

Elliott Lash, Fangzhe Qiu, David Stifter (Eds.)

Morphosyntactic Variation in Medieval Celtic Languages

Trends in Linguistics Studies and Monographs



Editors

Chiara Gianollo

Daniël Van Olmen

Editorial Board

Walter Bisang

Tine Breban

Volker Gast

Hans Henrich Hock

Karen Lahousse

Natalia Levshina

Caterina Mauri

Heiko Narrog

Salvador Pons

Niina Ning Zhang

Amir Zeldes

Editor responsible for this volume

Daniël Van Olmen

Volume 346

Morphosyntactic Variation in Medieval Celtic Languages



Corpus-Based Approaches

Edited by

Elliott Lash, Fangzhe Qiu, David Stifter

DE GRUYTER
MOUTON

This book was written as part of the project *Chronologicon Hibernicum*. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 647351). The editors of this volume also thank the Maynooth University Publications Fund for providing financial support for the publication of this volume in March, 2020, and the National University of Ireland Publications Scheme for providing financial support in July, 2020.



ISBN 978-3-11-068066-9

e-ISBN (PDF) 978-3-11-068074-4

e-ISBN (EPUB) 978-3-11-068079-9

DOI <https://doi.org/10.1515/9783110680744>



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. For details go to <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Library of Congress Control Number: 2020940693

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2020 Elliott Lash, Fangzhe Qiu, David Stifter, published by Walter de Gruyter GmbH, Berlin/Boston

The book is published open access at www.degruyter.com.

Typesetting: Integra Software Services Pvt. Ltd.

Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Contents

List of contributors — VII

Overview of linguistic annotation — XI

Elliott Lash, Fangzhe Qiu, and David Stifter

Introduction: Celtic Studies and Corpus Linguistics — 1

Part 1: Corpus tools for historical Celtic linguistics

Marius L. Jøhndal

1 Treebanks for historical languages and scalability — 15

Marieke Meelen

2 Annotating Middle Welsh: POS tagging and chunk-parsing a corpus of native prose — 27

Theodorus Fransen

3 Automatic morphological analysis and interlinking of historical Irish cognate verb forms — 49

Christopher Guy Yocum

4 Text clustering and methods in the *Book of Leinster* — 85

Part 2: Morphosyntactic variation and change in medieval Celtic languages

Liam Breatnach

5 The demonstrative pronouns in Old and Middle Irish — 115

Carlos García-Castillero

6 Paradigmatic split and merger: The descriptive and diachronic problem of Old Irish Class B infixed pronouns — 143

Elisa Roma

- 7 Nasalisation after inflected nominals in the Old Irish glosses:
Evidence for variation and change — 179**

Jürgen Uhlich

- 8 On the obligatory use of a nasalising relative clause after an
adjectival antecedent in the Old Irish glosses — 195**

Aaron Griffith

- 9 The “Cowgill particle”, preverbal *ceta* ‘first’, and prepositional cleft
sentences in the Old Irish glosses — 239**

Britta Irlslinger

- 10 The functions and semantics of Middle Welsh *X hun(an)*:
A quantitative study — 269**

Joseph F. Eska and Benjamin Bruch

- 11 Prolegomena to the diachrony of Cornish syntax — 313**

References — 339

Index — 365

List of contributors

The papers in this book arose from lectures given by the following contributors, hosted at the Department of Early Irish, Maynooth University by the project *Chronologicon Hibernicum* (Maynooth University, ERC Consolidator Grant 2015, H2020 #647351).

Liam Breatnach, MRIA is a Senior Professor at the School of Celtic Studies, Dublin Institute for Advanced Studies, and co-editor of the journal *Ériu*, published by the Royal Irish Academy. His main research interests are Old Irish, Middle Irish and the historical development of Irish, law texts, and poets, poetry and metrics. A recent publication on Early Irish law is *Córus Bésnái* (Breatnach 2017a). A recent publication on poetry and metrics is about the *Trefocal* tract (Breatnach 2017b). A recent publication on language is “Lebor na hUidre: Some linguistic aspects” (Breatnach 2015).

Benjamin Bruch teaches literature, history, and linguistics at the Pacific Buddhist Academy (Honolulu, Hawai’i). His research interests include metrics and versification, the historical phonology and syntax of the Celtic languages, and the preservation and revitalisation of endangered languages. Bruch received a Ph.D. in Celtic Languages and Literatures at Harvard University with a dissertation titled *Cornish verse forms and the evolution of Cornish prosody, c. 1350–1611* (Bruch 2005). He has published on medieval Cornish literature in verse: “Medieval Cornish versification: An overview” (Bruch 2009). He has also worked on Cornish historical phonology: “Nucleus length and vocalic alternation in Cornish diphthongs” (Bock and Bruch 2010) and “New perspectives on vocalic alternation in Cornish” (Bock and Bruch 2012). Bruch is also a co-author of the new standard orthography for Revived Cornish: *An outline of the standard written form of Cornish* (Bock and Bruch 2008), officially adopted by the Cornish Language Partnership.

Carlos García-Castillero teaches various subjects related to Indo-European linguistics and historical and comparative linguistics at the Faculty of Arts of the University of the Basque Country. He studied Classical Philology at the University of the Basque Country and wrote his Ph.D. thesis on Indo-European linguistics (*La formación del tema de presente primario osco-umbro*; García-Castillero 1999) under the supervision of Prof. Jürgen Untermann and Prof. Joaquín Gorrochategui. His main fields of research are comparative Indo-European linguistics (with several papers on pronominal and verbal morphosyntax), Old Irish morphosyntax (illocutionary force and clause types, templatic character of the verbal complex), and pragmatics, as well as diachronic linguistics. His publications can be found in a wide range of journals, such as *Ériu*, *Zeitschrift für celtische Philologie*, *Historische Sprachforschung*, *Indogermanische Forschungen*, *Journal of Historical Pragmatics*, and *Diachronica*.

Joseph F. Eska is Professor of Linguistics at the Department of English at Virginia Polytechnic Institute and State University. He has worked on all aspects of the linguistic history of the Celtic languages, in particular on the ancient Celtic languages of Continental Europe. He is the author (with Don Ringe) of *Historical linguistics: Toward a twenty-first century reintegration* (Ringe and Eska 2012) and over 80 articles and book chapters on diachronic Celtic linguistics. He is currently completing a monograph on the syntax of the Continental Celtic languages and is investigating the architecture of the left periphery in the Insular Celtic languages within a

Cartographic framework. He is the editor of the *North American Journal of Celtic Studies* and co-editor of *Indo-European Linguistics*.

Theodorus Fransén was awarded his Ph.D. from Trinity College Dublin in 2020 for a thesis entitled *Past, present and future: Computational approaches to mapping Old and Modern Irish cognate verb forms*. He is currently a postdoctoral researcher on the Cardamom project (Comparative deep models for minority and historical languages), led by John P. McCrae, at the Data Science Institute, National University of Ireland, Galway. While his focus so far has been on computational morphology for Old Irish verbs, he has a growing interest in the broader area of Natural Language Processing and its applications in fields such as lexicography and Digital Humanities. He hopes to explore new, digital avenues to facilitate systematic investigation of linguistic developments between Old and Modern Irish in a future research project. His forthcoming publications include “Automatic morphological parsing of Old Irish verbs using finite-state transducers” (to appear in *Leeds Working Papers in Linguistics and Phonetics*).

Aaron Griffith is Assistant Professor at the Department of Celtic Languages and Culture of Utrecht University. After having finished a Ph.D. at the University of Chicago on various problems in Insular Celtic historical phonology and morphology, he took up a postdoctoral position at the University of Vienna, where he created a digital database and new edition of the Milan Glosses (Griffith and Stifter 2013). The glosses remain a research interest of his, as do the pronominal systems, syntax, and typological profiles of the Insular Celtic languages.

Britta Irslinger is a researcher in the project *Deutsche Wortfeldetymologie in europäischen Kontext* at the Saxon Academy of Sciences and Humanities in Leipzig. She is an Indo-Europeanist and Celticist. Her dissertation on abstract nouns with dental suffixes in Old Irish (*Abstrakta mit Dentalsuffixen im Altirischen*) appeared in 2002. Her major publications include *Nomina im indogermanischen Lexikon* (Wodtko, Irslinger and Schneider 2008) and articles on various topics in historical and comparative linguistics, such as “The gender of abstract noun suffixes in the Brittonic languages” (Irslinger 2014a) and “More tales of two copulas” (Irslinger 2019). Further topics are pre-modern and modern concepts of Proto-Indo-European and Celtic Culture and the history of linguistics, cf. “Medb ‘the intoxicating one?’” (Irslinger 2017a) and “Geographies of identity” (Irslinger 2017b). Her contribution to this book arises from a previous project, “Detransitivity in the Brittonic languages: reflexivity, reciprocity and Middle voice constructions”.

Marius L. Jøhndal obtained his Ph.D. in 2012 from the University of Cambridge and was previously employed at the University of Oslo, where he worked on the PROIEL and Syntacticus treebanks of Indo-European languages. His research has focused on Latin syntax, in particular non-finiteness and reflexivity, and on computational methods for historical linguistics. He currently works for Google.

Elliott Lash is a postdoctoral researcher on the ERC-funded *Chronologicon Hibernicum* project at Maynooth University. He obtained his Ph.D. from the University of Cambridge in 2011 for a thesis entitled *A synchronic and diachronic analysis of Old Irish copular clauses*. Afterwards, he was an O’Donovan Scholar at the Dublin Institute for Advanced Studies from 2011 to 2014, and a ZIF-Marie Curie Fellow at the University of Konstanz from 2014 to 2016. His research

interests are syntax and language change, with a special focus on the history of the Irish language. He is currently writing an introduction to Old Irish syntax. Major journal articles published by him include “Coordinate subjects, expletives, and the EPP in early Irish” (Lash and Griffith 2018), “A quantitative analysis of e/i variation in Old Irish *eter* and *ceta*” (Lash 2017a), “Evaluating directionality in the internal reconstruction of pre-Old Irish copular clauses” (Lash 2017b), and “Subject positions in Early Irish” (Lash 2014b).

Marieke Meelen is a postdoctoral researcher and affiliated lecturer at the University of Cambridge and Fellow-Commoner at Trinity Hall. After completing her Ph.D. at Leiden University on word order change in the history of Welsh in June 2016, she moved to the UK for a postdoc in the ReCoS project led by Prof. Ian Roberts, working on comparative Tibeto-Burman syntax. After the completion of the project, she was awarded a British Academy postdoctoral fellowship to work on her own project “The emergence of V2 word order”, combining information structure and historical syntax with NLP techniques and building the Parsed Historical Corpus of the Welsh Language (PARSHCWL) and the Annotated Corpus of Classical Tibetan (ACTib).

Fangzhe Qiu received his Ph.D. degree in Early and Medieval Irish from University College Cork in 2015. He was an O’Donovan Scholar at the Dublin Institute for Advanced Studies from 2014 to 2015 and then a postdoctoral researcher in the project *Chronologica Hibernicum* in the Department of Early Irish, Maynooth University, Ireland from 2015 to 2019. He is currently a lecturer of Celtic Studies at University College Dublin. He has published widely on early Irish law, Old Irish language and medieval Irish manuscripts. His current research interests include medieval Irish annals, Celtic languages and quantitative historical linguistics. Some of his recent publications are: “Old Irish *ae* ‘descendant’ and its descendants” (Qiu 2019), “The first judgment in Ireland” (Qiu 2018), and “The Ulster Cycle in the law tracts” (Qiu 2017).

Elisa Roma is Associate Professor of Linguistics at the University of Pavia, where she earned her Ph.D. in Linguistics in 1998. The results of her Ph.D. research were published in a monograph (*Da dove viene e dove va la morfologia*; Roma 2000a) and an article (“How subject pronouns spread in Irish”; Roma 2000b). Her main research interests cover typologically oriented historical and comparative linguistics, Celtic philology and in particular Gaelic morphosyntax. She is currently involved in the Italian National Project “Transitivity and argument structure in flux” (Universities of Naples and Pavia). Her strong commitment to multilingualism has led her to translate scholarly works in Irish (*L’irlandese antico e la sua preistoria* and *Il medioirlandese/Middle Irish*; respectively McCone and Roma 2005; Breatnach and Roma 2013). Recent publications include: *Linguistic and Philological Studies in Early Irish* (Roma and Stifter 2014), “Nasalization after inflected nominals in the Old Irish glosses: A reassessment” (Roma 2018a), “Old Irish pronominal objects and their use in verbal pro-forms” (Roma 2018b), “On the origin of the absolute vs. conjunct opposition in Insular Celtic” (Budassi and Roma 2018).

