair after forget.VN

'It is obvious that Mair has forgotten.'

## b. (Thomas &amp; Thomas, 1989: 138 ll. 41–42)

on nw'n gweud y stori 'ny bod  
be.PST 3PL=PTC tell.VN ART.SG story(F).SG there be.VN  
e<sub>i</sub>'n itha gwir, chwel, [bod e 'di ifed  
3SG.M=PTC quite true you see be.VN 3SG.M after drink.VN  
gormod]<sub>i</sub>  
too much

'they told that story, that it was quite true, you see, that he had been drinking too much'

<sup>23</sup>Cf. examples (65) in §5.7.1.7 and (76) in §5.7.1.10.

A second possible factor in the observed developments is contact with English. As discussed in §1.2.2, various attrition phenomena have been observed and reported on, whereby younger speakers tend to adopt forms and structures which represent the “lowest common denominator” between Irish and English, reinforcing shared features and syntactic structures to the detriment of those that have no counterpart in English. It was also noted there that in the final stages of Manx and Cornish gender agreement was largely pronominal and semantically motivated.

There is of course a major difference between pronominal agreement in Irish and in English: the former has two genders while the latter has three, so that masculine pronouns in Irish fulfil the functions of both the neuter and masculine pronouns in English.<sup>24</sup> This is, however, also consistent with the convergence hypothesis. Structurally, the use of masculine pronominal forms in Irish and Welsh such as exemplified in (22), (23), (24b) and (25b) parallels that of the neuter pronoun in English, which is of course also consistently used for inanimate referents. Therefore, the actual direction of pronominal gender resemanticization, whereby masculine forms take over a semantic space that was previously shared with the feminine, is consistent with the convergence hypothesis, in that it represents a synthesis of what the two systems in contact already had in common or, paraphrasing Toribio (2004: 167), the most parsimonious pronominal agreement system that serves both contact languages.

## **7.6 Grammatical gender as an afunctional category**

As we saw in §4.2, it is often suggested that the function of gender is to increase discourse cohesion and thus enhance the chances of successful communication through agreement (reference tracking). However, the view was also presented that gender, at least in small gender systems such as are found in Indo-European languages, has no longer any function and is a morphosyntactic relic, one which is the more persistent the more cognitively en-

---

<sup>24</sup>But in the final stages of Manx, according to Phillips (2004), a three-way distinction was found in pronominal agreement (cf. §5.6.1.6). If his analysis is correct, it would lend further support to the idea of a pronominal agreement system more closely matching that of the contact variety after being restructured.

trenched it is. A useful question which it is useful to repeat here is whether the cognitive cost of maintaining a gender system is, so to speak, offset by its utility as a reference-tracking device (Trudgill, 1999: 141). Trudgill's answer is negative: he observes that, especially in Indo-European languages with three gender classes at most, the reference-tracking power of gender agreement is not great, compared to the richness of the gender-related morphosyntactic apparatus that the speaker has to master, and the occasions in which reference tracking is helped by gender agreement are not many, limited—to be precise—to those contexts in which the potential referents of an anaphoric pronouns are all of different genders. He also observes that when pidgin varieties are formed, grammatical gender is always lost in the process; and that when creoles emerge from pidgins, gender is never reintroduced, either (§4.2); from this he concludes that gender is a weakly functional or afunctional category, which languages can very easily dispense with.

Our findings show that the only domain in which grammatical gender agreement is fully preserved in contemporary varieties of Irish and Welsh is article-noun agreement. This is, as discussed in chapter 6, a domain where agreement is (now almost exclusively) marked on the head rather than on the dependent; dependent-marked agreement has become less frequent or altogether ceased to be productive. In other words, morphosyntactic information pertaining to the controller is no longer copied on other constituents but stays on the controller itself; pronominal agreement, as we saw, now reflects semantic and/or pragmatic information. Pronominal gender in Insular Celtic has undergone a resemantization process ending in a new division of labour between masculine and feminine forms: the latter are now (almost exclusively) used to co-refer with semantically feminine (i.e. female) antecedents, whereas the former are used for any other kind of referent, whether animate or inanimate. As we have seen, one of the most remarkable results of this process is that the distribution of masculine *vis-à-vis* feminine pronominal forms has been fundamentally altered: the former no longer occur outside the realm of animate reference, which means that in those contemporary varieties reference tracking via gender agreement is no longer possible as far as inanimate referents are concerned.

If we assumed that helping the speakers track reference through agreement is the main function of gender, this would certainly be an unexpected

state of affairs. This seems to lend further support to Trudgill's (1999) argument that gender system is an afunctional category, i.e. one whose persistence cannot be explained in functional terms.

## 7.7 Gender agreement: lexical vs. contextual information

In this chapter I proposed that in order to model anaphoric agreement it is necessary to distinguish the different types of information that can affect gender agreement. It was argued that such information is either grammatical or semantic; grammatical information (grammatical gender) is system-internal and, as argued in §6.2.3, stored in the lexical entry of at least the most common nouns; semantic information is either lexically encoded as part of a noun's definition (semantic gender) or extra-linguistic, i.e. retrieved from the discourse. As was also shown in this chapter, semantic gender, even where relevant, may be left unspecified in the lexical entry, as in the case of Irish *madadh*; this gap is filled by extra-linguistic information, where available.

As explained in §2.4, extra-linguistic information enters the hearer's Contextual Component, where it is stored to be used in the ongoing discourse, via the Grammatical Component (in comprehension mode); the latter, in the FDG model, does not have a memory, so it has to retrieve this type of information, where available, from the Contextual Component. Extra-linguistic information of this type about the natural gender of a referent is short-term knowledge derived from the actual communicative situation (Hesp, 1990a: 16f.).

Lexically encoded gender information, on the other hand, is long-term knowledge (Hesp, 1990a: *ibid.*). As explained in this chapter, Wechsler & Zlatić (2003) propose that lexical gender is encoded in two distinct sets of features, reflecting the distinction between syntactic and semantic agreement respectively: concord features, depending primarily on morphosyntactic factors such as the noun's declension, and index features, which depend on, or are open to, the influence of semantics. In chapter 4 this distinction was captured by referring to formal vs. semantic assignment criteria: grammatical gender is therefore encoded as part of a noun's concord features and semantic gender as part of its index features. For most nouns, concord fea-

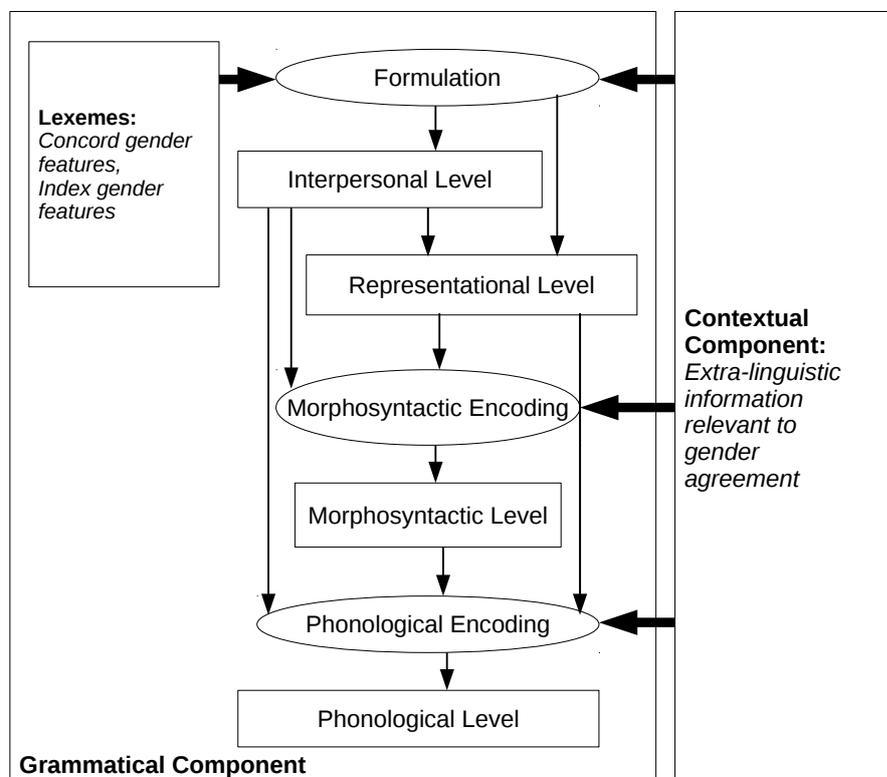


Figure 7.2: Sources of gender agreement information in FDG

tures and index features match, which in relation to gender means either that there is no clash between formal and semantic assignment criteria or that semantic criteria are not relevant for a particular noun. These two types of lexical information must be encoded separately for each Lexeme.

Figure 7.2 shows from where the different kinds of information relevant to gender agreement (*italicized*) enter the Formulation and Encoding components of the system. Lexical information enters the derivation during Formulation and is encoded in the RL representation of the intended utterance (cf. §6.3), while information from the Contextual Component may enter the derivation at any stage, as observed in §7.3.

## 7.8 Summary

In this chapter we have seen how the distribution of feminine agreement forms of anaphoric pronouns in contemporary varieties of spoken Irish and Welsh differs from that of older traditional varieties and a number of questions were addressed: first and foremost, in relation to the research questions set out in §1.1, how we can characterize the variation in anaphoric gender agreement found in the data and in the literature on Irish, Welsh and the other Insular Celtic languages; in particular, we wanted to establish whether this variation is random or systematic, and in the latter case, how it can be modelled. As was shown, the variation we observe is systematic: in the contemporary varieties, feminine forms of anaphoric pronouns appear to be used virtually only for co-reference with semantically feminine antecedents, i.e. for reference to female beings, a finding which is in line with previous claims—based on both quantitative and non-quantitative analyses—made in the literature. This was interpreted as resulting from a resemantization of pronominal gender and was observed to be fully consistent with typological expectations regarding the distribution of syntactic versus semantic agreement forms.

Arising from the above is the question where linguistic and extra-linguistic information that might influence pronominal agreement is stored, and how it can be represented. The model proposed here is based on a distinction between grammatical and semantic information; the former kind is lexically specified, while the latter can either be lexically specified (e.g. the natural gender or sex of *cailín* ‘girl’ is part of its lexical meaning) or extra-linguistic, e.g. the natural gender of an epicene noun like *madadh* ‘dog’. The model also distinguishes, following Wechsler & Zlatić (2003), between concord and index features, and the resemantization of pronominal agreement in Irish and Welsh is analyzed as a switch in index feature values for inanimate nouns.

This led to the questions what might have caused the resemantization of pronominal agreement and why it was the masculine rather than the feminine agreement forms that were generalized for inanimate antecedents. As regards the first question, it was argued that the resemantization of pronominal agreement may be explained as a result of the gradual demise of gender agreement marking within the noun phrase, which was the object of chapter

6, and cross-linguistic parallels were drawn in support of this hypothesis. As regards the other question, the switch leading to the overgeneralization of masculine forms was explained as a combination of two factors, language-internal (markedness) and contact-induced (convergence).

All in all, the developments discussed in this chapter and in the previous one lead to varieties in which dependent-marked gender agreement is declining as far as noun-phrase targets are concerned, and has undergone resemantization in the anaphoric domain. This is particularly significant when considered in the context of the debate on the function of gender. The resemantization of anaphoric gender agreement in particular, leading to a substantial reduction in the discriminatory power of anaphoric pronouns as a reference-tracking device, invites the conclusion that reference tracking is not the reason that keeps grammatical gender systems in existence in languages with few gender subdivisions such as Indo-European languages in general, as already argued by Trudgill (1999).