David Stifter is Professor of Old and Middle Irish at Maynooth University. He is founder and editor of the interdisciplinary Celtic Studies journal *Keltische Forschungen* (Vienna 2006–present) and founding member of the Societas Celtologica Europaea (European Association of Celtic Studies scholars). His research interests are language variation and change in Old Irish and comparative Celtic linguistics. Research projects include a dictionary

of the Old Irish glosses in the Milan manuscript *Ambr. C301 infr.*, *Lexicon Leponticum*, and the ERC-funded project *Chronologicon Hibernicum*. His introductory handbook *Sengoídelc: Old Irish for beginners* (Stifter 2006) has been adopted for teaching Old Irish in universities worldwide.

Jürgen Uhlich is a lecturer in Early Irish language and literature at Trinity College Dublin. A monograph based on his Ph.D., entitled *Die Morphologie der komponierten Personenamen des Altirischen*, appeared in 1993 (Uhlich 1993). Jürgen Uhlich has published on Early Irish and Celtic phonology and nominal morphology, the linguistic position of the earliest attested Celtic language Lepontic, Early Irish textual criticism as well as stylistics, most recently on the use of linguistic registers for stylistic purposes in the early Middle Irish text *Fingal Rónáin*. These areas also represent his ongoing research interests. He has furthermore prepared an edition of the Armenian translation of part of the acts of the *Ecumenic Concilium Ferrariense-Florentinum-Romanum* (1438–1445). He is currently working on a *Handbook of early Old Irish*, as well as on various linguistic and textual aspects of that linguistic period of Irish individually.

Christopher Yocum obtained his Ph.D. in Celtic Studies from the University of Edinburgh in 2009 for a thesis entitled *The literary figure of Fíthal*, which focused on the literary aspects of the early Irish judge Fíthal. His current research interests are in the application of semi-structured database and linked data concepts to the early Irish genealogical corpus. He has published articles in *Studia Celtica* and *Éigse*.

Overview of linguistic annotation

Linguistic examples quoted in the chapters are given interlinear glosses and English translations. The glossing conventions followed here are laid out in the following sections.

1 Glossing of Old Irish examples

Nouns are glossed with their translational equivalent and followed by the case (_{NOM, ACC, GEN, DAT}) in subscript small capitals. Singular number is viewed here as default and is not glossed. Plural nouns are glossed with the tag _{PL}, added after the case abbreviation following a full stop (e.g. _{NOM.PL}).

(1) *feruib*
men_{DAT.PL}

(2) *geinti*
gentiles_{NOM.PL}

Adjectives are glossed with their translational equivalent and followed by case, number (_{SG, PL}), and gender (_{MASC, FEM, NEUT}) in subscript small capitals, each tag separated by a full stop.

(3) *móir*
big_{ACC.SG.FEM}

The definite article and other pronominal modifiers (such as quantifiers) are, generally speaking, glossed in the same way as an adjective. However, when the definite article is found immediately before a stressed demonstrative, no gender features are tagged since the demonstrative itself lacks clearly discernible gender features.

(4) a. *in fer*
the_{NOM.SG.MASC} man_{NOM}
b. *in só*
the_{NOM.SG} this_{NOM}

The unstressed demonstrative particles, *-sin* distal (‘that’) and *-so* proximal (‘this’) are glossed respectively as DIST and PROX. These tags are attached to the preceding item with the equals sign. Stressed demonstratives are tagged as nouns, as in (4b) above.

- (5) a. *in* *fer-sin*
 the_{NOM.SG.MASC} man_{NOM}=DIST
 b. *in* *fer-so*
 the_{NOM.SG.MASC} man_{NOM}=PROX

The stressed anaphoric pronoun, *suide* (in all case forms) is glossed with the tag ANAPH followed by case and number tags in subscript capitals with full stops between each tag type. Note that, as with nouns, singular is default and is not tagged. The unstressed anaphoric particle, which has the forms *side*, *sidi*, *ade*, *de*, *adi*, *di*, is only glossed with the tag ANAPH.

- (6) a. *trisodin*
 through=ANAPH_{ACC}
 b. *achotlud* *adi*
 his=sleep_{NOM} ANAPH

Prepositional pronouns are glossed with the translational equivalent of the basic preposition followed by tags for person, number, gender, and case (in that order) in subscript small capitals. Tags for gender and case are separated from the tags for person and number with a full stop. The case tag is only used to disambiguate between the two possible cases (accusative and dative) governed a subset of prepositions which can govern both of these cases. If the preposition only ever governs one case, the case is not indicated in the glossing.

- (7) a. *dóib*
 tO_{3PL}
 b. *foir*
 oN_{3SG.MASC.ACC}
 c. *for*
 oN_{3SG.MASC.DAT}

Verbs are glossed with their translational equivalent and followed by abbreviations in subscript small capitals for agreement, tense, mood, passive and relative (in that order) with a full stop between each abbreviation. The abbreviations

used are listed in (8). Note that indicative mood is here conceived of as the default and is not glossed.

- (8) a. Tense: ^{PRES} (present), ^{IMPF} (imperfect), ^{PST} (past, only in past subjunctive), ^{PRET} (preterite), ^{FUT} (future).
 b. Mood: ^{SUBJ} (subjunctive), ^{CND} (conditional), ^{IMPV} (imperative).
 c. Passive forms are tagged ^{PASS}; relative forms are tagged ^{REL}.
 d. Agreement: ^{1SG}, ^{2SG}, ^{3SG}, ^{1PL}, ^{2PL}, ^{3PL}.
 e. The augment is tagged ^{AUG} or ^{AUG} (see below).

The sequence of glosses in verbs and examples of the method of glossing is given in (9). ^{AUG} has two positions. If it is the first preverb in the verbal complex it is treated as a PV (see below), consider (9a). If it is not the first preverb in verbal complex, it is glossed as in (9c).

- (9) a. *ro·berthae*
^{AUG}·bring^{3SG.PST.SUBJ.PASS}
 b. *berthar*
 bring^{3SG.PRES.SUBJ.PASS.REL}
 c. *inroigrainn*
 PV·persecute^{AUG.3SG.PRET}

For compound verbs, the lexical preverb is glossed separately as PV in capitals. Preverbs are separated from verbal roots by a raised dot in the glossing, even when the dot does not appear in the quoted example. Where present, infixed pronouns (glossed as ^{1SG}, ^{2SG}, ^{3SG_{MASC}}, ^{3SG_{FEM}}, ^{3SG_{NEUT}}, ^{1PL}, ^{2PL}, ^{3PL}) are inserted after the PV (or ^{AUG}) after a hyphen. If relevant the class type is added in parentheses in superscript afterwards (e.g. ^{3SG_{NEUT(A)}}, ^{3SG_{NEUT(B)}}, ^{3SG_{NEUT(C)}}). The hyphen is also used for the infixed relative, which is glossed ^{REL}, in prepositional relatives at after the preverbs *imm* and *ar*.

Consonant mutations play an important role in all Insular Celtic languages. In Old Irish, there are two prominent ones: lenition and nasalization. Lenition causes an initial stop to become a fricative; nasalisation causes initial voiceless stops to become voiced and prefixes a homorganic nasal to initial voiced stops and vowels. The mutations are glossed as superscript ^{LEN} and ^{NAS} respectively before the mutated form. Examples that follow these rules are given in (10).

- (10) a. *as·beir*
 PV·say^{3SG.PRES}

b. *at·beir*

PV-3SG_{NEUT}·say_{3SG.PRES}

c. *as·mbeir*

PV·^{NAS}say_{3SG.PRES}

d. *rondasaibset*

AUG-^{NAS}3SG_{FEM}·pervert_{3PL.PRET}

e. *immetét*

PV-REL·surround_{3SG.PRES}

Old Irish possesses a series of pronominal clitics that serve, roughly speaking, to emphasise items to which they cliticise. In traditional Irish grammar, these are called *notae augentes*. They are glossed with 1SG, 2SG, 3SG_{MASC}, 3SG_{FEM}, 3SG_{NEUT}, 1PL, 2PL, 3PL. These abbreviations are not in super/subscript. They are separated from the glosses for the stressed word with an equal sign (=) as in (11); see also below.

(11) *as·beir=som*

PV·say_{3SG.PRES}=3SG_{MASC/NEUT}

The example itself is presented using the editorial conventions of the edition cited. For example, if the edition does not use a raised dot to separate preverb from root, or a hyphen or equals sign to separate a *nota augens* from the verb, these are not inserted into the main text of the example. Punctuation is only inserted into the gloss as in (12).

(12) *asbeirsom*

PV·say_{3SG.PRES}=3SG_{MASC/NEUT}

In the gloss, an equal sign is used to separate an unstressed element from a stressed element (13), when the two are not separated by a space in the edition cited. A hyphen is used to separate an unstressed element from another unstressed element (14). A period is inserted between the words of translational equivalents where these consist of two or more words (15). An underscore is used between two possibly stressed items that are written without separation in the example (16).

(13) *isuidiu*

in=ANAPH_{DAT}

- (14) a. *arní*
for-NEG
b. *donaibferuib*
for-the_{DAT.PL.MASC}=men_{DAT.PL.}

- (15) *mórabba*
great.cause_{ACC}

- (16) *ísú*
DEICT_this_{DAT}

Note that (16) shows that the deictic particle *í* is glossed as DEICT. The negative particles are glossed NEG (main clause *ní*), NEG_{SUB} (non-main clause *na/nach/nad*) in subordinate non-relative clauses and NEG_{REL} in relative clauses.

2 Glossing of Brittonic examples

The glossing of Brittonic examples is somewhat different from the glossing of Old Irish. These differences are exemplified below.

Nouns and adjectives are glossed with their translational equivalent only.¹

- (17) a. *gwin*
wine
b. *riuedi*
numbers

- (18) *margh uskis*
horse swift

The definite article is glossed as DEF.

- (19) *'r llys*
DEF court

¹ Very occasionally, subscript small capital _{PL} is used to disambiguate a plural form of an adjective from a non-plural form (e.g. Welsh *eraill* is glossed other_{PL}). Certain numerals have feminine and masculine forms. These are distinguished with subscript small capital _{FEM} and _{MASC}, (e.g. *tri* three_{MASC} vs *tair* three_{FEM}).

All pronouns in Brittonic are tagged with the appropriate agreement tag (1SG, 2SG, 3SG, 1PL, 2PL, 3PL) and, if necessary, the following tags in subscript capitals: MASC (masculine), FEM (feminine), POSS (possessive), INFX (infix), INTS (intensifier), REFL (reflexive).

- (20) a. *y penn*
 3SG_{MASC.POSS} head
 b. *a 'e lladwn ef.*
 PTCL 3SG_{MASC.INFX} kill_{1SG.SUBJ.IMPF} 3SG_{MASC}
 c. *dy hun*
 2SG_{INTS}
 d. *dy hun*
 2SG_{REFL}

All demonstratives in Brittonic are tagged as either DIST (distal) or PROX (proximal).

- (21) a. *henna*
 DIST
 b. *an den ma*
 DEF man PROX
 c. *hynny*
 PROX

Verbs are glossed with their translational equivalent and followed by abbreviations in subscript small capitals for agreement, tense, mood, and impersonal (in that order) with a full stop between each abbreviation. The abbreviations used are listed in (22). Note that indicative mood is here conceived of as the default and is not glossed.

- (22) a. Agreement: 1SG, 2SG, 3SG, 1PL, 2PL, 3PL.
 b. Tense: PRES (present), PRET (preterite), FUT (future), IMPF (imperfect), PLPF (pluperfect), HAB (habitual).
 c. Mood: SUBJ (subjunctive), COND (conditional), IMPV (imperative).
 d. IMPS (impersonal)
 e. The perfective particle *re, ry, 'r (etc.)* is tagged PERF.

The sequence of glosses in verbs and examples of the method of glossing is given in (23).

- (23) a. *ledy*
 kill_{2SG.PRES}
 b. *deuthant*
 come_{3PL.PRET}
 c. *lladwn*
 kill_{1SG.IMP.F.SUBJ}
 d. *wnathoed*
 do_{3SG.PL.F}
 e. *bythynt*
 be_{3PL.HAB}

The particle *ym-* (also spelled *em-*) is glossed PV. This is separated from verbal roots by a raised dot in the glossing. Infixd pronouns (glossed as 1SG_{INF}, etc.) are separated from the verb and supporting particles by whitespace. Examples that follow these rules are given in (24).

- (24) a. *ym·dodant*
 PV·melt_{3PL.PRES}
 b. *re gowsys*
 PERF·speak_{3SG.PRET}
 c. *ny 's gwna e hun*
 NEG 3SG_{MASC.INF} make_{3SG.PRES} 3SG_{MASC.INTS}

Other verb-related glosses are: _{VN} (verbal noun), PST-PTCPL (past participle), PTCPL (participle), all subscript small capitals.

Negative particles are glossed NEG, with subscript SUB used for the subordinate negative, where necessary. The predicative particle (*yn* in Welsh) is glossed PRED. The progressive particle (*ow* in Cornish) is glossed PROG. Other particles are glossed PTCL.

3 List of abbreviations

1	1st person
2	2nd person
3	3rd person
A	Class A pronouns
ACC	Accusative
ANAPH	Anaphor
AUG	Augment
B	Class B pronouns

C	Class C pronouns
CND	Conditional
DAT	Dative
DEF	Definite
DEICT	Deictic particle í
DIST	Distal Demonstrative
FEM	Feminine
FUT	Future
GEN	Genitive
HAB	habitual
IMPF	Imperfect
IMPS	Impersonal
IMPV	Imperative
INF	Infinitive
INFX	Infix
INTS	Intensifier
LEN	Lenition
MASC	Masculine
NAS	Nasalization
NEG	Negation
NEUT	Neuter
NOM	Nominative
PASS	Passive
PERF	Perfect
PL	Plural
PLPF	Pluperfect
POSS	Possessive
PRED	Predicative Particle
PRES	Present
PRET	Preterite
PROG	Progressive
PROX	Proximal Demonstrative
PST	Past (Subjunctive)
PST-PTCPL	Past passive participle
PTCPL	Participle
PV	Preverb
REFL	Reflexive
REL	Relative
SG	Singular
SUB	Subordinate (Negative)
SUBJ	Subjunctive
VN	Verbal Noun

Elliott Lash, Fangzhe Qiu, and David Stifter

Introduction: Celtic Studies and Corpus Linguistics

1 Background to the volume

This volume is a collection of eleven chapters that showcase the state of the art in corpus-based linguistic analysis of the old, middle and early modern stages of Celtic languages (specifically, Old and Middle Irish, Middle Welsh, and Cornish). The contributors offer both new analyses of linguistic variation and change as well as descriptions of computational tools necessary to process historical language data in order to create and use electronic corpora. On the whole, the volume represents a platform for the exploration of corpus approaches to morphosyntactic variation and change in the Celtic languages and, for the first time, situates Celtic linguistics in the broader field of computational and corpus linguistics.

These chapters were originally prepared for lectures hosted by the Chronologicon Hibernicum project (ChronHib), an ERC-funded project at Maynooth University, Ireland (ERC Consolidator Grant 2015, H2020 #647351). The lectures occurred at three separate workshops (December 15, 2016, April 4, 2017, October 13–14, 2017), which brought together an international group of researchers with various backgrounds to help the ChronHib team gain insight into preparing linguistically marked-up text for statistical research on language variation in Old Irish. At the first event, all aspects of corpus building and use, such as morphological tagging, syntactic parsing and maintenance and sustainability of online databases, were discussed. In subsequent events, two main themes emerged: first, the necessity of developing computational tools such as morphological taggers/analysers and lemmatisers, and second, that careful use of corpora with a focus on new search queries yields progress on previously intractable problems of Celtic morphosyntax.

2 ChronHib and CorPH

The overall goal for ChronHib is to develop a statistical methodology of linguistic dating in order to more precisely date the diachronic development of the Early Irish language (Old Irish: seventh to ninth century, Middle Irish: tenth to twelfth century) and thereby to predict the age of the large number

of anonymous, dateless Irish texts. In many ways, too, the early stages of Brittonic languages present the same problems of anonymous, as yet undated text (Rodway 2013). In traditional studies of both Goidelic and Brittonic material, linguistic dating has typically been a matter of philological and linguistic analysis of manually curated data. ChronHib aims at advancing the methods used for linguistic dating of Early Irish by contributing to a chronologically more precise description of linguistic variations and by employing corpus linguistic and advanced statistical methods. It also endeavours to improve, by means of digital humanities techniques, on the availability and reliability of the material basis relevant to the chronology of linguistic developments and of the literature of early medieval Ireland (see Qiu et al. 2018 for a more in-depth discussion of ChronHib).

Essentially, ChronHib will produce a new linguistically tagged corpus of Old Irish texts. This corpus, called the *Corpus Palaeohibernicum* (CorPH, Stifter et al. 2015–) is in the development stage and will soon be freely accessible online. It will, firstly, unify some of the existing resources for the study of Old Irish texts under one annotation scheme, and secondly, expand the amount of electronic materials by digitising and annotating data that have only been available previously in printed media or manuscripts. Scholars working on Old Irish, for example, have, until now, mainly relied on the data found in the two-volume printed edition of *Thesaurus Palaeohibernicus* (Thes. = Stokes and Strachan 1901–1910). The existing digital resources for medieval Irish texts come in a variety of forms: annotated lexicons, digital glossaries, text with XML markup, treebanks, and fully digital dictionaries. For extensive discussion of some of these materials, see Griffith, Stifter, and Toner (2018). These heritage data together constitute the corpus on which the contributions in this volume are based, and a brief description of them is pertinent here.

The main online dictionary of Early Irish is eDIL (Toner et al. 2019). It enables research into semantic, morphological, and syntactic usage of Irish lexemes in sources written between the seventh century and 1700. There are, in addition, two major digital collections of early Irish texts: the *Corpus of Electronic Texts (CELT)* hosted by University College Cork (Färber 2012) and the *Thesaurus Linguae Hibernicae (TLH)* hosted by University College Dublin (Kelly and Fogarty 2006–2011). These corpora consist of analytically and structurally XML-marked up texts following the TEI guidelines. The usefulness of these textual resources for the corpus-linguist is only indirect, since no linguistic information is tagged. A prominent treebank is the *Parsed Old and Middle Irish Corpus* (Lash 2014a), a UPenn-style syntactically tagged treebank of fourteen Old Irish texts. The two online annotated lexicons are the *Milan Glosses* database (Griffith and Stifter 2013) and the *Priscian Glosses* database (Bauer 2015; see also Bauer, Hofman, and Moran 2018). These are fully annotated

for morphological and lexical information. Griffith and Stifter's (2013) database consists of around 50,000 morphologically and POS-tagged tokens from the Old Irish glosses in the Milan manuscript *Aubr. C301 infr.* (ML). Bauer's (2015) database consists of around 20,000 morphologically and POS-tagged tokens from the Old Irish glosses in several manuscripts of Priscian's *Institutiones Grammaticae*, with the St Gall *Stiftsbibliothek manuscript 904* (Sg.) containing the most extensive collection of these glosses. These two databases, along with the *Lexicon of the Old Irish glosses in the Würzburg manuscript of the Epistles of St. Paul* (Wb.; Kavanagh 2001, available in print and .pdf formats), have been the catalyst for much research into linguistic variation in Old Irish over the past eighteen years.

The above databases (ML, Sg.) and lexicon (Wb.) were used by most of the contributors in the present volume who studied variation in Old Irish in contemporary (eighth to ninth century) manuscripts. Moreover, many of the texts discussed in Liam Breatnach's and Christopher Yocum's contributions can be found in the CELT and TLH corpora. The ML and Sg. databases have now been incorporated into CorPH and stand beside other resources specifically made for CorPH such as the *Minor Glosses* database (Lash 2018), the *Annals of Ulster* database (Qiu 2019), and the *Poems of Blathmac* database (Barrett 2018a). In total, CorPH has over 120,000 fully annotated tokens of Old Irish text in various genres (glosses, annals, poetry, chief among them) and will allow researchers easy access to a large amount of data for research on linguistic variation. Some chapters in this volume (for example, Elisa Roma's and Theodorus Franssen's) have already made use of data from CorPH.

For the other well-attested medieval Celtic language, Middle Welsh (c. 1150–1500), authoritative editions have long served as the standard corpus for scholars. Meanwhile, two online, searchable corpora have been published, covering the majority of prose texts surviving from before 1425: *Rhyddiaith Gymraeg o Lawysgrifau'r 13eg Ganrif* (Isaac et al. 2013) and *Rhyddiaith Gymraeg 1300–1425* (Luft, Thomas, and Smith 2013). These form the basis of Britta Irslinger's investigation in this volume, and a more detailed description can be found in that contribution. The late medieval and early modern period of the Welsh language is represented by the *Corpus Hanesyddol yr Iaith Gymraeg 1500–1850* (Willis and Mittendorf 2004), which contains about 420,000 words from 30 texts in a variety of genres. However, these corpora have not been linguistically tagged and therefore their usefulness is somewhat limited. The contribution by Marieke Meelen aims to tackle this lacuna by developing tagging methods for part of the prose corpora mentioned above. The last medieval Celtic language dealt with in this volume, Cornish in its middle (c. 1200–1600) and late (c. 1600–1750) phases, survived mainly in versified religious plays and translated works, scholarly

editions of which constitute the corpus for the analysis in Joseph Eska and Benjamin Bruch's contribution.

3 Overview of themes

Digital corpora for medieval Celtic languages have certainly become a central part of the field of Celtic Studies in recent years but fully annotated corpora are still few in number and the application of computational linguistic methods in the analysis of Celtic languages is in its infancy. These languages represent a new frontier in the development of natural language processing tools, in part because they pose special challenges, such as complicated inflectional morphology with non-straightforward mappings between lemmata and attested forms, highly variable orthography, and initial consonant mutations. With so much data available in non-electronic form as the result of previous work and ongoing efforts to convert these data to computer-readable format, it is not surprising to find that the contributors employ both available digital corpora and printed editions or manuscripts in their research, and that quantitative studies are more often conducted in a data-based or data-inspired rather than data-driven manner. This approach shows great potential in revealing hitherto subtle generalisations over various aspects of medieval Celtic languages.

A significant aspect of the volume is that the quantitative studies all deal with aspects of syntactic structure, a subsection of the grammar of medieval Celtic languages (Irish in particular) that has suffered relative neglect, in favour of investigations focusing on phonology and morphology. Happily, more work on syntax has appeared since Isaac (2003) gave a short survey of the few works in the field and pronounced a handbook of Old Irish syntax to be a desideratum. Much of the work of the last decade and a half (e.g. García-Castillero 2013; Griffith 2008; Lash and Griffith 2018; Roma 2014) draws directly on the increasing availability of searchable corpora that enable easy access to the fundamental dataset. This explosion in research is set to continue with the development of CorPH. Bringing the results produced by central scholars participating in this endeavour together in one place emphasises the potential that corpus approaches have in aiding research and underlines many points in need of further investigation.

With its concentration on computational corpus linguistics and morphosyntactic data from historical language stages, this volume is a first in the discipline of Celtic Studies, which has been mainly focussed on traditional philological work such as the editing of texts and literary/historical explication of these

texts. Additionally, it contrasts with and complements other recent volumes of interest to scholars working in Celtic Studies, such as *Formal Approaches to Celtic Linguistics* (Carnie 2011), *Linguistic and Philological Studies in Early Irish* (Roma and Stifter 2014), the proceedings of the fourth *International Congress of Celtic Studies, held in Maynooth University, 1–5 August 2011* (Breatnach et al. 2015), and *Centres and Peripheries in Celtic Linguistics* (Bloch-Trojnar and Ó Fionnáin 2019). While each of these volumes consist of chapters analysing various stages of the Goidelic and Brittonic languages, very few use corpus data or deal with problems of corpus building. Moreover, many of these contain chapters that are more philological, historical, or literature-oriented than strictly linguistic in nature. The present volume, in contrast, reflects the increasing awareness of the usefulness of corpus data in Celtic linguistics, and its contributions show how corpora of Celtic languages can be most effectively constructed and exploited. In the meantime, scholars who focus mainly on philology should still find many of the chapters interesting, as they contribute to our knowledge of the grammars of medieval Celtic languages from fresh perspectives. It is also hoped that chapters such as Marieke Meelen’s and Theodorus Fransen’s, which showcase the development and testing of new computational tools for Celtic language data will also appeal to linguists in general, especially those who are interested in diachronic linguistic changes, computational linguistics, and corpora of historical languages.

4 Description of chapters

The volume is divided into two thematically distinct but related parts. Part one consists of four chapters dealing with the design and creation of corpora for historical languages generally and Celtic languages in particular. Part two consists of seven chapters that are broadly united by the theme of description and qualitative/quantitative analysis of linguistic data derived from the available corpora of medieval Celtic languages. The division into two main parts is motivated by thematic concerns, since the contributions fall into two general groups. There are, firstly, detailed technical discussions of corpus construction, automatic annotation tools, and clustering methods (Marius Jøhndal, Theodorus Fransen, Marieke Meelen, and Christopher Yocum’s chapter), and secondly, primarily corpus-based analyses of particular phenomena (Liam Breatnach, Carlos García-Castillero, Jürgen Uhlich, Elisa Roma, Aaron Griffith, Joseph Eska and Benjamin Bruch, and Britta Irslinger’s chapter). The first part of the book is therefore, roughly speaking, practical with its concentration on computational research tools and methods, while the second is analytical in focus.