## Chapter 8

# Summary and conclusions

### 8.1 Overview

The present chapter provides a brief summary of the contents of this work (§8.2) and a review of its main findings in relation to the research questions (§8.3). The significance of this study is reviewed and discussed in §8.4, where the need for additional elaboration of the FDG model is also argued for. Directions for further research on grammatical gender in Insular Celtic are suggested in §8.5.

### 8.2 Summary of the previous chapters

The aims of this study are presented in chapter 1, which also provides some background information about the sociolinguistic status of Irish and Welsh and introduces the concepts of language contact, obsolescence and convergence.

Chapter 2 is devoted to the theoretical framework of this study, Functional Discourse Grammar (FDG), and presents the reasons why it was adopted for this study. The emergence of the FDG model is delineated, with reference to both the broad functionalist movement and its immediate predecessor, Simon Dik's theory of Functional Grammar; theoretical notions central to my analysis of the data are presented and discussed. The details concerning the methodology are given in chapter 3, where the composition of the data sets used, and the analytical process followed, are described and motivated.

The grammatical category under investigation is described in some detail in chapters 4 and 5, from a typological point of view and with specific reference to the Insular Celtic group, respectively. Chapter 4 introduces two important issues: whether gender information is lexically stored or rather worked out “on line” based on language-specific algorithms; and whether grammatical gender can be considered a functionally motivated system. In chapter 5 the systems of gender assignment and agreement are outlined for each of the six languages in the Insular Celtic group; initial mutations, which are central to gender agreement in Insular Celtic, are also discussed there. Arguments against phonological and morphophonological models of initial mutations are presented in §5.3.2 and contrasted with a number of morpholexical models, which provide a more satisfactory account of the relevant phenomena. A morpholexical account of initial mutations compatible with FDG is then proposed.

In chapter 6 I dealt with gender agreement within the noun phrase in Irish and Welsh. The findings of the analysis carried out on the data were presented, and it was shown that contemporary varieties differ from traditional ones in terms of agreement consistency within the noun phrase. An important difference between traditional and contemporary varieties is that grammatical gender agreement in the latter is observed significantly more consistently between the article and the noun than between the noun and the attributive elements, a differential treatment of agreement targets which is not found in the older data. The agreement system was modelled in terms of the top-down and incremental model of Functional Discourse Grammar and it was shown that such a model can account for the observed variation; the consequences of these findings for the discussion on processual vs. lexical gender assignment were also examined.

Chapter 7 is devoted to anaphoric agreement with lexical antecedents: the findings of the analysis were presented and the patterns of variation observed therein detailed. A remarkable divergence between traditional and contemporary varieties was found as regards the distribution of feminine pronominal forms: in traditional varieties the distribution of gendered anaphoric forms responds to the grammatical gender of their lexical antecedents (with a few well-established exceptions), whereas in contemporary varieties feminine agreement forms seem to be employed almost exclusively with semantically feminine antecedents, i.e. when the referent of the lexical

antecedent is semantically feminine. This was interpreted as resulting from a resemantization of pronominal gender and its outcome was found to be in line with typological expectations regarding the distribution of syntactic versus semantic agreement forms.

A model of anaphoric agreement which takes into account the different kinds of information that can have a bearing on pronominal agreement was proposed, drawing on the distinction between concord and index features already put forward by Wechsler & Zlatić (2003); this also implies a refinement of the lexical specifications assumed so far in FDG theory. The resemantization of anaphoric agreement was then considered in the light of previous typological studies which point to a correlation between the salience of gender agreement within the noun phrase and anaphoric agreement variation.

The role of both language-internal and language-external factors in determining the direction of the resemantization process was also discussed.

### **8.3 Review of research goals and main findings**

The main research goals set out in chapter 1 were (i) to characterize gender agreement in Insular Celtic, by providing a descriptively adequate model of its complex system, and (ii) to provide a quantitative analysis of the gender agreement variation that has been reported in the literature, in order to identify and characterize systematic patterns. By considering the phenomena under investigation in the wider typological context, I also intended to fill the gap noted in §1.2.2, where it was observed that Insular Celtic varieties have not featured in any typological studies of grammatical gender despite their peculiarity in the Indo-European context.

Following a detailed description of traditional gender systems in Insular Celtic (chapter 5), agreement marking in the noun phrase was discussed in chapter 6 and anaphoric agreement in chapter 7, focussing on Irish and Welsh as representative of the two main branches of this language group, Goidelic and Brythonic respectively.

In line with a basic tenet of functional linguistics (§2.2), the analysis was conducted on linguistic data drawn from actual use; the main findings were that grammatical agreement with feminine nouns is not observed as regularly as with masculine nouns, and that there are clear patterns as regards the distribution of grammatical agreement forms across different target types

from both a synchronic and a diachronic viewpoint.

Diachronically, the contemporary data sets differ remarkably from the older ones in their treatment of agreement. Generally speaking, the latter show a fairly consistent agreement system in both languages: grammatical agreement forms are equally likely to appear with controllers of both genders in the contexts of article–noun and attribute–noun agreement; in the context of anaphoric agreement, they are equally likely with both masculine and feminine antecedents in Welsh, slightly less likely with feminine antecedents in Irish, irrespective of whether the nouns refers to animates or inanimates.

In the contemporary data sets, on the other hand, grammatical agreement forms appear as frequently with feminine controllers as with masculine ones only in the context of article–noun agreement (however, even in this context, they are slightly less frequent with feminine controllers in Irish); in the case of adjective–noun agreement, grammatical agreement with feminine controllers is significantly less likely than with masculine controllers in both languages (and below chance level in Irish). As regards anaphoric agreement, feminine agreement forms are very rare when reference was not to a female being (semantic agreement), and masculine forms are commonly used for anaphoric reference to lexical antecedents with inanimate reference.

It appears therefore that while grammatical agreement forms are found across the board in the older data sets, the distribution of feminine agreement forms in the contemporary ones depends on target type; furthermore, the contemporary data reveal that the semantic factor is irrelevant to agreement within the noun phrase but relevant to anaphoric agreement.

The pattern of variation observed in the noun phrase poses a problem for a model in which agreement is characterized uniquely by the notion of feature copying. This is discussed in chapter 6, where emphasis is laid on an important difference between article–noun and adjective–noun agreement, i.e. that the former is marked on the head (and marginally, in Goidelic, also on the dependent) whereas the latter is only marked on the dependent. In the model proposed here, feature copying only applies to dependent-marked agreement: information pertaining to the head noun is copied onto targets like the attributive adjective, whose mutation is, to use the terminology discussed in chapter 5, syntactic rather than lexical. The mutation of the noun after the article is itself seen as syntactic rather than lexical, therefore it does not depend on feature copying because the relevant feature (gender)

is inherent to the noun. Since head marking does not depend on feature copying, its persistence is not linked to the decline of dependent marking, which can be interpreted as the gradual decline of feature copying within the noun phrase.

In chapter 7 the distribution of feminine pronominal forms in contemporary vis-à-vis traditional varieties of Irish and Welsh was considered. The fact that feminine forms of anaphoric pronouns now appear to co-occur almost exclusively with semantically feminine antecedents was interpreted as the result of a resemantization process with typological parallels in languages in which gender agreement within the noun phrase is not particularly salient or has lost salience. Arising from the perceived inaccuracies of existing accounts within the functionalist literature, a model was also proposed of how different kinds of information affecting anaphoric gender agreement are stored and accessed by the speakers. The model draws on an important distinction made by Wechsler & Zlatić (2003) between concord and index features. The issue of directionality was also examined and it was suggested that a combination of language-internal and contact-induced factors (markedness and convergence, respectively) may account for the fact that masculine rather than feminine forms of the anaphoric pronouns are overgeneralized as a result of the resemantization process.

Finally, I return to the two collateral questions asked in chapter 1: whether the existence and persistence of gender may be explained on functional grounds and whether gender information is part of the lexical inventory of a language or, as argued by Corbett (1991), it is computed on line when required, based on language-specific sets of rules.

The first question (§7.6) is complicated by the difficulty inherent in distinguishing between the function of gender and the function of agreement; the two are easily confused since there would be no gender without agreement, which is the crucial difference between gender and other nominal categories like number and case, whose expression does not depend on agreement. Furthermore, unlike number and case, gender cannot always be interpreted semantically. A frequently made statement is that gender, as a generator of agreement distinctions, introduces redundancy into the discourse, which in turns helps decoding the message by facilitating reference tracking and eliminating ambiguity. However, although gender does increase the number of distinctions that may be made anaphorically, this appears to be of limited

utility in languages such as the Indo-European ones, in which the number of genders varies between two and three and paradigmatic syncretism often neutralizes these distinctions in contexts in which they would be just as useful, e.g., commonly enough, in the plural; to this it should be added that the overgeneralization of masculine forms in our data hits first and foremost anaphoric targets, which means that the number of possible distinctions is reduced precisely where they would be most useful to keep track of reference, but maintained in the noun phrase, where such distinctions are less, if at all, useful. Furthermore, the literature shows that gender agreement is abandoned when languages are pidginized and never restored when creoles emerge from pidgins, suggesting a weakly functional or altogether afunctional category. To conclude, it seems plausible to conclude (with Trudgill, 1999, cf. §4.2) that in the case of Indo-European-style gender systems the cognitive load they impose is not offset by any gain and that such systems should therefore be regarded, synchronically, as afunctional features which survive for as long as language learners are exposed to them, but decline rapidly in circumstances, such as pidginization and language obsolescence, where the input available to them is deteriorated in both qualitative and quantitative terms.

The second question (§6.2.3) is whether grammatical gender information is stored as part of the lexical entry of each noun, or computed on line (when required) based on language-specific rules. The advocates of this latter position argue that the on-line solution is more economical in terms of mnemonic load for the speakers and maintain that if gender information were memorized we would witness a higher incidence of unexpected agreement forms, due to erroneous gender recall (Corbett, 1991: 7). However, it has been called into question precisely why recalling lexical information in the case of grammatical gender should lead to more errors than would running an algorithm on line (e.g. van Berkum, 1996: ch. 2).

The psycholinguistic evidence reviewed in §4.4 and the evidence presented in §6.2, taken together, support the lexical position, or at least its weaker version that grammatical gender is stored lexically for the most commonly occurring nouns; they also support the suggestion that gender information might be represented as the form of the article–noun sequence, made more salient by the patterns of initial mutation occurring within it, rather than as an abstract feature.

## 8.4 Significance of this study

The main contribution offered by this study to our comprehension of grammatical gender is its theoretically-informed treatment of the Insular Celtic system, which brings together evidence pertaining to the various languages of this group. Particularly important in this regard is the observation that the same simplification patterns are evident in all six languages, regardless of the degree of support (or lack thereof) in terms of revitalization efforts and language promotion; this points to the important and consistent role of obsolescence in the bilingual context.

The findings of the analysis of actual spoken data are significant both from the point of view of assignment (resemantization) and from the point of view of agreement, two obviously interconnected aspects of the same phenomenon, and confirm the results of previous research suggesting that as the latter becomes less salient within the noun phrase, resemantization ensues.

As a contribution to the study of contemporary Irish in particular, a specific data set (CC-I) was designed and put together as an integral part of this research project. The significance of this contribution is demonstrated by the unavailability of any similar data set to date, and it is envisaged that the CC-I data will be made available to the research community in the near future as part of a larger data collection project which is currently in its inception. As explained in §8.5, some of the important questions regarding grammatical gender in Irish and Welsh, to which this study could not provide an answer, might be addressed when a larger corpus is available.