Within each part of the book, chapters are themselves grouped thematically. Part one begins with two chapters (by Marius Jøhndal and Marieke Meelen, respectively) that originate from discussions at the first and second ChronHib workshops about the building and sustainability/maintenance of linguistically annotated corpora. Additionally, as a description of a new Welsh treebank, Meelen’s chapter responds to some of the concerns about the need for better ways of doing research on problems of Celtic syntax, as was expressed by participants at the second and third ChronHib workshops. The next two chapters in part one concentrate on the creation and use of computational tools in order to analyse particular aspects of the Old Irish corpora (verbal morphology in Theodorus Fransen’s chapter and stylistic clustering in Christopher Yocum’s chapter).

Part two begins with two chapters (by Liam Breatnach and Carlos García-Castillero, respectively) that investigate the diachronic syntax and morphology of pronouns and demonstratives in Old Irish. The following three chapters (by Elisa Roma, Jürgen Uhlich, and Aaron Griffith, respectively) are all united through their investigation of grammaticalised consonant mutations in Old Irish, whether in the context of relative clauses (Griffith and Uhlich) or after nominals (Roma). The final two chapters in part two (by Joseph Eska and Benjamin Bruch on the one hand and Britta Irslinger on the other) deal with some syntactic phenomena in the Brittonic languages.

4.1 Description of Part 1

Marius Jøhndal’s “Treebanks for historical languages and scalability” presents both a general overview of the motivations for and practice of corpus building as well as a detailed overview of the PROIEL family of treebanks. This group of treebanks includes annotated texts from older Indo-European languages and is one of the most ambitious recent corpus-related projects for these languages. It includes the original core, the PROIEL (Pragmatic Resources in Old Indo-European Languages) itself, which is a corpus of New Testament texts in Ancient Greek, Latin, Classical Armenian, and Gothic, as well as some other texts in some of these languages. Additionally, the PROIEL family also includes the ISWOC Treebank, consisting of texts in Old English and Old Romance (Spanish, Portuguese), and the TOROT database with texts in Old Slavic (Old Church Slavonic, Old Russian). One of the goals of the chapter is the introduction of a new interface for browsing and searching the PROIEL Treebank and related treebanks called *Syntacticus* (<http://syntacticus.org>). This expansion of the PROIEL family of treebanks increases its visibility and is a crucial way of

achieving long-term maintenance. It is also an exemplary open-source infrastructure that can be used for future projects. The chapter is therefore programmatic and practical, since the kinds of technical, linguistic, and manpower related challenges it describes serve as both a guideline to best practice and an inspiration for future research on Celtic languages. Although the chapter does not discuss Celtic languages in particular, in many respects it sets the tone for the volume since many of the issues mentioned in it, being characteristic of less-resourced historical languages, will be familiar to scholars of medieval Celtic languages and it is hoped that the chapter may serve as a call to collaboration.

“Annotating Middle Welsh: POS tagging and chunk-parsing a partial corpus of native prose” by Marieke Meelen demonstrates the process of creating an annotated corpus of some Middle Welsh native prose (as against translated works), and the challenges and potentials of building such a corpus. The corpus contains only literary narratives and some law texts at present but will be extended to other genres and registers. Digitalised texts were pre-processed with punctuation and tokenisation, which was done automatically by a POS tagger and a Memory-Based Tagger. The text was then marked up with a simplified version of the TEI P5 header. The author adopts the UPenn annotation scheme modified with Welsh-specific tags that enable further queries concerning agreement patterns and change in Information Structure. A Memory-Based Tagger assigns morpho-syntactic tags to tokens automatically and a modified rule-based chunk-parser is deployed to annotate syntax and information structure. This chapter presents the first systematic approach to annotating historical Welsh, and the corpus it describes ultimately aims to provide a starting point to build a fully annotated Welsh historical treebank.

In “Automatic morphological analysis and interlinking of historical Irish cognate verb forms”, Theodorus Fransen describes a computational approach to understanding how the Irish verbal system develops diachronically. The author’s major contribution is to propose a morphological analyser for Old Irish verbs and to discuss ways this analyser can be incorporated into a framework of computational resources for various stages of Irish. This proposal dovetails with Jøhndal’s and Meelen’s chapters in dealing with ways of expanding the current computational toolset for a historical language (specifically historical stages of Irish) and in its concerns with scalability. These concerns are reflected in his detailed investigation of the challenges encountered by a methodology that incorporates finite-state morphology as it applies to Old Irish. The challenges he details are two-fold. The first challenge has to do with word and morpheme division as encountered in “real” text, i.e. editions or manuscript transcriptions. In many cases, multiple morphemes may be written as a concatenated string, resulting in the need to find a way to encode licit combinatorial possibilities of multiple

morphemes. This is a so-called *generation* problem, where generation means the ability of the analyser to generate all and only the licit inflected forms of any given stem. In other cases, whitespace is found between morphemes leading to potential parsing ambiguities since the analyser is word-based (where a word is understood to be an element between whitespace). This is a so-called *analysis* problem, which may result in the wrong morphological tag being assigned to any given string. The second challenge has to do with the complex interaction between phonology (especially stress) and morphology in Old Irish since stress alternations can result in syncope and the presence or absence of palatalisation of stem-final and ending-initial consonants. These challenges impinge on the choices made for implementing the finite-state transducer. For instance, does one rely on a strictly rule-based approach to specify certain licit combinations and handle stem variants induced by stress alternations, using “flag” morphemes or upper-level filters for instance to deal with the generation problem? Or does one hard-code (i.e. list) such stem variation or parts of paradigms? Fransen carefully weighs the advantages of different approaches in order to ensure the applicability of his analyser. He also envisions a fully functioning POS-tagger suitable for both Old and Middle Irish by making some suggestions for allowing interoperability of resources, especially between his morphological analyser and Dereza’s (2018) Old Irish lemmatiser.

Christopher Yocum’s chapter “Text clustering and methods in the *Book of Leinster*” uses machine-learning techniques to cluster the texts in the *Book of Leinster* (LL), and tries to identify the reason for the clustering. The author extracts individual texts from the electronic edition of LL, tags the function words and calculates the frequency of function words in each text. The frequencies are then turned into a matrix of vectors, which goes through the *k*-medoids algorithm, subject to normalisation and “Principal Component Analysis”. The result is a clustering scatter plot. The clustering can be caused by the variables of author, scribe or genre, and these three factors are tested in turn. The result suggests that authorship is the main factor in clustering, and that the traditional ascriptions to certain authors do not fit the clustering and may need to be revised. The methods used are innovative within Celtic Studies and contrast with the traditional philological approach to text clustering. The chapter is a useful addition to the large body of work on the history of the manuscript and the clusters of text reported on deserve further investigation. If specific linguistic usages can be associated with particular clusters, this may be useful for the study of idiolect/style at particular periods.

4.2 Description of Part 2

In “The demonstrative pronouns in Old and Middle Irish”, Liam Breatnach uses a corpus of Old Irish verse texts that are largely available online in TLH and CELT. The author first observes that there is a split between the unstressed enclitic demonstrative particles *-sin* ‘that’, *-so/se* ‘this’ and their stressed pronominal variants, *sin* ‘that’, *só/sé* ‘this’ (dative *sund/síu*). The rest of the chapter deals with a diachronic investigation of the morphophonology, syntax, and semantics of the stressed demonstrative pronouns. The results of this investigation map the distribution of demonstratives according to four main features: syntactic function, singular/plural number, inanimate/animate reference and period (i.e. Old versus Middle Irish). The main contribution of the chapter is that it highlights subtle differences between Old and Middle Irish usages. First, while the stressed demonstratives on their own (without the addition of the particle *í*) could be construed as plural in both Old and Middle Irish, plural reference was very restricted in Old Irish, but much expanded in Middle Irish. Specifically, plural reference is found in Old Irish when the demonstrative acts as a subject of a copular sentence and in later Old Irish as the complement of an agreeing preposition. Middle Irish allows plural reference in some other contexts. Second, demonstratives with inanimate and animate reference are likewise found in both Old and Middle Irish, but animate reference in Old Irish once again is restricted to subjects of copular sentences whereas it is found in other contexts in Middle Irish. The chapter closes with some discussion of the possibility that the independent, personal pronoun *sé* ‘he’ developed during Middle Irish from the demonstrative *sé* in contexts where it had animate reference.

Carlos García-Castillero’s chapter is titled “Paradigmatic split and merger: The descriptive and diachronic problem of Old Irish class B infixed pronouns”. This contribution replaces García-Castillero’s lecture “Synonymy (a^N / *aní* ‘that (what)’, a^N / *inta(i)n* ‘when’) and homonymy (a^N ‘that (what)’ and a^N ‘when’) in the Old Irish glosses” presented at the third workshop, because the author had already submitted the lecture for publication elsewhere. The contribution in this volume explains the diachronic origin of the Old Irish class B infixed pronouns, which are used in a declarative clause after pretonic lexical preverbs of the structure (-)VC-. The author firstly clarifies the relevant notions in Old Irish (clause types, verbal complex, phonotactic structure of preverbs, etc.), and then illustrates the use of non-third person infixed pronouns with instances collected from the corpus of the contemporaneous Old Irish glosses. This corpus-based approach yields the interesting observation that, in the language of the contemporaneous Old Irish texts, non-third person infixed pronouns are much less regular than the third person infixed pronouns in making a distinction between declarative and relative forms,

especially when the lexical preverb after which the infixed pronoun appears is of type (-)VC-. Such asymmetry in distribution between the persons raises a question, which, in the author's opinion, is directly related to the diachronic origin of the class B infixed pronouns. The author argues that class B infixed pronouns arose to distinguish a verbal complex with a third person singular masculine or neuter infixed pronoun in a declarative clause from a complex without an infixed pronoun in a relative clause. More specifically, a process of morphological split in the original class C paradigm has given rise to two forms in the third persons, and tentatively in the other persons.

Elisa Roma presents her findings on the distribution of nasalisation after nominals in Old Irish glosses in "Nasalisation after inflected nominals in the Old Irish glosses: Evidence for variation and change", where her main interest lies in the possibility of mapping variation in nasalisation to chronological or diatopic criteria. All instances of nasalisation after nominals from four Old Irish corpora of glosses have been collected (Wb., Ml., and Sg. and the *Minor Glosses Database*). The phonetic contexts for nasalisation are categorised, as well as the word class of the nasalising/nasalised word. The frequency of nasalisation in each combination of phonetic context and word class has already been reported in Roma (2018a). Firstly, the data show that the absence of nasalisation after inflected nominals in Old Irish cannot be due merely to the loss of a nasal consonant in consonant clusters. Secondly, individual texts show different frequencies of nasalisation in the same context. The variation between Old Irish texts in nasalisation after inflected nominals suggests not only diachronic strata but also probable regional differences that led to later developments in Modern Irish and Scottish Gaelic. The chapter is comparable to other corpus-based investigations of morphophonology, such as Griffith (2016a) and Lash (2017a). Together with these papers, Roma's chapter is illustrative of the impact lexicons and corpora have had on Celtic linguistics.

In "On the obligatory use of a nasalising relative clause after an adjectival antecedent in the Old Irish glosses", Jürgen Uhlich uses a corpus consisting of the main Old Irish glosses (Wb., Ml., Sg.) to explore the extent to which adjectives having a modal adverbial reading must be followed by a nasalising relative clause in cleft sentences (e.g. *arndip maith nairlethar a muntir* 'so that he may well order his household', lit. 'that it may be good how he orders'). The author argues that, save for some well-defined exceptions, the nasalising relative clause is an absolute prerequisite of this construction. His approach is at once quantitative, since he has systematically and exhaustively collected all instances of modal adjective cleft sentences from the glosses and studied their distribution, and qualitative, since he also carefully establishes and describes the varying types of "exceptions" to the generalisation. The exceptions to the

generalisation include (a) cases in which the verb in the clause following the adjective has an object marked with a class A or B infixed pronoun, (b) instances of mixed antecedents in coordination where the antecedent farthest from the embedded clause is the modal adjective, (c) clauses involving what Uhlich terms “syntactic raising”, essentially multiple dependencies, where the modal adjective and another constituent simultaneously act as the antecedent to the embedded clause, and (d) some possibly innovative instances of leniting rather than nasalising relative clauses. The paper is an important contribution to a long-standing debate in Old Irish studies dealing with the rather complex syntax of relative clauses and its conclusion that a nasalising relative clause is an essential component in a modal adjective cleft revises the previous consensus that nasalising relative clauses were optional across much of the domain in which they could be used.