Finally, this study contributes to the refinement of the FDG theory by providing a more detailed structural description of the lexical entry and of the connection between Formulation-level lexical primitives (Lexemes) and Encoding-level lexical primitives (suppletive forms). By suggesting a more refined analysis of the different types of agreement features encoded in the lexical entry, it allows to characterize anaphoric agreement more effectively. In relation to the connection between Lexemes and suppletive forms, I further suggested a new formulation of Dik's principle of lexical priority which describes a more efficient system and provide the missing link between the set of Lexemes and the set of suppletive forms in FDG. A model of attributive-Np agreement in Insular Celtic is also put forward.

As explained in chapter 2, Hengeveld & Mackenzie's (2008) FDG is a model of language production informed by cross-linguistic evidence. As such, it provides a theoretical framework at the same time typologically sound and flexible enough to permit the analysis of language-specific phenomena such as initial mutation and attributive-Np agreement. As a functional model, FDG is sophisticated enough to avoid the simplistic conception that linguistic structure is always shaped by functional motivations (the necessity to convey meaning), and recognizes the existence of afunctional (no longer functional) linguistic systems, e.g. grammatical gender as attested in Indo-European. Like FG, it rests on the assumption that the synchronic constraints captured by typological research are operational in diachronic change as well (principle of diachronic adequacy) and that any theoretically valid description of linguistic change should be consistent with this principle. Accordingly, my own account of change in gender agreement marking and of pronominal gender resemantization was supported by evidence drawn from the study of comparable changes in other languages.

The model proved adequate to describe the two different mechanisms of gender agreement found in Insular Celtic within the noun phrase; in terms of anaphoric agreement, it was argued that the model can be refined by incorporating Wechsler & Zlatić's (2003) distinction between concord and index features at the level of the Lexeme (Representational Level).

There is still room for improving FDG. Further work should be devoted in particular to describing in some detail how the "non-grammar" components work (cf. Butler, 2008: 19), and indeed some attention to the Conceptual and Contextual Components has been paid in more recent publications (e.g. Rijkhoff, 2008; Cornish, 2009). Also to be undertaken is a study of the interface between the Grammatical and the Contextual Component, where the exchange of extra-linguistic information relevant to anaphoric agreement takes place, and of the interface between the Morphosyntactic and the Phonological Levels, crucial to the production of initial mutation.

## **8.5 Directions for future research**

The analysis of the data sets described in chapter 3 made it possible to address the two main issues of providing a descriptively adequate model of gender agreement in Irish and Welsh as representative of the Insular Celtic

language group and to account for the variation in gender agreement that had been reported in the literature; it also permitted to contribute to the ongoing debate surrounding the possibility of a functional explanation for the existence of grammatical gender systems and the ways in which gender information is treated and processed by the language users.

Other questions remain, however, open, and the investigation of additional data is required to answer them. In what follows, I will mention but three, all of which require the analysis of larger data sets, and I will suggest that the convergence hypothesis could be tested by collecting new material for two other varieties of Insular Celtic.

The first question concerns gender agreement variation at the phonological and phonetic levels—if and to what extent variation depends on which type of sound undergoes which type of mutation. As far as language-internal factors are concerned, we noted in §1.2.2 a suggestion by Comrie (1979) that the greater the articulatory distance between the initial sound and its mutated counterpart, the greater the probability that the mutation will not take place; as regards contact-induced change, on the other hand, it could be hypothesized that failure to apply initial mutation is most likely in the case of sounds whose mutated counterparts have no phonemic status in the competing variety. However, the data sets I used were not large enough to allow a differential analysis of mutation rates across mutable segments.

The second question has to do with the investigation of gender agreement in relation to the decline of case marking in Irish. As explained in §§5.4.1.5 and 6.2.1.2, gender agreement may be marked word-finally on the attributive adjective in the genitive singular, but due to the decline of the attributive position and of genitive marking in spoken Irish there were simply too few instances of adjectives in the genitive singular to assess how productive word-final gender agreement marking is today. A much larger corpus of spoken Irish might help shed light on both issues.

Finally, the third question concerns the rate of grammatical agreement with anaphoric pronouns as a function of the different type of pronoun involved, as it has been suggested that grammatical agreement might be more frequent with simple pronominal forms than with prepositional ones (cf. §3.3.4); this suggestion seems to be at odds with Corbett's (1991: 238) statement that semantic agreement in the nominative is as likely or more likely than semantic agreement in oblique forms. However, it was observed

that this was not the case in either Irish or Welsh as far as the data in the older components were concerned; while in relation to the contemporary components, the number of prepositional forms did not allow for this difference to be tested. Likewise, not enough possessive forms were found in either the older or the contemporary components to test the hypothesis, also implicit in §3.3.4, that they would show agreement rates significantly different from those displayed by simple or prepositional pronouns on account of the different agreement-marking mechanisms involved. Again, a larger data set might allow to test these hypotheses as well.

To conclude, it should not escape our attention that there are two varieties of Insular Celtic, both on the Brythonic side, which are spoken in a context of bilingualism not with English, but with two Romance varieties, French and Spanish: Breton and Patagonian Welsh, respectively.<sup>1</sup> These two varieties are particularly important because they might provide the means to test the convergence hypothesis, given that both French and Spanish have a grammatical gender system considerably more similar to the traditional Insular Celtic one, with two genders (masculine and feminine) and syntactic agreement both within the noun phrase and pronominally. If it were possible to demonstrate that the traditional gender agreement system is better preserved today in these two spoken varieties, the convergence hypothesis would be significantly corroborated. The construction of a corpus of Patagonian Welsh is currently underway at the ESRC Centre for Research on Bilingualism in Theory & Practice at University of Wales Bangor,<sup>2</sup> and it is to be hoped that similar initiatives will be undertaken to record a substantial corpus of spoken Breton while such an enterprise is still possible.

---

<sup>1</sup>For Patagonian Welsh, see for instance Jones (1998b) and the introduction provided by Johnson (2009: §1).

<sup>2</sup>Cf. <http://siarad.org.uk/> for details and <http://siarad.org.uk/patagonia.php> for samples; both pages accessed on 29 November 2010.

# Appendix A

## The Irish data: further analysis

### A.1 Article agreement

#### A.1.1 Article agreement and syntactic case in OC-I

Since the genitive is marked in a different way from the nominative and the prepositional case (i.e., not only by means of lenition but also by a different form of the article), and because gender agreement in the prepositional cases is only possible for a limited subset of nouns (those beginning in lenitable *s-*), it is interesting to repeat the test for each syntactic case.

Table A.1: Agreement with the article in OC-I by case

|            | Nom.         |      | Gen.        |      | Prep.        |      |
|------------|--------------|------|-------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.       | Fem. | Masc.        | Fem. |
| Agreeing   | 173          | 78   | 40          | 28   | 14           | 1    |
| Total      | 175          | 80   | 42          | 28   | 15           | 2    |
| % agreeing | 99           | 97.5 | 95          | 100  | 93           | 50   |
|            | $p = 0.5918$ |      | $p = 0.513$ |      | $p = 0.2279$ |      |

As Table A.1 shows, if we consider only nominative noun phrases we get roughly the same agreement rates as when the overall data are considered, with no significant difference between the two proportions. In the genitive, the difference appears to be larger, but not significant for these data ( $p =$

0.513); in the prepositional case the difference in agreement rate is even larger but still tests as statistically not significant ( $p = 0.2279$ ) on account of the very small counts involved.

### A.1.2 Article agreement and syntactic case in CC-I

If we consider agreement rates for each syntactic cases, as shown in Table A.2, we see that the difference in agreement rates tests as statistically significant only in the nominative ( $p = 0.0035$ ), while there are too few feminines in the prepositional case to make any meaningful generalization.

Table A.2: Agreement with the article in CC-I by case

|            | Nom.         |      | Gen.         |      | Prep.        |      |
|------------|--------------|------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 124          | 91   | 36           | 53   | 16           | 2    |
| Total      | 129          | 106  | 36           | 55   | 17           | 4    |
| % agreeing | 96           | 86   | 100          | 96   | 94           | 50   |
|            | $p = 0.0035$ |      | $p = 0.5165$ |      | $p = 0.0797$ |      |

When the nominative is considered on its own, as can be seen from Table A.3, the difference is statistically significant in the non-RnaG sub-component ( $p = 0.0091$ ), but borderline significant, at best, in the RnaG sub-component ( $p = 0.071$ , c.i. 0.83 to 21.08). On the other hand, as shown in Tables A.2 and A.4, the difference is not statistically significant in the genitive singular.<sup>1</sup>

However, more than half of the 53 occurrences of the expected genitive singular *na* in CC-I include frequent and/or formulaic phrases, such as *na hÉireann* ‘of Ireland’ and *na Gaeilge* ‘of the Irish language’, *mí na Nollag* ‘December’, and the official names of *Raidió na Gaeltachta* and of a research institute (*Institiúid Taighde na Fiontraíochta Digití* ‘The Digital Enterprise Research Institute’): more examples are listed in Table A.5. Therefore, the productivity of *na* as a marker of gender agreement is open to question.

<sup>1</sup>As regards the prepositional case, the counts are too small to be relevant: there are only three feminine tokens in the RnaG sub-component and only one in the non-RnaG one.

Table A.3: Agreement with the article in CC-I, RnaG vs. others, nominative only

|            | RnaG        |         | Others       |         |
|------------|-------------|---------|--------------|---------|
|            | Nom. m.     | Nom. f. | Nom. m.      | Nom. f. |
| Agreeing   | 77          | 65      | 47           | 16      |
| Total      | 80          | 74      | 49           | 22      |
| % agreeing | 96          | 88      | 96           | 73      |
|            | $p = 0.071$ |         | $p = 0.0091$ |         |

Table A.4: Agreement with the article in CC-I, RnaG vs. others, genitive only

|            | RnaG    |         | Others  |         |
|------------|---------|---------|---------|---------|
|            | Gen. m. | Gen. f. | Gen. m. | Gen. f. |
| Agreeing   | 27      | 37      | 9       | 16      |
| Total      | 27      | 38      | 9       | 17      |
| % agreeing | 100     | 97      | 100     | 94      |
|            | $p = 1$ |         | $p = 1$ |         |

### A.1.3 The semantic factor

Since gender assignment in Irish can be either formally or semantically motivated, the question might be asked whether we would observe different agreement rates for feminine nouns if we did not consider cases of semantic agreement, that is if we excluded from the count all the feminine tokens that have semantically feminine (i.e., female) referents. However, only two occurrences of thus-defined semantic agreement were found, one in OC-I and one in CC-I, both of which show the expected agreement form. The impact of the semantic factor was therefore deemed negligible.

### A.1.4 Proper names

The comparison between the two components was carried out a second time after removing proper names from the samples (cf. §3.3.2). (More precisely, 13 such observations with feminine controllers were removed from CC-I, while none was found in OC-I.) By doing so, however, we do not alter the picture significantly (Table A.6); agreement rates are essentially unchanged

Table A.5: Genitive singular *na* in CC-I: frequent tokens and repetitions

| Times | Token         | Example  | Gloss  |
|-------|---------------|--|--|
| 5     | na cathrach   | lár na cathrach ( $\times 2$ )<br>fud fad na cathrach                    | the city centre<br>across the city               |
| 5     | na Mumhan     | timpeall na cathrach ( $\times 2$ )<br>Gaelainn na Mumhan ( $\times 4$ ) | around the city<br>Munster Irish                 |
| 4     | na háite      | Gaeilge na Mumhan<br>áilleacht na háite                                  | Munster Irish<br>the beauty of the place         |
| 3     | na bliana     | timpeall na háite ( $\times 3$ )<br>i rith na bliana                     | around the place<br>during the year              |
| 3     | na Gaeltachta | in aghaidh na bliana ( $\times 2$ )<br>Raidió na Gaeltachta              | per year<br>(name of a radio station)            |
| 3     | na Nollag     | aimsir na Nollag<br>roimis na Nollag                                     | Christmas time<br>before Christmas               |
| 2     | na Mí         | mí na Nollag<br>Contae na Mí   | December<br>County Meath                         |
| 2     | na hÉireann   | Ollscoil na hÉireann<br>stair na hÉireann                                | University of Ireland<br>the history of Ireland  |
| 2     | na Gaeilge    | trí mheán na Gaeilge<br>cúig lá na seachtaine                            | through the medium of Irish<br>the five weekdays |
| 2     | na seachtaine | i rith na seachtaine<br>os comhair na teilifíse                          | during the week<br>in front of the TV            |
| 2     | na teilifíse  | gnó na teilifíse   | the television business                          |
| Total |               | 33   |  |

and the difference between them is still statistically significant.

Table A.6: Article agreement with feminine controllers, OC-I vs. CC-I

|            | Overall      |           | Excluding proper names |           |
|------------|--------------|-----------|------------------------|-----------|
|            | Fem. (OC)    | Fem. (CC) | Fem. (OC)              | Fem. (CC) |
| Agreeing   | 107          | 136       | 107                    | 123       |
| Total      | 110          | 155       | 110                    | 142       |
| % agreeing | 97           | 88        | 97                     | 87        |
|            | $p = 0.0059$ |           | $p = 0.0029$           |           |

## A.2 Adjective agreement

### A.2.1 Proper names

The comparison between OC-I and CC-I was carried out a second time after excluding all occurrences of agreement forms which occurred as part of proper names. In OC-I we find seven such occurrences, like the two place names in (1) and (2).

(1) (Wigger, 2000 1-01-06)

Aill                      Bhuí  
 cliff(F).NOM.SG    yellow.NOM.F.SG

‘Yellow Cliff’

(2) (Wigger, 2000 1-01-09)

Cuan                      an                      Fhir                      Mhóir  
 harbour(M).NOM.SG    ART.GEN.M.SG    man(M).GEN.SG    big.GEN.M.SG

‘The Big Man’s Harbour’

All of these show agreement as expected. However, to exclude them from the count does not significantly alter the picture, as shown in Table A.7: the rate of agreement with masculines does not significantly differ from the rate of agreement with feminines ( $p = 0.67$ ) in OC-I, while it does in CC-I, where agreement with masculines is still more frequent than with feminines ( $p < 0.0001$ ).

Table A.7: Agreement with the adjective, excluding proper names

|            | OC-I       |      | CC-I         |      |
|------------|------------|------|--------------|------|
|            | Masc.      | Fem. | Masc.        | Fem. |
| Agreeing   | 47         | 28   | 49           | 8    |
| Total      | 50         | 31   | 54           | 19   |
| % agreeing | 94         | 90   | 90           | 42   |
|            | $p = 0.67$ |      | $p < 0.0001$ |      |

Excluding proper and place names, the difference between the two components in terms of agreement with feminine controllers is still statistically significant ( $p = 0.0007$ ), as shown in Table A.8.

Table A.8: Adjective agreement with feminine controllers, OC-I vs. CC-I, excluding proper names

|            | Fem. (OC)    | Fem. (CC) |
|------------|--------------|-----------|
| Agreeing   | 28           | 8         |
| Total      | 31           | 19        |
| % agreeing | 90           | 42        |
|            | $p = 0.0007$ |           |

## Appendix B

# The Welsh data: further analysis

### B.1 Article agreement

#### B.1.1 The semantic factor

If semantic agreement is excluded from the analysis (Table B.1), the difference in agreement rate remains statistically not significant in both OC-W ( $p = 0.0507$ ) and CC-W ( $p = 0.1928$ ).

Table B.1: Non-semantic agreement with the article

|            | OC           |      | CC           |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 160          | 45   | 245          | 64   |
| Total      | 160          | 47   | 247          | 65   |
| % agreeing | 100          | 96   | 99           | 98   |
|            | $p = 0.0507$ |      | $p = 0.1928$ |      |

The difference in agreement rate across the two components remains statistically not significant with both masculine and feminine controllers ( $p = 0.5217$  and  $p = 1$ , respectively).

### B.1.2 Proper names

As we have seen in §3.3.2, mutations occurring within the boundaries of a noun phrase that is also a proper name—such as (1) in OC-W (the title of a song) and (2) in CC-W (the name of a bird, quoted here as the title of a TV series)—may be regarded as unanalyzed wholes that speakers learn and store in the lexicon as such.

- (1) (Thomas & Thomas, 1989: 142 l. 4)

Tôn y botel (cf. *potel*)  
tune(F).SG ART.F.SG bottle(F).SG

‘the Bottle’s Tune’

- (2) (*sianthomas.cha* 199–200)

Glas y dorlan (cf. *torlan*)  
blue ART.F.SG bank(F).SG

‘The Kingfisher’

If such instances are excluded from the analysis along with semantic agreement, the difference in agreement rate between masculine and feminine gender is at best borderline significant in OC-W ( $p = 0.0459$ ), still not significant in CC-W (Table B.2).

Table B.2: Non-semantic agreement with the article, excluding proper names

|            | OC           |      | CC           |      |
|------------|--------------|------|--------------|------|
|            | Masc.        | Fem. | Masc.        | Fem. |
| Agreeing   | 152          | 40   | 239          | 57   |
| Total      | 152          | 42   | 241          | 59   |
| % agreeing | 100          | 95   | 99           | 97   |
|            | $p = 0.0459$ |      | $p = 0.1746$ |      |

The difference in agreement rate with feminine nouns across the two components remains statistically not significant ( $p = 1$ ), as shown in Table B.3.

Table B.3: Non-semantic article agreement with feminine controllers, OC-W vs. CC-W, excluding proper names

|            | Fem. (OC) | Fem. (CC) |
|------------|-----------|-----------|
| Agreeing   | 40        | 57        |
| Total      | 42        | 59        |
| % agreeing | 95        | 97        |
|            | $p = 1$   |           |

## B.2 Adjective agreement

### B.2.1 The semantic factor

Semantic agreement accounts for just two observations in OC-W and for 11 in CC-W. Only in one case, in CC-W, does a feminine noun with a female referent (*merch* ‘girl’) fail to trigger soft mutation onto a following adjective. If we disregard semantic agreement, as illustrated in Table B.4, agreement rates remain very similar to overall rates and the difference is still statistically significant ( $p = 0.0302$ ).

Table B.4: Adjective agreement with feminine controllers, OC-W vs. CC-W)

|            | Overall      |           | Excluding semantic agreement |           |
|------------|--------------|-----------|------------------------------|-----------|
|            | Fem. (OC)    | Fem. (CC) | Fem. (OC)                    | Fem. (CC) |
| Agreeing   | 26           | 47        | 25                           | 42        |
| Total      | 27           | 61        | 26                           | 55        |
| % agreeing | 96           | 77        | 96                           | 76        |
|            | $p = 0.0314$ |           | $p = 0.0302$                 |           |

### B.2.2 Proper names

Adjective agreement within proper names is found once in OC-W and twice in CC-W. Its contribution to agreement rates is marginal and does not significantly alter the picture, as shown in Table B.5.

Table B.5: Adjective agreement with feminine controllers, OC-W vs. CC-W, excluding semantic agreement and proper names

|              | Fem. (OC) | Fem. (CC) |
|--------------|-----------|-----------|
| Agreeing     | 24        | 40        |
| Total        | 25        | 53        |
| % agreeing   | 96        | 75        |
| $p = 0.0297$ |           |           |

### B.3 Agreement with demonstrative pronouns

#### B.3.1 The older component

Demonstrative pronouns are rarely found in OC-W: there are just 17 instances, 11 with masculine antecedents and six with feminine antecedents. Gender agreement is consistently observed. Semantic agreement accounts for two occurrences, one with a masculine and one with a feminine noun.

#### B.3.2 The contemporary component

Demonstrative pronouns are rather infrequent in CC-W too (21 occurrences altogether). As shown in Table B.6, agreement with masculine nouns is found 15 times out of 15 and agreement with feminine nouns four out of six. However, as also shown in Table B.6, semantic agreement accounts for two thirds of the pronouns referring to masculine nouns and half of those referring to feminine ones.

Table B.6: Agreement with demonstrative pronouns in CC-W

|            | Overall |      | Excl. sem. agr. |      |
|------------|---------|------|-----------------|------|
|            | Masc.   | Fem. | Masc.           | Fem. |
| Agreeing   | 15      | 4    | 5               | 2    |
| Total      | 15      | 6    | 5               | 4    |
| % agreeing | 100     | 67   | 100             | 50   |

## B.4 Agreement with numeral pronouns

Only in three cases were numeral pronouns found in cataphoric position, twice in OC-W and once in CC-W; in all three cases grammatical agreement was observed. The occurrence of cataphoric numeral pronouns was only observed in the partitive structure of the type represented by example (3), drawn from CC-W: in these types of structures the numeral pronoun is followed by a partitive prepositional phrase, where the controller is found (*drefydd* < *trefydd* in the example).

(3) (*jonsi.cha* 2058–2059)

un neu ddwy o drefydd yn y gogledd  
 one or two.F of town(F).PL in ART.M.SG north(M).SG

‘one or two towns in the north’

### B.4.1 The older component

There are only two occurrences of pronominal *un* ‘one’ where gender marking is possible, both with a feminine controller and presenting the expected mutation (e.g. *un dda* < *da* ‘a good one’): both are cases of semantic agreement, where reference is made to a woman.

There are 12 occurrences of numeral pronouns in the ‘2’–‘4’ range, ten with a masculine antecedent and all showing the expected masculine form, and two with a feminine controller, also showing the expected feminine form. Agreement is semantically motivated in two cases, both with masculine controllers.

### B.4.2 The contemporary component

There is only one occurrence of pronominally used *un*, which was considered uncertain as its antecedent, *stumog* ‘stomach’, is of oscillating gender according to Thomas (2002). The occurrence, recorded in CC-W, is shown in (4).

## (4) (gaynor.cha 949–951)

GAY: mae rhaid bo(d) gymnach chi stumog eitha  
 be.PRS.3 necessity be.VN with.2PL 2PL stomach(?).SG quite  
 cryf Rhys  
 strong(M) Rhys  
 ‘you must have quite a strong stomach, Rhys’

RHY: well am wn i mae gynna i un  
 well inasmuch as know.1SG 1SG be.PRS.3 with.1SG 1SG one.M.SG  
 da  
 good.M.SG  
 ‘well, as far as I know, I have got a good one’

It might be argued that *stumog* is considered masculine by these speakers, and that agreement is consistently masculine both within and outside the noun phrase: *cryf* belongs in the restricted set of adjectives marking gender agreement by vowel shift and in the noun phrase *stumog eitha cr-f* the feminine form of the adjective would be *cref*, with no mutation expected because of the adverb (*eitha*) intervening between the noun and the adjective. However, because no example of agreement by vowel shift was found in either component, and because this type of agreement is regarded as a literary trait, we cannot rely on the form *cryf* as a decisive argument. The masculine gender of pronominal *un* is revealed by the absence of soft mutation on the following adjective (*da*).