In Aaron Griffith’s chapter, “The ‘Cowgill particle’, preverbal *ceta* ‘first’, and prepositional cleft sentences in the Old Irish glosses”, he connects what he calls “three seemingly unrelated” phenomena: the phonological shape of the adverbial preverb *ceta* ‘first’, evidence for the so-called Cowgill Particle (**eti*), and the usage of relative verbs in PP-clefts. The author investigates both the first and second vowel in *ceta* using a combination of a quantitative corpus-based approach and a qualitative comparative approach. In his discussion of the variation in the initial syllable of *ceta* (attested as both *ceta* and *cita*), he shows that the usage of the *i*-variant increases over time. He then argues that the final vowel of *ceta*, together with the final vowel of the preverb *ocu* (in *ocu-ben*) could provide further, previously unexamined, evidence for the Cowgill Particle, if the initial vowel of **eti* was not elided after preverbs ending in *u* (i.e. **kintu-eti*, not **kintu-ti* > *ceta*, **onku-eti*, not **onku-ti* > **ocu*). Because the preverb *ceta* is predominately found in relative clauses, where the Cowgill Particle would in fact not be expected, the paper then shifts to a discussion of two examples in which a verb containing *ceta* is arguably non-relative. These two examples are both prepositional cleft sentences (e.g. *ar is do thabirt díglae berid in claideb sin* ‘for it is to wreaking revenge that he carries that sword’), where a non-relative verb typically follows the prepositional phrase (PP). The author surveys the evidence for PP-clefts in the corpus of glosses and shows that, despite the general rule, the Milan Glosses have innovative relative verbs after the PP. While this leaves the status of the two examples containing *ceta* uncertain (they could either be non-relative, and therefore evidence for the Cowgill Particle, or relative), the chapter is, like Uhlich’s, a useful contribution to the perennial debate on the syntax of cleft sentences and relative clauses in Old Irish.

Britta Irslinger, in “The functions and semantics of Middle Welsh X *hun(an)*: a quantitative study”, uses two untagged corpora of Middle Welsh – *Rhyddiaith*

Gymraeg 1300–1425 / Welsh Prose 1300–1425 and Rhyddiaith y 13eg Ganrif: Fersiwn 2.0 – to investigate an innovative usage of the collocation *X hun(an)* (where *X* is a possessive pronoun) as a reflexive pronoun in Middle Welsh. The author shows that the collocation *X hun(an)* was generally used as an intensifier in the corpora, in a manner similar to English *myself* in *I saw him myself*, but there is some evidence of its grammaticalisation as a reflexive pronoun. This new function of *X hun(an)* appears in fourteen instances out of a total of 1908 unique tokens of *X hun(an)*, where it is used instead of the usual reflexive markers, the verbal prefix *ym-* or plain pronouns. The fourteen examples of reflexive usage come from translation literature, but it does not appear that the collocation *X hun(an)* corresponds to any particular intensifier marker in the base language. This suggests that the examples display a real innovation in Welsh grammar. The study is part of an ongoing effort (see references cited in the chapter) to understand the expression of reflexivity, reciprocal action, and middle voice in Welsh and also contributes to the debate over the extent to which English *-self* as an expression of reflexivity arose as the result of contact with Welsh. According to the author, the use of *-self* as a reflexive in English expanded from the mid-twelfth to the seventeenth century. Although this is not explicitly stated by the author, the fact that there are so few examples of *X hun(an)* used as a reflexive before 1425, i.e. after the first signs of the innovation in English, could suggest that the contact with Welsh was not the only factor in the development of *-self*.

In “Prolegomena to the diachrony of Cornish syntax”, Joseph Eska and Benjamin Bruch discuss the diachronic development of the configuration of the Cornish affirmative root clause with comparison to other Brittonic languages. Since verbal sequences do not occur in Old Cornish, examples from Old Welsh and Old Southwest Brittonic, showing VSO and V2 orders, are quoted, with the assumption that these languages behaved similarly to Cornish. The affirmative root clauses in Middle Welsh and Middle Breton are generally V2, and surface V2 (along with V3) is also found in Middle Cornish. The authors then analyse the architecture of the left periphery and the preverbal Object DP, pointing out that the exceptions to V2 in Middle Cornish are caused by metrical considerations overriding the grammar, and despite the corpus of Middle Cornish being composed largely of verse, the Middle Cornish affirmative root clause was V2 of the “relaxed” type. The authors then examine the corpus of Late Cornish texts and find that these are of dubious evidential value because the corpus is very small and consists of translations by a native speaker and texts by non-native speakers.

Part 1: Corpus tools for historical Celtic linguistics

Marius L. Jøhndal

1 Treebanks for historical languages and scalability

1 Introduction

Historical linguistics, whether synchronic or diachronic, is by definition based on corpora. Since we do not have access to the intuitions of native speakers we can only test linguistic hypotheses about historical languages by systematically collating information from our corpus of texts.

For questions that typically concern linguists, this often means identifying every occurrence of a particular phenomenon in the corpus, analysing, classifying and counting the occurrences and then using this for testing hypotheses about the structure of the language. This can be done manually, but this is time-consuming and error-prone. As Haug (2015) points out, while reading the text and manually collating information from it is essential for *hypothesis formation* it is much less useful for *hypothesis testing*. Even if the text is in electronic form, it is easy to overlook an example, record it incorrectly or fail to apply test criteria consistently over time.

This paper focuses on treebanks, which are corpora that have been annotated with morphosyntactic information so that we can extract linguistic structures like ‘verb with an accusative noun’. High-quality treebanks for a range of historical languages now exist and are widely used in historical linguistic research. This includes treebanks that follow the Penn-style of annotation, e.g. the Penn-Helsinki Parsed Corpus of Middle English (Kroch and Taylor 2000), the Penn-Helsinki Parsed Corpus of Early Modern English (Kroch, Santorini, and Delfs 2004), the Penn-Helsinki Parsed Corpus of Modern British English (Kroch, Santorini, and Diertani 2016), the Tycho Brahe Parsed Corpus of Historical Portuguese (Galves and Britto 2002) and the Icelandic Parsed Historical Corpus (Wallenberg et al. 2011), as well as dependency-based treebanks, e.g. the Index Thomisticus (Passarotti 2007), the Ancient Greek and Latin Dependency Treebanks (Bamman and Crane 2011; Celano, Crane and Almas 2014), the PROIEL Treebank (Haug and Jøhndal 2008, Haug, Eckhoff et al. 2009), the ISWOC Treebank (Bech and Eide 2014) and the TOROT Treebank (Eckhoff and Berdičevskis 2015).

A key challenge in building treebanks for historical languages is lack of resources. Funding is limited and there are few existing computational language resources like taggers and parsers available. At the same time, the task is complex and experts on the language have to devote a significant amount of time to

the annotation task. This comes on top of the complexity of designing a suitable annotation scheme that balances the desire to capture philological and linguistic detail with an approach that is reliable, scalable and technically feasible.

A key motivator behind treebank efforts is to facilitate reuse of resources and to provide access to large data sets that make hypothesis testing robust and encourage replication of published research, but as funding for construction of a treebank tends to be tied to a time-limited research project, it is challenging to fulfil such long-term aspirations and achieve scale and long-term consistency.

This paper describes these challenges in the context of the PROIEL, ISWOC and TOROT treebanks, and how this has motivated efforts to use automated tools like taggers and parsers to scale the annotation process. The paper also describes *Syntacticus* (<http://syntacticus.org>), which now serves as a shared front-end for PROIEL, ISWOC and TOROT, but whose long-term aim is to integrate automated taggers and parsers with our existing annotation tools and offer this as an open infrastructure platform that can be used by researchers working on other less-resourced, historical languages within the Indo-European family, such as the Celtic languages.

Section 2 briefly introduces the PROIEL, ISWOC and TOROT treebanks and some key properties of the annotation scheme. Section 3 describes the challenges involved in maintaining these treebanks, expanding them and making them accessible for researchers, and how this has motivated us to set up *Syntacticus*. Section 4 describes in more detail current efforts aimed at evaluating how the annotation process can be scaled using automated taggers and parsers.

2 The PROIEL, ISWOC and TOROT treebanks

The PROIEL-family of treebanks currently includes the PROIEL, ISWOC and TOROT treebanks. Together they contain text samples from a number of old Indo-European languages (see Table 1) which, when consolidated into one treebank, contains around one million words that have been lemmatized, morphologically analysed and annotated with syntactic dependencies.

The original PROIEL Treebank stems from a research project called *Pragmatic Resources in Old Indo-European Languages* at the University of Oslo (2008–2012), which was set up to study information packaging in ancient Indo-European languages. A major part of this was to compile a treebank containing the New Testament in its original and translations, as the New Testament is a natural

Table 1: Languages and token counts in the PROIEL Treebank release 20180408, the TOROT Treebank release 20180919 and the ISWOC Treebank release 20160620.

Language	Number of tokens	Number of sentences	Treebank
Ancient Greek	250,455	18,173	PROIEL
Latin	225,064	19,425	PROIEL
Gothic	57,211	5,457	PROIEL
Classical Armenian	23,513	1,916	PROIEL
Old Russian	235,275	24,716	TOROT
Old Church Slavonic	58,269	6,350	PROIEL
Old Church Slavonic	82,007	8,371	TOROT
Old English	29,406	2,536	ISWOC
Old French	2,340	137	ISWOC
Old Portuguese	36,595	2,027	ISWOC
Old Spanish	54,661	2,615	ISWOC

parallel text that allows for cross-linguistic comparison of phenomena like word order, anaphoric expressions, definiteness, background events and discourse particles.

To achieve this the New Testament texts were annotated with morphosyntactic and information-structure annotation, and then aligned so that words in, for example, the *Vulgate* were linked to the words that they translate to in the Greek *New Testament*.

The PROIEL Treebank has since been expanded with other texts in Latin and Ancient Greek, which have been morphosyntactically annotated. Since the end of the original PROIEL project, the long-term objective has been to expand the treebank to the point where it contains – to the extent it is practically possible – representative samples from different periods and genres. This is why, for example, the Latin section of the treebank now includes not just the *Vulgate* and texts from the classical canon, like Caesar’s *Gallic War*, but also works like the Late Latin *Peregrinatio Aetherae* and sections of Palladius’ agricultural handbook, and at the time of writing Petronius’ *Satyricon* and samples from Plautus are being prepared.

In parallel to the continued expansion of the PROIEL Treebank, the ISWOC Treebank and the TOROT Treebank were set up. The ISWOC Treebank contributes

samples from Old English, Old French, Old Spanish and Old Portuguese, while the TOROT Treebank contributes a large and expanding selection of texts from Old Russian and Old Church Slavonic. Both are modelled on the PROIEL Treebank and were designed to be fully compatible. They therefore adhere to the same annotation scheme, were built using the same annotation process and rely on the same data representation (Eckhoff et al. 2018).

Using the same annotation scheme offers a range of advantages. For linguists using the treebank the main advantage is that it becomes possible to test cross-linguistic hypotheses, but it also significantly simplifies the process of building a treebank if resources can be combined to design shared guidelines and build shared annotation infrastructures that reflect best practices.

The Universal Dependencies project (Nivre et al. 2016) is today the largest collection of treebanks that have been harmonised in this manner, and Universal Dependencies have become the de facto standard within computational linguistics. The PROIEL Treebank predates Universal Dependencies and uses a different annotation scheme, but the PROIEL-style of annotation can be automatically converted to Universal Dependencies. The conversion relies on some heuristics but work is ongoing to align the PROIEL-style of annotation with Universal Dependencies so that these heuristics can be eliminated.

2.1 The annotation scheme and the annotation process

The PROIEL-style of annotation is based on multiple levels of annotation. Lemma, part of speech and morphological features are annotated at the morphological annotation level. The syntactic annotation level includes labelled dependencies, as well as a combination of enhanced (or ‘secondary’) labelled dependencies and empty elements for representing syntactic phenomena that involve gaps, coindexing or displacement. The information structure level has annotation for givenness and anaphoric reference chains. The alignment level contains links between elements that are translational equivalents in two texts. Finally, the semantic level is used for free classification of data according to criteria like aspect or lexical semantics.

Each annotation level allows for annotation of individual tokens. Some levels are also defined for larger textual units like sentences or paragraphs, but the annotation process itself is designed around sentences as the minimal unit. As annotation of a text progresses, each sentence is individually assigned an ‘annotated’ or ‘reviewed’ status, where ‘annotated’ indicates that the sentence has been annotated by the primary annotator and ‘reviewed’ indicates that it has also been approved by the secondary annotator. A sentence has to have

complete annotation on both the morphological and syntactic levels before it can be assigned the ‘annotated’ or ‘reviewed’ status, while the other levels of annotation are optional and can be added independently.

The annotation scheme used on the morphosyntactic level is broadly aligned with ‘school grammar’ in the sense that assumptions about morphology and syntax are not too different from what would be expected by students who have studied the language but not necessarily formal linguistics. The scheme by default also tries to adhere to linguistically informed conventions for the language and its philological traditions. For Latin, for example, lemmatising is based on the *Oxford Latin Dictionary* but has been adapted to make the relationship between headwords and parts of speech more predictable so that each lemma in the treebank has one and only one normalized headword form and one and only one part of speech.

Although no linguistic annotation is ever completely theory-independent, morphological annotation is generally uncontroversial as philologists and linguists of different persuasions generally follow the same conventions. Syntactic annotation is a different matter with wide-ranging disagreement among researchers. The syntactic annotation in PROIEL-style treebanks is based on dependency grammar. Dependency grammar is not well developed as a linguistic theory, but the PROIEL-flavour of dependency grammar has been enriched with formal devices that can handle syntactic structures like raising and control. The implementation of these devices and the specific analyses of structures with ‘gaps’ or long-distance dependencies is based on Lexical-Functional Grammar (Kaplan and Bresnan 1982, Bresnan 2001), whose functional structures were in turn influenced by dependency grammar. Grammatical functions like subject and object are primitives in Lexical-Functional Grammar and this assumption has also been carried over into PROIEL-style dependency grammar along with Lexical-Functional Grammar’s criteria for identifying these grammatical functions.

Dependency grammar-based annotation was chosen over an annotation scheme rooted in constituency structure in part because of its near-universal adoption in current computational work, and in part because it makes it possible to annotate free-word-order languages consistently. Haug (2015) discusses the latter point in more detail, as well as broader methodological motivations and the practical implications of this choice.

The details of the syntactic annotation scheme and the precise handling of specific syntactic structures are complex and well beyond the scope of this chapter, which aims to give only a brief overview of the key characteristics of the treebanks. For further details on the morphosyntactic annotation scheme the reader is directed to the overviews by Haug and Jøhndal (2008), Haug,

Jøhndal et al. (2009), Haug, Eckhoff et al. (2009) and Eckhoff et al. (2018), while the design of the annotation scheme for information structure is described in Haug et al. (2014).

3 Long-term scalability and maintenance challenges

A number of early design choices contributed to the success of treebanks that use the PROIEL-style of annotation. Annotation requires specialist knowledge, so it is crucial to be able to recruit students and researchers across the world as annotators. This requires a tool that supports distributed annotation and that does not have to be installed on the annotator's computer, as this would have required us to provide technical support to annotators. We also needed a tool that could be tailored to the evolving annotation scheme and allow us to make continuous improvements to the software without disrupting annotators. No such tool existed in 2008 when work on the PROIEL Treebank started. We therefore opted to develop our own annotation tool as a web application.

The use of dependency grammar and the organisation into multiple levels of annotation, in which each level is independent and can be conceptualised either as a graph with nodes and edges or as pairs of tokens and feature structures, allowed for a flexible data model that could be mapped onto standard technologies for data representation and storage like XML and relational databases, and it permitted researchers to work independently, adding other annotation levels when resources and expertise became available.

Treating the sentence as the smallest unit that can be annotated and reviewed on its own is also a design decision that has worked well in practice as it made it possible to release data in batches, even when texts were not completely ready, and to preserve the history of changes in a practical way.

Finally, the Lexical-Functional Grammar-influenced variety of dependency grammar has proven to be easy for annotators with philological training to learn and apply consistently. It also allows for some flexibility in designing consistent analyses of syntactic structures across languages when there is disagreement in the linguistic literature on what the correct analysis is.

Other design choices have in hindsight proven to be suboptimal or have blocked progress. The model of having a primary annotator with a secondary annotator as a reviewer was put in place to ensure consistency while the annotation scheme was still being developed, and was subsequently used to ensure that the three treebanks were compatible and used formal devices in the same

way. This relied on extensive coordination between reviewers and centralised training of annotators. This approach worked well when several annotators were working intensively on annotating multiple texts in parallel but is not cost-effective today when only a few annotators occasionally work on expanding the treebank.

The process for developing documentation was not integrated with the annotation tool itself. Unfortunately, documentation efforts have therefore not kept up with annotation and the documentation is neither consolidated nor complete.

On the technological side, the annotation tool is monolithic, so it is hard to break it up or replace components. This makes it challenging to modify it or the data model that it uses. This is a particular issue in two areas. First, it has hampered integration with external automated taggers and parsers, which is necessary since the tool itself only has built-in support for generating suggestions using finite-state transducers or by looking up the annotation that an annotator has already chosen for a token with the same surface form. Second, it has slowed down efforts to address weaknesses in the data model, which is a particular concern as the data model lacks support for sub-token annotation, e.g. annotation of compound words or infixes.

In combination these challenges now constitute a significant barrier to further expansion of the treebanks and are risk factors when it comes to long-term maintenance and accessibility.

3.1 Syntacticus

To address the long-term scalability, maintenance and accessibility challenges, we launched Syntacticus in 2018. The aims of Syntacticus are (1) to increase the visibility, accessibility and discoverability of the PROIEL, ISWOC and TOROT treebanks, (2) to develop processes for long-term maintenance, (3) to improve the scalability of the annotation process and (4) to provide an open infrastructure platform for other researchers working on less-resourced, historical languages. These are ambitious aims that will take time to achieve. Aim 4, in particular, is a long-term aspiration. Aims 1 and 2, on the other hand, are crucial for ensuring that the treebanks remain accessible and reliable. Aim 3, in turn, is a requirement if continued expansion is going to be economically feasible.

Visibility, accessibility and discoverability (aim 1) have been addressed by setting up a dedicated website for Syntacticus (<http://syntacticus.org>) that provides much more direct access to data from the treebanks than before. Crucial elements include removing all registration barriers, incorporating elements of

the familiar search-engine paradigm in the user interface and making more of the treebank data indexable by search engines. We have also included direct access to data that have been synthesised from treebank data like dictionary resources that are automatically generated from the morphosyntactic annotation. The *Varangian Rus Project* (Eckhoff and Berdičevskis 2016) has in turn built an Old Russian dictionary with glosses in Russian and English on top of the synthesised dictionary for Old Russian.

At the time of writing much work remains to be done before the Syntacticus site is mature and satisfies our requirements, but the process for achieving this is well understood and achievable given recent advances in web technology and the broad availability of suitable open-source software components. The remainder of this paper is devoted to discussing how we aim to address annotation scalability (aim 3), which presents significant challenges for low-resourced languages.

4 Scaling morphosyntactic annotation

Manual annotation of lemma, part of speech and morphological features is time-consuming, error-prone and very tedious for annotators. The practical experience from PROIEL, ISWOC and TOROT has shown that annotation speed increases and the error rate decreases when annotators are provided with some automated assistance, such as pre-populated annotation fields that they can correct or a list of suggested annotations that they can choose from. The effect is positive even when this assistance is very crude and generated using simple methods, such as looking up annotations that have already been made earlier in the text, ranking them by frequency and serving them to annotators as suggestions.

More sophisticated and higher-accuracy assistance can be provided if we use automated taggers, parsers and other techniques in natural-language processing (NLP). The difficulty here is that historical languages are, in NLP jargon, low-resource languages. This means that the data sets and models that are prerequisites for applying many NLP techniques do not usually exist and have to be built largely from scratch. For example, in order to use a statistical part-of-speech tagger you would have to train the tagger using a corpus that has already been annotated with parts of speech.

While some required language resources, like part-of-speech-tagged corpora, do exist for the most widely studied historical languages, they may not be suitable for the task. It is common for such resources to be too small, or to suffer

from inconsistent quality or licensing incompatibility. Even when high-quality, freely reusable resources do exist, different design decisions or idiosyncratic technological choices can make reuse a complex and time-consuming task.

This chicken-and-egg problem, posed by having to annotate a substantial amount of data before automated methods can be used to assist with the task, is a particular challenge if, as is often the case, the annotated corpus itself is one deliverable within a larger research project whose primary aim is actually to answer linguistic and philological questions.

4.1 Rule-based tagging and ambiguity

Lack of suitable resources has been the situation for most of the languages in Syntacticus. The approach we have taken when starting the annotation process for a language is to rely on a combination of (1) a crude mechanism for looking up existing morphosyntactic annotation weighted by frequency, (2) rule-based morphological analysers that provide guesses for inflectional forms that have not been annotated before, and (3) hand-crafted rules for deriving probable syntactic labels from the morphological analysis.

Our rule-based morphological analysers are written using finite-state morphology (Beesley and Karttunen 2003), which is a well-understood technique for mapping surface inflectional forms to morphological analyses. Writing a complete finite-state morphology for a language is a large undertaking, but while a finite-state morphology with high coverage may be desirable for other purposes, we have found that in practice we only need a finite-state morphology with limited coverage and mainly benefit from it in the initial phase of annotating texts in a new language.

The finite-state morphology can build on a combination of a manually compiled list of high-frequency function words with analyses, and rules for high-frequency inflectional classes. If an electronic lexicon is available, it may be possible to combine this with the rules for inflectional classes. If no such lexicon is available or it lacks details of inflectional classes, we can instead use a stem guesser that allows us to guess unknown words based on what a likely stem is. A stem guesser will, however, over-generate unless accurate rules for possible stems and possible combinations of stem and affix can be formulated. If the stem guesser over-generates, annotators will be faced with a range of nonsense annotation suggestions. Unfortunately, the information in reference grammars is in practice not detailed enough to formulate precise constraints, even for a well-documented language like Latin, so a phase of experimentation is necessary to achieve the right balance.

4.2 Statistical approaches and data sparsity

For every surface form there may be a number of possible analyses. A finite-state morphology, or any other method that simply looks at a single inflectional form in isolation and maps it to possible analyses, cannot on its own disambiguate them. For Syntacticus this means that when the finite-state morphology is applied annotators are served a list of tuples of lemma, part of speech and the ten morphological features of the PROIEL annotation scheme. Depending on the properties of the language, the list of suggestions can be very long and it can require time and concentration for the annotator to pick the right combination, especially since many candidate analyses differ only in one morphological feature.

Because of the potential for multiple possible analyses to degrade annotation speed and accuracy, it is paramount to perform disambiguation. As the annotated part of the corpus grows, this becomes possible using statistical NLP techniques.

The canonical method for statistical part-of-speech or morphological tagging uses supervised machine learning. In supervised machine learning the system is given a training set which consists of an input with features and their correct labels. In this case the features are the surface word forms to be tagged, or parts of those word forms, and the labels are the parts of speech, lemmas or morphological features. Then, using a machine-learning algorithm, the system produces a classifier that can assign labels to new inputs. In other words, the system is given the correct answers for part of the data and then uses this to infer a model that can generalize to unseen data.

For historical reasons, our existing toolchain only allows off-line tagging. This means that statistical tagging is done as a separate batch operation in which the morphological annotation level is populated with the output of the tagger. This takes place before annotators start their work and may have to be repeated at regular intervals as annotators correct suggestions from the tagger. Our experiments with this ‘pre-annotation’, which have mostly used the TnT tagger (Brants 2000), show a significant positive effect on annotator performance (Skjærholt 2011). The process is, however, very inflexible and does not allow us to make full use of modern taggers. Instead, automated tagging should be done on-the-fly as annotators work on the text. We also need to understand better what affects accuracy so that we can tune parameters and build a pipeline that does text normalisation when this has a positive effect on accuracy.

The key challenge when applying statistical techniques to a historical language is data sparsity. Morphological complexity is one contributing factor as it makes each individual inflectional form less frequent than in less morphologically

complex languages, and for morphologically more complex languages the models that perform well are not necessarily the same as those that perform well for morphologically less complex languages. Despite this, reasonable results can be achieved. As an illustration, Celano, Crane and Majidi (2016) report 88% average accuracy for Ancient Greek part-of-speech tagging.

Another reason for data sparsity is lack of standardised orthography. The level of standardisation differs significantly between historical languages and is a complex issue that spans the degree of variation in manuscripts, the philological traditions and conventions of published editions for a particular historical language. For texts, whose orthography shows significant variation, normalisation of the text before training a model and before tagging may significantly improve results. Berdičevskis, Eckhoff and Gavrilova (2016) report for a Slavic corpus that tagging accuracy improved significantly with text normalisation (89.5% accuracy for POS-tagging and 81.5% for a ten-feature morphology).

As a rule of thumb, a small training set will lead to low accuracy, and historical corpora in general tend to be small. While there are large corpora for historical languages like the Bibliotheca Teubneriana Latina, which contains around 13 million words of Latin, we have to keep in mind that such collections cover a multitude of genres, a range of registers and sociolinguistic variation and, importantly, texts produced at very different times. The effect of such intra-corpus variation on accuracy is not clear. Birnbaum and Eckhoff (2018) find that for tagging Byzantine Greek results improve when the tagger is trained on a corpus that contains a combination of Ancient Greek, Koine and Byzantine Greek (91.3% accuracy for POS-tagging and 94.0% accuracy for ten-feature morphological tagging), despite the internal variation within this corpus. On the other hand, Adesam and Bouma (2016), in work on Old Swedish, show that when a tagger is trained on one section of a text and then used to tag another section of the same text, accuracy is very high (94.2% average accuracy for POS-tagging and 83.2% for POS and morphology), but when it is used to tag a *different* text with similar properties the results are vastly inferior even when various forms of text normalisation are applied (69.9% for POS-tagging and 49.0% for combined POS and morphology).