There are nine instances of gender-marking pronominal numerals other than *un*, of which four co-refer with masculine antecedents and five with feminine ones. The former four all agree with their masculine antecedents. Of the latter five, only four agree with their feminine antecedents, two of which are semantically feminine as they denote females. Somewhat surprisingly, the non-agreeing numeral pronoun in this group has a semantically feminine antecedent: this is *dau* in example (5), which refers to two members of a female quintet first introduced by their stage name and then referred to by the generic noun *merched* ‘girls’ (feminine plural).

## (5) (ecdau.cha 125–141)

DAF: felly be(th) am Girls Aloud?  
 DAF: wnest ti gyfarfod nhw (y)n iawn?  
 BBA: do wnes i gyfarfod (gy)da nhw.

BBA: a fi (y)n lico nhw.

DAF: oh.

BBA: ond [/] ond fi (we)di clywed bod nhw (y)n merched feisty iawn yn gallu gael bach o tantrum neu un [/] un neu ddau.

BBA: ond na dim yn fan (hyn)ny o gwbl.

BBA: ond <o'n nhw (y)n ffansïo Chr(is)> [//] oedd [?] un neu ddau yn ffansïo Chris.

*Translation:*

DAF: so what about Girls Aloud?

DAF: did you meet them properly?

BBA: yes, I met them.

BBA: and I like them.

DAF: oh.

BBA: but I have heard that they are very feisty girls, can have a little bit of tantrum, one or two.

BBA: but no, not there at all.

BBA: but they fancied Chris, one or two(M) fancied Chris.



# Bibliography

- Abney, S. P. 1987. *The English Noun Phrase in its sentential aspect*. Ph.D. thesis, M.I.T.
- Acquaviva, P. 2006. Goidelic inherent plurals and the morphosemantics of number. *Lingua*, 116, 1860–1887.
- Aikhenvald, A. Y. 2000. *Classifiers: A typology of noun categorization devices*. Oxford University Press, Oxford.
- Aikhenvald, A. Y. 2003. Mechanisms of change in areal diffusion: new morphology and language contact. *Journal of Linguistics*, 39, 1–29.
- Aikhenvald, A. Y. & Dixon, R. M. W. 1998. Dependencies between grammatical systems. *Language*, 74, 56–80.
- Aitchison, J. & Carter, H. 1994. *A geography of the Welsh language 1961–1991*. University of Wales Press, Cardiff.
- Akutagawa, M. 1987. A linguistic minority under the protection of its own ethnic state: a case study in an Irish Gaeltacht. In Mac Eoin, G., Ahlqvist, A., & Ó hAodha, D. (Eds.), *Third International Conference on Minority Languages: Celtic papers*, pp. 125–146. Multilingual Matters, Clevedon.
- Anderson, S. R. 1999. A formalist's reading of some functionalist work in syntax. In Darnell, M., Moravcsik, E., Newmeyer, F., Noonan, M., & Wheatley, K. (Eds.), *Functionalism and Formalism in Linguistics*, Vol. 1, pp. 111–135. John Benjamins, Amsterdam. URL of the cited version: <http://bloch.ling.yale.edu/Files/Milwaukee.pdf>, accessed 13 August 2011.
- Anderson, S. R. & Lightfoot, D. W. 1999. The human language faculty as an organ. *Annual Review of Physiology*, 62, 697–722. URL of the cited version: [http://bloch.ling.yale.edu/Files/Organ\\_annual\\_review.pdf](http://bloch.ling.yale.edu/Files/Organ_annual_review.pdf), accessed 13 August 2011.
- Aronoff, M. 1994. *Morphology by itself*. The MIT Press, Cambridge, Massachusetts.
- Audring, J. 2008. Gender assignment and gender agreement: evidence from pronominal gender languages. *Morphology*, 18(2), 93–116.

- Audring, J. 2009. *Reinventing pronoun gender*. Netherlands Graduate School of Linguistics / Landelijke – LOT. URL: <http://dare.ubvu.vu.nl/bitstream/1871/15413/5/8136.pdf>, accessed 28 December 2009.
- Awbery, G. M. 1986. Moves towards a simpler, binary mutation system in Welsh. In Mac Eoin, G., Ahlqvist, A., & Ó hAodha, D. (Eds.), *Sandhi phenomena in the languages of Europe*, pp. 161–166. Mouton de Gruyter, Berlin.
- Awbery, G. M. 1987. The position of the Welsh language in Wales. In Andersen, H. (Ed.), *Third International Conference on Minority Languages: Celtic papers*, pp. 1–12. Multilingual Matters, Clevedon.
- Ayoun, D. 2007. The second language acquisition of grammatical gender and agreement. In Ayoun, D. (Ed.), *French applied linguistics*, pp. 130–170. John Benjamins, Amsterdam and Philadelphia.
- Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press, Cambridge.
- Backus, A. 2004. Convergence as a mechanism of language change. *Bilingualism: Language and Cognition*, 7(2), 179–181.
- Baerman, M., Brown, D., & Corbett, G. G. 2005. *The syntax-morphology interface: A study of syncretism*. Cambridge University Press, Cambridge.
- Ball, M. J. 1982. Stylistic variation in radio broadcast: an introductory study. *Papurau Gwaith Ieithyddol Cymraeg Caerdydd/Cardiff Working Papers in Welsh Linguistics*, 2, 17–24.
- Ball, M. J. 1988. Accounting for linguistic variation: dialectology. In Ball, M. J. (Ed.), *The use of Welsh*, pp. 24–38. Multilingual Matters, Clevedon.
- Ball, M. J. 1993. Initial-consonant mutation in modern spoken Welsh. *Multilingua*, 12(2), 189–205.
- Ball, M. J. & Müller, N. 1992. *Mutation in Welsh*. Routledge, London.
- Bechert, J. 1990. The structure of the noun in European languages. In Bechert, J., Bernini, G., & Buridant, C. (Eds.), *Toward a typology of European languages*, pp. 115–140. Mouton de Gruyter, Berlin and New York.
- Berlin, B. & Kay, P. 1969. *Basic color terms: Their universality and evolution*. University of California Press, Berkeley.
- Boroditsky, L., Schmidt, L. A., & Phillips, W. 2003. Sex, syntax, and semantics. In Gentner, D. & Goldin-Meadow, S. (Eds.), *Language in mind: Advances in the study of language and thought*, pp. 61–79. The MIT Press, Cambridge, Massachusetts.

- Borsley, R. D., Tallerman, M., & Willis, D. 2007. *The syntax of Welsh*. Cambridge University Press, Cambridge.
- Breatnach, R. A. 1964. The Irish revival reconsidered. *Studies: An Irish Quarterly Review*, 53(209), 18–30.
- Breatnach, R. B. 1958. Initial mutation of substantives after preposition + singular article in Déise Irish. *Éigse*, 9, 217–222.
- Broderick, G. 1984. *A handbook of late spoken Manx*, Vol. 1. Grammar and texts. Max Niemeyer, Tübingen.
- Broderick, G. 1993. Manx. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 228–285. Routledge, London.
- Bullock, B. E. & Toribio, A. J. 2004. Introduction: convergence as an emergent property in bilingual speech. *Bilingualism: Language and Cognition*, 7(2), 91–93.
- Butler, C. S. 2003. *Structure and function: A guide to three major structural-functional theories. Part 1: Approaches to the simplex clause*. No. 63 in Studies in Language Companion Series. John Benjamins, Amsterdam and Philadelphia.
- Butler, C. S. 2008. Cognitive adequacy in structural-functional theories of language. *Language Sciences*, 30, 1–30.
- Butler, C. S. 2009. Criteria of adequacy in functional linguistics. *Folia Linguistica*, 43(1), 1–66.
- Butterworth, B. 2004. Lexical representation. In Katamba, F. (Ed.), *Morphology: Critical concepts in linguistics*, Vol. VI: Morphology: its place in the wider context, pp. 1–40. Routledge, London and New York. First appeared in *Language production*, edited by B. Butterworth (1983, London: Academic Press).
- Bybee, J. 2006. From usage to grammar: the mind's response to repetition. *Language*, 82(4), 711–733.
- Caighdeán 1979. *Gramadach na Gaeilge agus litriú na Gaeilge: An caighdeán oifigiúil* (6th edition). Oifig an tSoláthair, Baile Átha Cliath.
- Calder, G. 1990. *A Gaelic grammar*. Gairm Publications, Glasgow.
- Carnie, A. 2008. *Constituent structure*. Oxford University Press, Oxford.
- Chomsky, N. 1965. *Aspects of the theory of syntax*. The MIT Press, Cambridge, MA.

- Chomsky, N. 1980. *Rules and representations*. Columbia University Press, New York.
- Comrie, B. 1979. Morphophonemic exceptions and phonetic distance. *Linguistics*, 17(215/216), 51–60.
- Comrie, B. 1999. Grammatical gender systems: a linguist's assessment. *Journal of Psycholinguistic Research*, 28, 457–466.
- Consten, M., Knees, M., & Schwarz-Friesel, M. 2007. The function of complex anaphors in texts: evidence from corpus studies and ontological considerations. In Schwarz-Friesel, M., Consten, M., & Knees, M. (Eds.), *Anaphors in text: Cognitive, formal and applied approaches to anaphoric reference*, pp. 81–102. John Benjamins, Amsterdam and Philadelphia.
- Corbett, G. G. 1991. *Gender*. Cambridge University Press, Cambridge.
- Corbett, G. G. 1994. Gender and gender systems. In Asher, R. E. & Simpson, J. M. Y. (Eds.), *The encyclopaedia of language and linguistics*, Vol. 3, pp. 1347–1353. Pergamon Press, Oxford.
- Corbett, G. G. 2003. Types of typology, illustrated from gender systems. In Plank, F. (Ed.), *Noun phrase structure in the languages of Europe*, pp. 289–334. Mouton de Gruyter, Berlin and New York.
- Corbett, G. G. 2006. *Agreement*. Cambridge University Press, Cambridge.
- Cornish, F. 2009. *Text and discourse as context: Discourse anaphora and the FDG Contextual Component*. *Web Papers in Functional Discourse Grammar*, 82, 97–115. URL: [http://home.hum.uva.nl/fdg/working\\_papers/WP-FDG-82\\_Cornish.pdf](http://home.hum.uva.nl/fdg/working_papers/WP-FDG-82_Cornish.pdf), accessed 5 October 2010.
- Coupland, N. & Ball, M. J. 1989. Welsh and English in contemporary Wales: sociolinguistic issues. *Contemporary Wales*, 3, 7–40.
- Craig, C. (Ed.). 1986. *Noun classes and categorization*. John Benjamins, Amsterdam and Philadelphia.
- CSO 2006. *Census 2006*, Vol. 9: Irish language. Central Statistics Office Ireland. URL: [http://www.cso.ie/Census/census2006\\_volume9.htm](http://www.cso.ie/Census/census2006_volume9.htm), accessed 16 January 2010.
- Curzan, A. 2003. *Gender shifts in the history of English*. Cambridge University Press, Cambridge.
- Dahl, Ö. 1999. Animacy and the notion of semantic gender. In Unterbeck, B. & Rissanen, M. (Eds.), *Gender in grammar and cognition*, pp. 99–115. Mouton de Gruyter, Berlin and New York.