Existing work on historical languages has mainly focused on part-of-speech tagging, morphology and lemmatising, but we know from work on other languages that the challenges for dependency parsing are similar. The conventional approach is to perform dependency parsing as a separate step after tagging. As an illustration of the range of accuracy that can be achieved in this way, results from a large-scale experiment that included the Universal Dependencies version of the PROIEL Treebank were in the range of 80% unlabelled accuracy and 75% labelled accuracy for Latin and Ancient Greek (Alberti et al. 2017). It is interesting

to note, however, that current state-of-the-art parsing approaches that have been designed to work with raw text as input, and which have been trained using data from many languages, perform worse on historical languages than on living languages primarily because it is difficult for them to determine sentence boundaries from the inconsistent punctuation and spelling conventions in historical data (Zeman et al. 2017).

5 Conclusion and future work

In this chapter we have described some scalability challenges that the PROIEL, ISWOC and TOROT treebanks face, discussed the motivation for setting up Syntacticus and reviewed some early results from relevant studies on automated tagging and parsing of the historical languages that Syntacticus covers.

Our work on scaling the annotation process with taggers and parsers is in its early phase, and while the studies on automated techniques reviewed here show some promising results, it is not certain that these techniques will lead to actual improvements in annotation speed and accuracy. Before we can determine this we have to integrate these tools with our existing annotation tool-chain and conduct experiments with our online annotation tool in realistic online annotation scenarios.

Work on Syntacticus as a platform is also still in its early phase. In particular, we have to scale the annotation process to new languages. This may require applying more specialised techniques, such as annotation projection, an approach in which existing annotation in one language is mapped onto another language. Sukhareva et al. (2017) have recently demonstrated that this can be used successfully to induce a part-of-speech tagger for Hittite using Hittite texts that had been aligned with German texts.

Another unsolved issue for new languages concerns the interaction with tokenisation. For annotating Sanskrit, for example, it is necessary to do tokenisation and tagging at the same time since Sanskrit texts show the surface forms that result after the application of sandhi, which often removes the word boundaries that taggers and parsers tend to assume are present. Moreover, work by Inglese, Molina and Eckhoff (2018) show that the data model itself may need to be revised substantially to support sub-token annotation for languages like Hittite, where the relationship between text and annotatable unit is particularly complex.

Marieke Meelen

2 Annotating Middle Welsh: POS tagging and chunk-parsing a corpus of native prose

1 Introduction

For a study on syntactic changes in Middle Welsh (see Meelen 2016), it was desirable to compile a searchable corpus of Middle Welsh, at least a partial one including the most important narrative literature from the medieval period. This chapter presents this first annotated corpus of historical Welsh. For the present study, a selection of Middle Welsh texts was used, based on their popularity among Welsh philological and linguistic researchers (the tales of the *Mabinogion*) and a less-researched excerpt of a law text (the *Laws of Women*) to compare the results in a different genre. This selection forms a first step towards the creation of a much larger, and well-balanced, Parsed Historical Corpus of the Welsh Language (PARSHCWL, see Meelen and Willis forthc.).

The *White Book of Rhydderch* and the *Red Book of Hergest* manuscripts, both dating from the 14th century, contain the most famous collection of Middle Welsh native¹ literature: the *Mabinogion*. These tales (of unknown authorship) derive in part from an oral literary tradition. They are thought to have their origin in the early medieval period but were only put down in writing several centuries later (see, among others, Davies 1998). For the present chapter, all extant tales of the *Mabinogion* (11 in total) were annotated, representing the narrative prose of the Middle Welsh period of the language, c. 1150–1500 AD. High-definition photographs of both of these manuscripts are available online via the websites of the National Library of Wales (www.llgc.org.uk – *White Book manuscript* Peniarth 4–5) and Jesus College, University of Oxford (<https://digital.bodleian.ox.ac.uk/> – *Red Book manuscript* Jesus College 111). The earliest text for the present corpus is an excerpt from a law text mainly concerned with the rights of women. This excerpt of the Early Welsh laws is taken from the BL manuscript Add. 22356 (S), one of the most important manuscripts in the tradition of the Welsh *Law of Hywel*. This *Law of Hywel* was the system of law in use in medieval Wales, based on an

¹ The term ‘native’ refers to the texts that are assumed to be originally composed in Middle Welsh, as opposed to ‘translated’ texts that are translations from other languages, such as chronicles translated from Latin.

older custom system. It is named after Hywel Dda ('Hywel the Good'), a 10th-century Welsh king. The British Library Law manuscript Add. 22356 (S), however, is from the mid-15th century. As with all Middle Welsh law texts, although the manuscripts are mostly dated from the late medieval period, (parts of) the texts go back many centuries (Davies 1966). The latest edition is accessible online via www.cyfraith-hywel.org.uk.

Although many of these manuscripts are available as digital photographs, not all of them have been converted to (online) editions with search options that facilitate philological and linguistic research. A large collection of Middle Welsh texts, including the texts of the *Mabinogion* included in the present study, is now available online through the Welsh Prose project (www.rhyddiaithganoloesol.caerdydd.ac.uk), but the search function is limited, and the texts are not annotated in any way. Searchable corpora are indispensable tools in historical linguistic research, both from a quantitative and qualitative perspective. Qualitative corpus research, for example, comprises investigations of the distribution of the different forms and constructions that are attested in a wide variety of texts. When analysing large amounts of texts or corpora in this way, linguists need to be extremely consistent in their approach to get meaningful and testable results. Categorising and labelling forms or structures in large amounts of data can be prone to error, because even the most careful researcher can change their ideas about the features and characteristics they use to disambiguate categories as they are confronted with more and more material. As the number of texts in need of investigation grows, it is no longer feasible to simply read and make notes. A further disadvantage of manual notes is that the results are much harder to verify and replicate, which is problematic in quantitative studies in particular. But qualitative studies can also benefit from searchable (annotated) corpora, as with improved computational methods specific or rare forms under investigation (for the purposes of philological research or comparative reconstruction, for example) are much easier to find. Therefore, it is useful to employ methods from the field of Natural Language Processing (NLP) and the tools created by computational linguists to build annotated searchable corpora. Because of their computational nature, these tools are designed to consistently deal with large amounts of data in a very short period of time. The results are consistent, following strict rules that are well-described in the annotation manuals (see Meelen and Willis forthc.), and can then be made readily available for any (Welsh) linguist.

Having said this, however, as an inflected language without a standardised orthography, Middle Welsh poses some specific challenges for ready-built NLP tools like part-of-speech (POS) tagging and parsing algorithms that automatically add morphosyntactic information (see Sections 3 and 4). Initial consonant mutation, for example, can yield a word like *pawb* 'everyone' in three different

ways: *pawb* ‘everyone’ (no mutation), *a phawb* ‘with everyone’ (aspirate mutation) and *i bawb* ‘to everyone’ (soft mutation). Furthermore, with five tenses (Present, Past, Pluperfect, Present Subjunctive and Conditional/Imperfect), seven different person-number-gender suffixes, various sets of pronouns and clitics and a wide range of functional particles, Middle Welsh with its rich morphology and extensive orthographical variation is far more difficult to automatically annotate than a morphologically poor and orthographically standardised language like Present-Day English.

This chapter presents the first systematic approach to annotating historical Welsh, ultimately aiming to provide an excellent starting point to build a fully annotated Welsh historical corpus (see also Meelen et al. 2017 and Meelen and Willis forthcoming). This chapter discusses some of the challenges faced developing the methodology and annotating the first part of this corpus. Some of these challenges are specific to Welsh (or other Celtic languages), others inherent to working with historical corpora in general. Section 2 gives a brief overview of how the corpus was compiled and pre-processed. In Section 3, I discuss the procedure of part-of-speech (POS) tagging and of developing a tag set to annotate Welsh (and potentially other languages) with rich verbal and prepositional inflection. Sections 4 and 5 shed light on adding syntactic and information-structural features to the data so that it can be queried in various ways. In the final section, I demonstrate how this first annotated corpus of Middle Welsh prose can be extended and that this entails promising opportunities for future research.

2 Pre-processing

2.1 General philosophy and goals

Ideally, any corpus should be well-balanced in terms of text genre, length, origin etc. When working with historical data, however, the choices are often limited, and creating detailed annotation is extremely time-consuming. Therefore, this first annotated historical corpus of Middle Welsh focussed on the most commonly used editions² of the 11 native tales of the *Mabinogion* and some excerpts from the laws only. Future extension of the corpus will include alternative manuscript versions

² The editions used for this first annotated corpus are Williams (1951) for the *Four Branches*, Bromwich and Evans (1992) for *Culhwch and Olwen*, Thomson (1997) for *Gereint*, Thomson (1968) for *Owein*, Goetinck (1976) for *Peredur*, Roberts (1975) for *Lludd and Llefelys*, Williams (1908) for *Breuddwyd Maxen*, Richards (1948) for *Breudwyt Ronabwy*. For the new and complete

of each of these texts. In addition to that, the corpus will be extended to include more texts from different genres and registers such as the historical chronicles of the kings and princes and translated texts such as chronicles or Bible translations.

The main aim of this project was not to give a complete syntactic analysis or to provide a detailed parsed structure. The part-of-speech (POS) tags contain highly detailed morphological information (see Section 3), but the phrasal annotation is confined to a hierarchical chunk parse (see Section 4). In this way, the annotated corpus could remain theory-neutral. At the same time, queries for linear order and basic hierarchical phrase structure (e.g. to find noun phrases within prepositional phrases) are still possible. Finally, future enrichment of the chunk-parsed corpus is enabled, because of its flexible XML format (see Section 5).

Any controversial decisions were avoided as much as possible by backing off to a simpler form of annotation. For example, interjections were simply labelled as 'INTJ' instead of aiming to classify them further. Similarly, constructions that are changing over time were consistently annotated to facilitate future studies. A good example is the *sef*-construction³ in Welsh. The information-structural status of this construction changes from initial identificatory focus to plain predicate focus in the course of the Middle Welsh period (see Meelen 2016: 272–284). Since most texts are difficult to date exactly, throughout the corpus the specific tag SEF for any occurrence of this type of sentence was used (see also Section 3 on POS tagging). In this way, all sentences with the *sef*-construction can be easily found and examined in context.

2.2 Preparation and text formatting

There are various orthographical inconsistencies in the various versions of the texts of the *Mabinogion* (cf. Thomson 1986: xi), e.g. *yniuroed~y niuroed* 'numbers', *mywn~mewn* 'in', etc. For the present study, the texts were not extensively pre-processed through text normalisation or orthographical regularisation, because there was no stemmer available yet for Middle or Early Modern Welsh. A stemmer can automatically detect the root or stem of inflected and conjugated

Parsed Historical Corpus of the Welsh Language (PARSHCWL), we will create new editions so that all annotated texts can be deposited and made freely available to any scholars.

³ This construction originated from a reduced identificatory cleft construction *ys ef* 'it is it', but developed in the course of the Old and Middle Welsh period into an adverb *sef* meaning 'namely'.

forms, which is extremely useful when automatically processing a new text in a particular language. With a stemmer the task of POS tagging is easier because the various forms are reduced to stems + their respective inflection/conjugation. Developing a good stemmer or carrying out the procedure manually during the pre-processing stage is a tremendous task, however, which is why the detection of inflected forms was left to the POS-tagging stage, where the automatic classification algorithm could use specific features to disambiguate words regardless of their varying orthography (see Section 3).

In order to prepare the texts for annotation, a minimum amount of preparation is always necessary, however. First of all, the markup was stripped from the digitised texts, which were then saved as plain text files (.txt) so that they would be in the right input format for the POS tagger. Further pre-processing involved the insertion of sentence-final punctuation (if that was not present in the edition already) and the deletion of sentence-internal full stops. Finally, utterance boundaries in the form of <utt> were inserted semi-automatically (automatically after a full stop and manually if the full stop did not exist). Without utterance boundaries, the POS tagger is not able to assign morphosyntactic tags to all tokens. In addition, utterance boundaries are useful units for subsequent NLP tasks like chunking, chinking⁴ and full syntactic parses. The only punctuation marks that were removed were the full stops preceding and following numbers, e.g. ‘.12.’ was turned into ‘12’ to optimise automatic token recognition. Tokenisation (the isolation of words) was done automatically by the POS tagger on the basis of word spacing and full stops at the end of an utterance.