- Davies, C. 1988. Cymraeg Byw. In Ball, M. J. (Ed.), *The use of Welsh*, pp. 200–210. Multilingual Matters, Clevedon.
- de Bhaldraithe, T. 1953. *Gaeilge Chois Fhairrge: An deilbhíocht*. Institiúid Árd-Léinn Bhaile Átha Cliath, Baile Átha Cliath.
- De Vos, L. & De Vogelaer, G. In press. Dutch gender and the locus of morphological regularisation. *Folia Linguistica*, 45(2).
- Denny, J. P. & Creider, C. A. 1986. The semantics of noun classes in Proto-Bantu. In Craig, C. (Ed.), *Noun classes and categorization*, pp. 217–239. John Benjamins, Amsterdam and Philadelphia.
- Dik, S. C. 1978. *Functional grammar*. North-Holland Publishing Co., Amsterdam.
- Dik, S. C. 1986. On the notion “functional explanation”. *Belgian Journal of Linguistics*, 1, 11–52.
- Dik, S. C. 1989. *The theory of functional grammar*. Foris, Dordrecht.
- Dik, S. C. 1997. *The theory of functional grammar. Part 1. The structure of the clause. (Second, revised edition prepared by Kees Hengeveld.)*. Mouton de Gruyter, Berlin.
- Dikker, S. 2004. On the whereabouts of gender and number assignment: location and accessibility. *Working Papers in Functional Grammar*, pp. 35–51.
- Dixon, R. M. W. 1972. *The Dyirbal language of North Queensland*. Cambridge University Press, Cambridge.
- Dixon, R. M. W. & Aikhenvald, A. 2002. Word: a typological framework. In Dixon, R. M. W. & Aikhenvald, A. (Eds.), *Word: A cross-linguistic typology*, pp. 1–41. Cambridge University Press, Cambridge.
- Dorian, N. C. 1973. Grammatical change in a dying dialect. *Language*, 49(2), 413–438.
- Dorian, N. C. 1976a. Gender in a terminal Gaelic dialect. *Scottish Gaelic Studies*, 12, 279–282.
- Dorian, N. C. 1976b. A hierarchy of morphophonemic decay in Scottish Gaelic language death: the differential failure of lenition. *Word*, 28(1–2), 96–109.
- Dorian, N. C. 1978a. *East Sutherland Gaelic: The dialects of Brora, Golspie, and Embro fishing communities*. Dublin Institute for Advanced Studies, Dublin.
- Dorian, N. C. 1978b. The Fate of Morphological Complexity in Language Death: evidence from East Sutherland Gaelic. *Language*, 54(3), 590–609.
- Dorian, N. C. 1981. *Language death: The life cycle of a Scottish Gaelic dialect*. University of Pennsylvania Press, Philadelphia.

- Dorian, N. C. 1994. Purism vs. compromise in language revitalization and language revival. *Language in Society*, 23(4), 479–494.
- Erman, B. 2007. Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12(1), 25–53.
- Ferguson, C. A. 1959. Diglossia. *Word*, 15, 325–340.
- Fife, J. 1986. Literary vs. Colloquial Welsh: problems of definition. *Word*, 37(3), 141–150.
- Fife, J. 1993. Introduction. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 3–25. Routledge, London.
- Filppula, M. 2003. The quest for the most ‘parsimonious’ explanations: endogeny vs. contact revisited. In Hickey, R. (Ed.), *Motives for language change*, pp. 161–173. Cambridge University Press, Cambridge.
- Fodor, I. 1959. The origin of grammatical gender. *Lingua*, 8, 1–41, 186–214.
- Fortescue, M. 2007. The non-linearity of speech production. In Hannay, M. & Steen, G. J. (Eds.), *Structural-Functional Studies in English Grammar: In honour of Lachlan Mackenzie*, pp. 337–351. John Benjamins, Amsterdam and Philadelphia.
- Fowkes, R. A. 1957. Gender redistribution in Keltic: a preliminary study. In Pulgram, E. (Ed.), *Studies presented to Joshua Whatmough on his sixtieth birthday*, pp. 39–46. Mouton, 'S-Gravenhage.
- Fowkes, R. A. 1987. Brythonic gender reduction – the Cornish picture. In Cardona, G. & Zide, N. H. (Eds.), *Festschrift for Henry Hoenigswald*, pp. 121–129. Gunter Narr Verlag, Tübingen.
- Frenda, A. S. Forthcoming. Gender in Irish between continuity and change. *Folia Linguistica*, 45(2).
- Friedmann, N. & Biran, M. 2003. When is gender accessed? A study of paraphasias in Hebrew anomia. *Cortex*, 39, 441–463.
- García Velasco, D. & Rijkhoff, J. (Eds.). 2008. *The noun phrase in Functional Discourse Grammar*. Mouton de Gruyter, Berlin and New York.
- Gathercole, V. C. M., Thomas, E. M., & Laporte, N. 2001. The acquisition of grammatical gender in Welsh. *Journal of Celtic Language Learning*, 6(Special issue: *First language acquisition*), 53–87.
- Genee, I. 1998. *Sentential complementation in a functional grammar of Irish*. No. 7 in LOT International Series. Holland Academic Graphics, The Hague.

- George, K. 1993. Cornish. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 410–468. Routledge, London.
- George, K. & Broderick, G. 1993. The revived languages: modern Cornish and modern Manx. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 644–663. Routledge, London.
- Giannelli, L. 1997. Tuscany. In Maiden, M. & Parry, M. (Eds.), *The dialects of Italy*, pp. 297–302. Routledge, London and New York.
- Giannelli, L. & Cravens, T. D. 1997. Consonantal weakening. In Maiden, M. & Parry, M. (Eds.), *The dialects of Italy*, pp. 32–40. Routledge, London and New York.
- Gillies, W. 1993. Scottish Gaelic. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 145–227. Routledge, London.
- Green, A. D. 2006. The independence of phonology and morphology: the Celtic mutations. *Lingua*, 116, 1946–1985.
- Greenberg, J. H. 1978. How does a language acquire gender markers?. In Greenberg, J. H. (Ed.), *Universals of human language*, Vol. III: Word structure, pp. 47–82. Stanford University Press, Stanford.
- Hamp, E. P. 1951. Morphophonemes of the Keltic mutations. *Language*, 27(3), 230–247.
- Hannahs, S. J. & Tallerman, M. 2006. At the interface: selection of the Welsh definite article. *Linguistics*, 44(4), 781–816.
- Hannay, M. & Martínez Caro, E. 2008. Last things first. In Gómez González, M. A., Mackenzie, J. L., & González Álvarez, E. M. (Eds.), *Languages and cultures in contrast and comparison*, pp. 33–68. John Benjamins, Amsterdam and Philadelphia.
- Harder, P. 2008. Determiners and definiteness: Functional semantics and structural differentiation. In Müller, H. & Klinge, A. (Eds.), *Essays on Nominal Determination: From morphology to discourse management*, pp. 1–25. John Benjamins, Amsterdam and Philadelphia.
- Harris, A. C. & Campbell, L. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.
- Harris, J. & Murtagh, L. 1987. Irish and English in Gaeltacht primary schools. In Mac Eoin, G., Ahlqvist, A., & Ó hAodha, D. (Eds.), *Third International Conference on Minority Languages: Celtic papers*, pp. 104–124. Multilingual Matters, Clevedon.

- Hemon, R. 1975. *A historical morphology and syntax of Breton*. The Dublin Institute for Advanced Studies, Dublin.
- Hengeveld, K. 2004a. The architecture of a Functional Discourse Grammar. In Mackenzie, J. L. & Gómez-González, M. A. (Eds.), *A new architecture for Functional Grammar*, pp. 1–21. Mouton de Gruyter, Berlin and New York.
- Hengeveld, K. 2004b. Epilogue. In Mackenzie, J. L. & Gómez-González, M. A. (Eds.), *A new architecture for Functional Grammar*, pp. 365–378. Mouton de Gruyter, Berlin and New York.
- Hengeveld, K. & Mackenzie, J. L. 2008. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford University Press, Oxford.
- Hesp, C. 1990a. A critique of FG-CNLU. *Working Papers in Functional Grammar*, 35.
- Hesp, C. 1990b. The Functional Grammar Computational Natural Language User and psychological adequacy. In Nuyts, J., Bolkestein, A. M., & Vet, C. (Eds.), *Layers and Levels of Representation in Language Theory*, pp. 295–312. John Benjamins, Amsterdam and Philadelphia.
- Hickey, R. 1995. Sound change and typological shift: initial mutation in Celtic. In Fisiak, J. (Ed.), *Linguistic Typology and Reconstruction*, pp. 133–182. Mouton de Gruyter, Berlin.
- Hickey, R. 2003. Reanalysis and typological change. In Hickey, R. (Ed.), *Motives for language change*, pp. 258–278. Cambridge University Press, Cambridge.
- Hindley, R. 1990. *The death of the Irish language: A qualified obituary*. Routledge, London and New York.
- Hjelmslev, L. 1970. *Animé et inanimé, personnel et non-personnel* (2nd edition). Nordisk Sprog- og Kulturforlag, Copenhagen.
- Holmer, N. M. 1957. *The Gaelic of Arran*. Dublin Institute for Advanced Studies, Dublin.
- Holmer, N. M. 1962. *The Gaelic of Kintyre*. Dublin Institute for Advanced Studies, Dublin.
- Honselaar, W. & Keizer, E. 2009. Lexicon and frames in FDG: A treatment of Dutch *bekend zijn* ‘to be familiar, well known’, *behandelen* ‘to treat’ and *trouwen* ‘to marry’. *Lingua*, 119, 1212–1241.
- Hoppenbrouwers, C. 1983. Het genus in een Brabants regiolect. *TABU*, 13, 1–25.

- Hughes, A. 1994. Gaeilge Uladh. In McCone, K., McManus, D., Ó Háinle, C., Williams, N., & Breatnach, L. (Eds.), *Stair na Gaeilge*, chap. VIII, pp. 611–660. Roinn na Sean-Ghaeilge, Coláiste Phádraig, Maigh Nuad.
- Hurford, J. R. 1988. The interaction between numerals and nouns. In Plank, F. (Ed.), *Noun phrase structure in the languages of Europe*, pp. 561–620. Mouton de Gruyter, Berlin and New York.
- Jackson, K. 1942. Some mutations in Blasket Irish. *Éigse*, 3, 272–277.
- Jenner, H. 1904. *A handbook of the Cornish language chiefly in its latest stages with some account of its history and literature*. David Nutt, London.
- Johnson, I. 2009. How green is their valley? Subjective vitality of Welsh language and culture in the Chubut Province, Argentina. *International Journal of the Sociology of Language*, 2009(195), 141–171.
- Jones, B. M. 1993. *Ar lafar ac ar bapur: cyflwyniad i'r berthynas rhwng yr iaith lafar a'r iaith ysgrifenedig*. Y Ganolfan Astudiaethau Addysg, Aberystwyth.
- Jones, D. G. 1988. Literary Welsh. In Ball, M. J. (Ed.), *The use of Welsh*, pp. 125–171. Multilingual Matters, Clevedon.
- Jones, M. C. 1994. A tale of two dialects: standardization in modern spoken Welsh. In Parry, M. M., Davies, W. V., & Temple, R. A. M. (Eds.), *The changing voices of Europe: Social and political changes and their linguistic repercussions, past, present and future*, pp. 243–264. University of Wales Press, Cardiff.
- Jones, M. C. 1998a. *Language obsolescence and revitalization: Linguistic change in two sociolinguistically contrasting Welsh communities*. Clarendon Press, Oxford.
- Jones, R. O. 1998b. The Welsh language in Patagonia. In Jenkins, G. H. (Ed.), *Language and community in the nineteenth century*, chap. 10, pp. 287–316. University of Wales Press, Cardiff.
- Kallen, J. L. 1979. Initial mutation in Modern Irish. Unpublished M.A. thesis, University of Washington.
- Kallen, J. L. 1986. *Linguistic fundamentals for Hiberno-English syntax*. Ph.D. thesis, Trinity College Dublin.
- Karmiloff-Smith, A. 1979. *A functional approach to child language: a study of determiners and reference*. Cambridge University Press, Cambridge.
- Keizer, E. & Honselaar, W. 2009. A Functional Discourse Grammar account of set nouns in Dutch and its implications for lexicography. *International Journal of Lexicography*, 22(4), 361–397.

- Kempe, V. & Brooks, P. J. 2005. The role of diminutives in the acquisition of Russian gender: can elements of child-directed speech aid in learning morphology?. *Learning Language*, 55, 139–176.
- Kintsch, W. 1988. The role of knowledge in discourse comprehension: a construction–integration model. *Psychological Review*, 95, 163–182.
- Kintsch, W. 1998. *Comprehension: A paradigm for cognition*. Cambridge University Press, New York.
- Kirchner, R. 2004. Consonant lenition. In Hayes, B., Kirchner, R., & Steriade, D. (Eds.), *Phonetically based phonology*, pp. 313–345. Cambridge University Press, Cambridge.
- Kneen, J. J. 1973. *A grammar of the Manx language*. Yn Cheshaght Ghailckagh (The Manx Gaelic Society).
- Kortlandt, F. 1982. Phonemicization and rephonemicization of the Old Irish mutations. *Ériu*, 33, 73–83.
- Kuter, L. 1989. Breton vs. French: language and the opposition of political, economic, and cultural values. In Dorian, N. C. (Ed.), *Investigating obsolescence: Studies in language contraction and death*, pp. 75–89. Cambridge University Press, Cambridge.
- Labov, W. 1972. *Sociolinguistic patterns*. University of Pennsylvania Press, Philadelphia.
- Labov, W. 1994. *Principles of linguistic change*, Vol. 1: Internal factors. Blackwell, Oxford UK & Cambridge USA.
- Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago University Press, Chicago.
- Langacker, R. W. 1987. *Foundations of cognitive grammar*, Vol. I: Theoretical prerequisites. Stanford University Press, Stanford.
- Lehmann, C. 1988. On the function of agreement. In Ferguson, C. A. & Barlow, M. (Eds.), *Agreement in natural language: Approaches, theories, descriptions*, pp. 55–65. Center for the Study of Language and Information, Stanford, CA.
- Leiss, E. 1999. Gender in Old High German. In Unterbeck, B. & Rissanen, M. (Eds.), *Gender in grammar and cognition*, pp. 237–258. Mouton de Gruyter, Berlin and New York.
- Levelt, W. J. M. 1989. *Speaking: From intention to articulation*. The MIT Press, Cambridge, Massachusetts.

- Levelt, W. J. M. 1999. Producing spoken language: a blueprint of the speaker. In Brown, C. M. & Hagoort, P. (Eds.), *The Neurocognition of Language*, pp. 83–122. Oxford University Press, Oxford.
- Levinson, S. C. 1997. From outer to inner space: linguistic categories and non-linguistic thinking. In Nuyts, J. & Pederson, E. (Eds.), *Language and Conceptualization*, pp. 13–45. Cambridge University Press, Cambridge.
- Lewis, H. & Pedersen, H. 1974. *A concise comparative Celtic grammar* (3rd edition). Vandenhoeck & Ruprecht, Göttingen.
- Lewis, M. P. 2009. *Ethnologue: Languages of the World* (Sixteenth edition). SIL International, Dallas, Tex. Online version: <http://www.ethnologue.com/>, accessed 16 November 2010.
- Lieber, R. 1983. New developments in autosegmental morphology: consonant mutation. In Barlow, M., Ferguson, C. A., Flickinger, D., & Wescoat, M. (Eds.), *Proceedings of the West Coast Conference on Formal Linguistics*, pp. 165–175. Stanford University Press, Stanford, CA.
- Lightfoot, D. 1997. Catastrophic change and learning theory. *Lingua*, 100, 171–192.
- Lleó, C., Cortés, S., & Benet, A. 2008. Contact-induced phonological changes in the Catalan spoken in Barcelona. In Siemund, P. & Kintana, N. (Eds.), *Language contact and contact languages*, pp. 185–212. John Benjamins, Amsterdam and Philadelphia.
- Loporcaro, M. 1997. Lengthening and *raddoppiamento fonosintattico*. In Maiden, M. & Parry, M. (Eds.), *The dialects of Italy*, pp. 41–51. Routledge, London and New York.
- Lucas, D. W. (Ed.). 1968. *Aristotle: Poetics. Introduction, commentary and appendices by D. W. Lucas*. Clarendon Press, Oxford.
- Mac Congáil, N. 2004. *Irish grammar book*. Cló Iar-Chonnachta, Indreabhán, Conamara.
- Mac Eoin, G. 1993. Irish. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 101–144. Routledge, London.
- MacAulay, D. 1992. The Scottish Gaelic language. In MacAulay, D. (Ed.), *The Celtic languages*, pp. 137–248. Cambridge University Press, Cambridge.
- Macnamara, J. 1971. Successes and failures in the movement for the restoration of Irish. In Rubin, J. & Jernudd, B. H. (Eds.), *Can language be planned? Sociolinguistic theory and practice for developing nations*, pp. 65–94. The University Press of Hawaii, Honolulu.

- MacWhinney, B. 2010. A tale of two paradigms. In Kail, M. & Hickmann, M. (Eds.), *Language Acquisition across Linguistic and Cognitive Systems*, pp. 17–32. John Benjamins, Amsterdam.
- Maguire, G. 1991. *Our own language: An Irish initiative*. Multilingual Matters, Clevedon.
- Mark, C. 2003. *The Gaelic-English dictionary*. Routledge, London and New York.
- Matthews, P. 2003. On change in ‘E-language’. In Hickey, R. (Ed.), *Motives for language change*, pp. 7–17. Cambridge University Press, Cambridge.
- McCone, K. 1994. An tSean-Ghaeilge agus a réamhstair. In McCone, K., McManus, D., Ó Háinle, C., Williams, N., & Breatnach, L. (Eds.), *Stair na Gaeilge*, chap. II, pp. 61–219. Roinn na Sean-Ghaeilge, Coláiste Phádraig, Maigh Nuad.
- McGahan, C. 2009. *Language obsolescence and language death: the Irish language in south-east Ulster*. Ph.D. thesis, School of Languages, Literatures and Performing Arts, The Queen’s University of Belfast, Belfast.
- Meillet, A. 1948. *Linguistique historique et linguistique générale*. Champion, Paris.
- Meurig Evans, H. & Thomas, W. O. (Eds.). 2007. *Y geiriadur mawr: The complete Welsh-English, English-Welsh dictionary*. 27th ed. Gwasg Gomer and Christopher Davies (Cyhoeddwyr) Cyf.
- Mithun, M. 1999. *The Languages of Native North America*. Cambridge University Press, Cambridge.
- Mittendorf, I. & Sadler, L. 2006. A treatment of Welsh initial mutation. In Butt, M. & King, T. H. (Eds.), *Proceedings of the LFG06 Conference*, pp. 344–364. CSLI Publications. URL: <http://csli-publications.stanford.edu/LFG/11/lfg06.html>, accessed 6 August 2011.
- Moravcsik, E. 2003. Inflectional morphology in the Hungarian noun phrase: a typological assessment. In Plank, F. (Ed.), *Noun phrase structure in the languages of Europe*. Mouton de Gruyter, Berlin and New York.
- Moseley, C. (Ed.). 2010. *Atlas of the World’s Languages in Danger* (3rd edition). UNESCO Publishing, Paris. Online version: <http://www.unesco.org/culture/en/endangeredlanguages/atlas>, accessed 29 November 2010.
- Mulally, R. 1994. The internal structure of the noun phrase in Scottish Gaelic. In *The noun phrase sketch book*, EUROTYP working papers No. VII/22. Konstanz.
- Nesset, T. 2006. Gender meets the Usage-Based Model: four principles of rule interaction in gender assignment. *Lingua*, 116, 1369–1393.

- Newmeyer, F. J. 2003. Formal and functional motivation for language change. In Hickey, R. (Ed.), *Motives for language change*, pp. 18–36. Cambridge University Press, Cambridge.
- Ní Chasaide, A. 1999. Irish. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, pp. 111–116. Cambridge University Press, Cambridge.
- Nic Pháidín, C. 2003. ‘Cén fáth nach?’—Ó chanúint go criól. In Ní Mhianáin, R. (Ed.), *Idir lúibíní: Aistí ar an léitheoracht agus ar an litearthacht*, pp. 113–130. Cois Life, Baile Átha Cliath.
- NIG 2004. *New Irish grammar by the Christian Brothers*. C J Fallon, Dublin.
- Nuyts, J. 1992. *Aspects of a Cognitive-Pragmatic Theory of Language*. John Benjamins, Amsterdam and Philadelphia.
- Nuyts, J. 2001. *Epistemic Modality, Language, and Conceptualization: A Cognitive-Pragmatic Perspective*. John Benjamins, Amsterdam and Philadelphia.
- Ó Baoill, D. P. 1986. *Lárchanúint don Ghaeilge*. Institiúid Teangeolaíochta Eireann, Baile Átha Cliath.
- Ó Baoill, D. P. 1988. Language planning in Ireland: the standardization of Irish. In Ó Riagáin, P. (Ed.), *Language planning in Ireland (International Journal of the Sociology of Language, 70)*, pp. 109–126. Mouton de Gruyter, Berlin.
- Ó Curnáin, B. 2007a. *The Irish of Iorras Aithneach County Galway*, Vol. I. Dublin Institute for Advanced Studies, Dublin.
- Ó Curnáin, B. 2007b. *The Irish of Iorras Aithneach County Galway*, Vol. II. Dublin Institute for Advanced Studies, Dublin.
- Ó Curnáin, B. 2007c. *The Irish of Iorras Aithneach County Galway*, 4 vols. Dublin Institute for Advanced Studies, Dublin.
- Ó Dochartaigh, C. 1992. The Irish language. In MacAulay, D. (Ed.), *The Celtic languages*, pp. 11–99. Cambridge University Press, Cambridge.
- Ó Dónaill, N. 1977. *Foclóir Gaeilge-Béarla*. Roinn Oideachais agus Eolaíochta, Baile Átha Cliath.
- Ó Giollagáin, C. & Mac Donnacha, S. 2008. The Gaeltacht today. In Nic Pháidín, C. & Ó Cearnaigh, S. (Eds.), *A new view of the Irish language*, pp. 108–120. Cois Life, Dublin.
- Ó hUiginn, R. 1994. Gaeilge Chonnacht. In McCone, K., McManus, D., Ó Háinle, C., Williams, N., & Breatnach, L. (Eds.), *Stair na Gaeilge*, chap. VII, pp. 539–609. Roinn na Sean-Ghaeilge, Coláiste Phádraig, Maigh Nuad.

- Ó Murchú, M. 1969. Common core and underlying forms. *Ériu*, 21, 42–75.
- Ó Murchú, M. 1988. Diglossia and interlanguage contact in Ireland. *Language, Culture and Curriculum*, 1(3), 243–249.
- Ó Raghallaigh, B. 2010. *Multi-dialect phonetisation for Irish text-to-speech synthesis: a modular approach*. Ph.D. thesis, Trinity College Dublin.
- Ó Sé, D. 2000. *Gaeilge Chorca Dhuibhne*. Institiúid Teangeolaíochta Eireann, Baile Átha Cliath.
- Ó Siadhail, M. 1984. A note on gender and pronoun substitution in modern Irish dialects. *Ériu*, 35, 173–177.
- Ó Siadhail, M. 1989. *Modern Irish: Grammatical structure and dialectal variation*. Cambridge University Press, Cambridge.
- Oftedal, M. 1962. A morphemic evaluation of the Celtic initial mutations. *Lochlann*, 2, 93–102.
- Phillips, J. D. 2004. *Manx*. Lincom Europa, München.
- Plank, F. 1994. Teaching myself noun phrases. In *The noun phrase sketch book*, EURO TYP working papers No. VII/22. Konstanz.
- Plaster, K. & Polinsky, M. 2007. Women are not dangerous things: gender and categorization. *Harvard Working Papers in Linguistics*, 12. URL: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:3209556>, accessed 23 March 2010.
- Polinsky, M. 2008. Gender under incomplete acquisition: heritage speakers' knowledge of noun categorization. *Heritage Language Journal*, 6(1), 40–71.
- Punch, A. 2008. Census data on the Irish language. In Nic Pháidín, C. & Ó Cearnaigh, S. (Eds.), *A new view of the Irish language*, pp. 43–54. Cois Life, Dublin.
- Regier, T. & Gahl, S. 2004. Learning the unlearnable: The role of missing evidence. *Cognition*, 93, 147–155.
- Rice, C. 2006. Optimizing gender. *Lingua*, 116, 1394–1417.
- Rijkhoff, J. 2002. *The noun phrase*. Oxford University Press, Oxford.
- Rijkhoff, J. 2008. Layers, levels and contexts in Functional Discourse Grammar. In García Velasco, D. & Rijkhoff, J. (Eds.), *The noun phrase in Functional Discourse Grammar*, pp. 63–115. Mouton de Gruyter, Berlin and New York.
- Roberts, I. G. 1993. *Verbs and diachronic syntax*. Kluwer, Dordrecht.
- Robins, R. H. 1997. *A short history of linguistics* (4th edition). Longman, London and New York.

- Rohlf, G. 1966. *Grammatica storica della lingua italiana e dei suoi dialetti*, Vol. I: Fonetica. Einaudi, Torino.
- Sadler, L. 1997. Clitics and the structure-function mapping. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.54.5865&rep=rep1&type=pdf>, accessed 13 December 2010..
- Sankoff, G. 2001. Linguistic outcomes of language contact. URL: <http://www.ling.upenn.edu/~gillian/Interlang.doc.pdf>, accessed 16 September 2009..
- Sayers, D. 2004. How narrow is narrowcasting? Are regional dialects standardised for national television?. URL: [http://privatewww.essex.ac.uk/~dsayers/academic/Sayers\\_narrowcasting.pdf](http://privatewww.essex.ac.uk/~dsayers/academic/Sayers_narrowcasting.pdf), accessed 19 December 2006.
- Sayers, D. 2005. Standardising diversity—The Language Revival Paradox: Can the Cornish language be revived without inhibiting its diversity?. URL [http://privatewww.essex.ac.uk/~dsayers/academic/Sayers\\_standardising\\_diversity.pdf](http://privatewww.essex.ac.uk/~dsayers/academic/Sayers_standardising_diversity.pdf), accessed 19 December 2006.
- Schmidt, A. 1985. *Young people's Dyirbal: An example of language death from Australia*. Cambridge University Press, Cambridge.
- Siemund, P. 2008. *Pronominal gender in English*. Routledge Studies in Germanic Linguistics. Routledge (Taylor and Francis), New York.
- Sinclair, J. 2005. Corpus and text – basic principles. In Wynne, M. (Ed.), *Developing linguistic corpora: A guide to good practice*, pp. 1–16. Oxbow Books, Oxford.
- Singer, R. 2010. Creativity in the use of gender agreement in Mawng: how the discourse functions of a gender system can approach those of a classifier system. *Studies in Language*, 34(2), 383–416.
- Sorace, A. 2004. Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7(2), 143–145.
- Steele, S. 1978. Word order variation: a typological study. In Greenberg, J. H. (Ed.), *Universals of human language*, Vol. IV: Syntax, pp. 585–623. Stanford University Press, Stanford.
- Stemberger, J. P. & MacWhinney, B. 2004. Are inflected forms stored in the lexicon?. In Katamba, F. (Ed.), *Morphology: Critical concepts in linguistics*, Vol. VI: Morphology: its place in the wider context, pp. 107–122. Routledge, London and New York. First appeared in *Theoretical morphology: Approaches in*

- modern linguistics*, edited by M. Hammond and M. Noonan (1988, London: Academic Press).
- Stephens, J. 1993. Breton. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 349–409. Routledge, London.
- Stewart, Jr., T. W. 2004. *Mutation as morphology: bases, stems, and shapes in Scottish Gaelic*. Ph.D. thesis, Ohio State University. URL: <http://www.ohiolink.edu/etd/send-pdf.cgi?osu1086046888>.
- Stokhof, M. & van Lambalgen, M. 2011. Abstractions and idealisations: The construction of modern linguistics. *Theoretical Linguistics*, 37, 1–26.
- SurrIDGE, M. 1989. Factors in the assignment of grammatical gender in Welsh. *Études Celtiques*, 26, 187–209.
- Taft, M. & Meunier, F. 1998. Lexical representation of gender: a quasiregular domain. *Journal of Psycholinguistic Research*, 27, 23–45.
- Tainturier, M. J., Leek, E. C., Schiemenz, S., Williams, C. A., Thomas, E., & Gathercole, V. C. 2005. Access to lexical phonology does not predict retrieval of grammatical gender in Welsh: implications for theories of language production. *Brain and Language*, 95, 52–53.
- Taylor, J. R. 1995. *Linguistic categorization: Prototypes in linguistic theory*. Clarendon Press, Oxford.
- Ternes, E. 1992. The Breton language. In MacAulay, D. (Ed.), *The Celtic languages*, pp. 371–452. Cambridge University Press, Cambridge.
- Thomas, A. R. 1992a. The Cornish language. In MacAulay, D. (Ed.), *The Celtic languages*, pp. 346–369. Cambridge University Press, Cambridge.
- Thomas, A. R. 1992b. The Welsh language. In MacAulay, D. (Ed.), *The Celtic languages*, pp. 251–345. Cambridge University Press, Cambridge.
- Thomas, B. & Thomas, P. W. 1989. *Cymraeg, Cymrâg, Cymrêg ... Cyflwyno'r Tafodieithoedd*. Gwasg Taf, Caerdydd.
- Thomas, E. M. 2001. *Aspects of gender mutation in Welsh*. Ph.D. thesis, University of Wales.
- Thomas, E. M. & Gathercole, V. C. M. 2005. Minority language survival: obsolescence or survival for Welsh in the face of English dominance?. In Cohen, J., McAlister, K. T., Rolstad, K., & MacSwan, J. (Eds.), *Proceedings of the 4th International Symposium on Bilingualism*, pp. 2233–2257. Cascadilla Press, Somerville, Ma.
- Thomas, R. J. (Ed.). 1950–2002. *Geiriadur Prifysgol Cymru: A dictionary of the Welsh language*. Gwasg Prifysgol Cymru, Caerdydd.

- Thomson, R. L. 1992. The Manx language. In MacAulay, D. (Ed.), *The Celtic languages*, pp. 100–136. Cambridge University Press, Cambridge.
- Thorne, D. A. 1993. *A comprehensive Welsh grammar*. Blackwell, Oxford.
- Thurneysen, R. 1980. *A grammar of Old Irish. Revised and enlarged edition with supplement*. Dublin Institute for Advanced Studies, Dublin.
- Toribio, A. J. 2004. Convergence as an optimization strategy in bilingual speech: evidence from code switching. *Bilingualism: Language and Cognition*, 7(2), 165–173.
- Trudgill, P. 1999. Language contact and the function of linguistic gender. *Poznań Studies in Contemporary Linguistics*, 35, 133–152. URL: <http://ifa.amu.edu.pl/psicl/files/35/10Trudgill.pdf>, accessed 12 March 2010.
- Tsimpli, I., Sorace, A., Heycock, C., & Filiaci, F. 2004. First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8(3), 257–277. URL: <http://ijb.sagepub.com/cgi/content/abstract/8/3/257>, accessed 11 April 2010.
- Tucker, G. R., Lambert, W. E., & Rigault, A. A. 1977. *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. Mouton, The Hague.
- Ua Súilleabháin, S. 1994. Gaeilge na Mumhan. In McCone, K., McManus, D., Ó Háinle, C., Williams, N., & Breatnach, L. (Eds.), *Stair na Gaeilge*, chap. VI, pp. 479–538. Roinn na Sean-Ghaeilge, Coláiste Phádraig, Maigh Nua.
- Unterbeck, B. & Rissanen, M. (Eds.). 1999. *Gender in grammar and cognition*. Mouton de Gruyter, Berlin and New York.
- van Berkum, J. J. A. 1996. *The psycholinguistics of grammatical gender: Studies in language comprehension and production*. Doctoral Dissertation, Max Planck Institute for Psycholinguistics. Nijmegen University Press, Nijmegen, Netherlands.
- Van Valin, Jr., R. D. & LaPolla, R. J. 1997. *Syntax: Structure, meaning, and function*. Cambridge University Press, Cambridge.
- Vendryes, J. 1921. *Le langage*. Le Renaissance du Livre, Paris.
- Vogel, P. M. 1999. Nominal abstracts and gender in Modern German: A “quantitative” approach towards the function of gender. In Unterbeck, B. & Rissanen, M. (Eds.), *Gender in grammar and cognition*, pp. 461–493. Mouton de Gruyter, Berlin and New York.

- Watkins, C. 1999. Two Celtic notes. In Anreiter, P. & Jerem, E. (Eds.), *Studia Celtica et Indogermanica: Festschrift für Wolfgang Meid zum 70. Geburtstag*, pp. 539–544. Archaeolingua, Budapest.
- Watkins, T. A. 1993. Welsh. In Ball, M. J. & Fife, J. (Eds.), *The Celtic languages*, pp. 289–348. Routledge, London.
- Watson, S. 1989. Scottish and Irish Gaelic: the giant's bed-fellows. In Dorian, N. C. (Ed.), *Investigating obsolescence: Studies in language contraction and death*, pp. 41–59. Cambridge University Press, Cambridge.
- Weber, D. 1999. On the function of gender. In Unterbeck, B. & Rissanen, M. (Eds.), *Gender in grammar and cognition*, pp. 495–509. Mouton de Gruyter, Berlin and New York.
- Wechsler, S. & Zlatić, L. 2003. *The many faces of agreement*. Center for the Study of Language and Information, Stanford, Calif.
- Wigger, A. (Ed.). 2000. *Caint Chonamara: Bailiúchán Hans Hartmann*. Imleabhar IX. *Ros Muc*. Universität Bonn.
- Williams, S. J. 1980. *Elfennau gramadeg Cymraeg*. Gwasg Prifysgol Cymru, Caerdydd.
- Zubin, D. & Köpke, K. M. 1986. Gender and folk taxonomy: the indexical relation between grammatical and lexical categorization. In Craig, C. (Ed.), *Noun classes and categorization*, pp. 139–180. John Benjamins, Amsterdam and Philadelphia.