The Natural Language Toolkit (NLTK, www.nltk.org) regular expression chunk-parser (see Section 4) requires a list of words and tags. Therefore, the POS-tagged text files of the format token/TAG were converted to the right input format for the chunk-parser using a simple text conversion script designed especially for this purpose (see Meelen 2016: 326 for a sample of the code and further details of the chunking process).

2.3 Splitting and joining tokens

Meelen and Beekhuizen (2013) in their pilot study observed that the huge amount of orthographic variation in Middle Welsh complicates the POS-tagging task tremendously. The advantage of using a Memory-Based Tagger (MBT) is that this

⁴ Chunking is defining a set of words that should be grouped into phrases; chinking is defining the set of words that should be excluded from those phrases.

type of tagger could filter those out on the basis of the context most of the time (see Section 3). Splitting and joining certain tokens still had to be done, however. When two words are merged together without spaces in the orthography, e.g. *ymangor* (*yn* + *Bangor*) ‘in Bangor’, the resulting combination contains a preposition ‘in’ and place name ‘Bangor’, but only receives one POS tag. If words like these are not split, it is difficult to decide which morphosyntactic tag this should be (preposition or place name). Joint tags like ‘P-NPR’ (‘preposition + proper noun’) could in theory be created, but the larger the amount of POS tags, the more difficult it will be for the algorithm to classify words correctly. Also, a researcher interested in place names would need the opportunity to automatically extract just place names from the corpus and would not like to be confronted with extra work sifting through examples of place names combined with prepositions (in case of combined tags) or missing examples altogether (if only the preposition tag was used). Some tokens, however, were particularly challenging for the automated tagger, since very few generalisations could be made from the small training set (see Meelen and Beekhuizen 2013). To overcome some of those very specific orthographic challenges, the following combinations were automatically split using so-called regular expression replacements. Regular expressions are sequences of characters that define search patterns. They allow for more detailed searches (and therefore quicker semi-automatic replacements) than simple string searches, because they can include logical operators. In this way, fixed combinations with prepositions that cause nasal mutation like *yn* ‘in’, e.g. *ymwyt* > *y#⁵+mwyt* ‘in food’, conjunctions combined with definite articles: *ar* > *a# + r* ‘and the’ and particles combined with pronouns, e.g. *ae* > *a# + e* ‘PCL PRO-A’ (particle + accusative pronoun).

3 POS tagging

A properly pre-processed version can be tagged automatically by a part-of-speech tagger. Although for this first attempt to create an annotated corpus pre-processing tasks were minimal, the tokenisation (splitting and combining words) alongside the insertion of sentence boundaries was enough to make a memory-based algorithm perform well. For Middle Welsh, there was no Part-of-Speech tagger available yet. I therefore generated an MBT that could subsequently be used to assign morphosyntactic tags to the Middle Welsh data. For this purpose,

⁵ The # sign was inserted to indicate word breaks so that researchers are still able to identify the original orthography.

decisions have to be made concerning the tag set, the list of morphosyntactic labels for each of the Middle Welsh words. A very detailed tag set facilitates more (and different types of) research. When working with a corpus of limited size, however, use of too many different tags leads to low frequencies and many hapaxes, which in turn complicates the automatic tagging task and yields degraded results. In this section I describe these challenges and furthermore offer some solutions that are not only useful for those working on Middle Welsh, but for anyone working with similarly complex historical data.

3.1 Establishing the morphosyntactic tag set

The rich morphology, the initial consonant mutations and the abundant orthographical variations of Middle Welsh pose significant problems for any automated task in Natural Language Processing. Extensive pre-processing including normalisation of spelling and mutations, for example, would simplify the POS tagging since the tagger would then be able to recognise a larger number of words. Apart from the fact that such regularisations are time-consuming, it entails a type of editorial intervention that has major implications for future research. When orthography is regularised, for example, editorial decisions have to be made concerning what this ‘regular’ spelling should be in the first place. It also does not allow for dialectal variation or differences in scribal practices that might give crucial insights in the linguistic history and geography at a given time. Another way to simplify the tagging task is to limit the tag set to a short list of broad morphosyntactic categories like ‘VERB’, ‘NOUN’, etc. However, this too limits the range of research opportunities tremendously. It is therefore worthwhile to develop an annotation scheme that gives as much morphosyntactic information as possible.

A commonly used annotation scheme adding morphosyntactic information to historical corpora is the ‘UPenn standard’ developed initially for Old and Middle English texts (see www.ling.upenn.edu/histcorpora). This annotation scheme, however, does not always provide enough information to answer certain research questions, mainly queries concerning agreement patterns and changes in information structure. To enable further research in these and other areas, I have used the standardised UPenn scheme, but extended the part-of-speech tag set where necessary. Starting from the already extended tag set used for the Icelandic corpus (cf. Wallenberg et al. 2011), I have examined the features of Middle Welsh grammar and systematically added extra features, i.e. more inflectional features such as person and number added to the root and tense/mood forms with a dash/hyphen.

3.1.1 Verbal tags

Verbal inflection in Welsh mainly occurs as a suffix to the verbal stem. Inflected verbs in the UPenn tag set are tagged VB. Past tense is indicated by the regular English past-tense ending in *-ed*, resulting in VBD. For Welsh, the VBD for the preterite tense was kept and in the same way tags for present (-P), future (-F) (only relevant for irregular verbs) and pluperfect (-G, for Welsh *gorberffaith* ‘pluperfect’), imperative (-I) and imperfect (-A, for Welsh *amherffaith* ‘imperfect’) etc. were added. Finally, a distinction was made between indicative (-I) or subjunctive (-S) mood for the present and imperfect tenses.⁶ This results in insightful systematic combinations like VBPI (present indicative), VBAI (imperfect indicative), VBG (pluperfect) etc. The same letters were systematically added to irregular verbs, resulting in for example DOPI (present indicative of the verb *gwneuthur* ‘to do’), GTI⁷ (imperative of the verb *cael* ‘to get’) or BEAS (imperfect subjunctive of the verb *bod* ‘to be’).

Apart from these more-detailed tense-aspect-mood markers, further information was added about the inflection to indicate person and number. Following standard glossing practices, person and number were represented as -1SG (first-person singular), -2PL (second-person plural) etc. Welsh has a further inflectional suffix for the ‘impersonal’ form of the verb that can be used in true impersonal contexts meaning ‘one’ or underspecified ‘they’; it is frequently translated into English as a passive. The number 4 was employed for this specific suffix and added to the verbal tags like the other personal endings, e.g. VBPI-4 (impersonal present indicative) or DOAI-4 (impersonal imperfect indicative of the verb *gwneuthur* ‘to do’).

3.1.2 Inflected and combined prepositions

Another feature of the grammar, specific to Welsh and other Celtic languages (but also seen in for example Semitic languages like Arabic or Hebrew), is inflected prepositions. Middle Welsh had a specific set of prepositions that could

⁶ Note that the ‘-I’ for “indicative” can only appear combined with ‘-P-’ (present) or ‘-A-’ (imperfect); the ‘-I’ for ‘imperative’ only appears directly after the core verbal tag (VB/GT/BE/DO). Since the imperative does not have different tenses this does not lead to any ambiguities in the annotation system.

⁷ The initial ‘HV-’ is for the auxiliary verb *have* in the UPenn standard annotation scheme. In Welsh, *cael* can mean ‘have, get’ with auxiliary functions as well, although it is not the exact equivalent of English *have*, which is why GT ‘get’ was chosen for Welsh *cael*.

be inflected for person, number and gender (the last of these in the third-person singular only and only ever with a pronominal object). There are also ‘uninflected’ prepositions in Welsh, but the inflected set includes very common prepositions like *i* ‘to’, *ar* ‘on’ and *yn* ‘in’. Following the verbal inflectional tags, Middle Welsh *iddi* ‘to her’ for example was tagged as P-3SGF ‘preposition third-person singular feminine’.

Some prepositions in Welsh could also be combined with other prepositions, e.g. *y dan* ‘under, below’ from *y* ‘to’ + *tan* ‘under’. These complex prepositions were tagged P + PX, so they could be recognised as separate, but also as combined prepositions. A further advantage of this is that the automatic tagger looking at the tags preceding and following the focus word, will not encounter the rare sequence of two prepositions. A disadvantage remains, of course, that the tag set is further extended and there are more homophonous forms that could render worse results if the complex preposition in question does not occur frequently in the training set.

Welsh also allows for some further types of complex prepositions: a combination of a preposition plus a grammaticalised noun. If the object of this type of preposition is a pronoun, it can appear in between the two prepositions as a possessive pronoun, e.g. *yn eu herbyn* ‘against/towards them’ from *yn* ‘in’ plus *eu* ‘their’ plus *erbyn* ‘opposition’. There are two possible ways to annotate constructions that are changing in historical corpora: we can annotate the original structure and form or the new construction as a whole. Since the exact date of grammaticalisation is often difficult to determine, it is not always easy to choose one or the other. As long as the construction is tagged consistently in one text (or one period of the historical corpus) and the annotation manual is clear about this, this should not be a problem. In that case, future researchers will always be able to find and, if necessary, change the annotation again. In this particular case of combined prepositions, a less conservative annotation scheme, disregarding the nominal origin of the construction yielding the tag sequence ‘P PRO-G PX’ (preposition – possessive pronoun – second part of combined preposition) was preferred to facilitate research into prepositional phrases.

3.1.3 Distinguishing different types of pronominal forms

Another part of grammar in which the tag set was extended significantly is pronominal forms. Since Welsh has various sets of pronouns for different (grammatical) contexts, a more fine-grained distinction here could enhance research not only in the pronominal domain, but also in Information Structure. Conjunctive pronouns like *ynteu* ‘he (then)’, for example, are used in contexts of

topic switch, meaning ‘but I’, ‘I, then,’ etc. Reduplicated pronouns like *tydi* ‘you’, on the other hand, are only used in focussed contexts. Separate tags for those are thus useful for finding the focus domain of sentences.

A further distinction in the pronominal domain was made between possessive pronouns and object pronouns. Since the infixed versions of these pronouns often exhibit the exact same form, a more fine-grained distinction in the tag set facilitates syntactic research here as well. Following the extensions of the tag set for the Icelandic parsed corpus (see Wallenberg et al. 2011), these pronominal tags receive case tags marked with a dash, for example, *fy* ‘my’ PRO-G (‘pronoun genitive’), or *e* ‘him’ PRO-A (‘pronoun accusative’).

3.1.4 Additional extensions of the tag set

Further extensions of the tag set include ADJQ for equative constructions, e.g. *cochet* ‘as red’ (from *coch* ‘red’ + equative *-et*) and ADJPL for plural adjectives, e.g. *gueisson ieueinc* ‘young servants’. More detailed tags like these are helpful to historical linguists and syntacticians looking at the structure and agreement patterns of noun phrases.

As described above, Welsh employs a wide range of particles. These too were tagged separately according to their function (e.g. PCL-QU ‘question particle’, PCL-FOC ‘focus particle’, PCL-NEG ‘negative particle’) to help distinguish different types of clauses. Aspectual particles like *yn* ‘progressive’ (PROGR) or *wedi* ‘perfective’ (PERF) were also distinguished from the homophonous predicative particles (PRED) and prepositions (P) respectively.

The verbal noun category characteristic of Celtic languages was tagged VN for regular verbs. Irregular verbs with verbal nouns that have specific functions in Welsh, e.g. *cael* ‘get’, also used for the passive, received specific verbal noun tags. The -N was added systematically to their base forms, e.g. GT- ‘have, get’ > GTN ‘verbal noun of the verb *cael* ‘to get’’. The verbal noun of the verb ‘to be’ was tagged as ‘BEN’, although it can also appear in this form in many other syntactic contexts, e.g. in complement clauses.

Finally, some additional lexical items with specific functions were tagged separately. An example of this is the above-mentioned petrified form *sef* (tagged SEF), which was used in earlier stages of the language to focus identificatory copular sentences. During the Middle Welsh period, it grammaticalised further until it became an adverbial element used in apposition to noun phrases meaning ‘that is’ (cf. Latin *id est* still used as the abbreviation *i.e.* in English).

One final problem that remained for Middle Welsh was the large amount of homophony alluded to in various cases above already. Because there is a large number of very short words in Middle Welsh that are spelled the same with a wide range of meanings, e.g. *a* which can mean ‘and’, ‘with’, preverbal particle, etc. this poses problems for automatic classification algorithms. The tagger, however, was often able to distinguish between up to five different possible meanings of, for example, Middle Welsh *y* ‘the, his, her, to, to his/her, in’ etc. on the basis of the preceding and following context.

3.2 Morphosyntactic annotation

With the MBT from TiMBL⁸ it is possible to generate a tagger for any language based on a training set consisting of a systematic token/TAG format with utterance boundaries at the end of every sentential unit. A memory-based tagger uses memory-based algorithms to disambiguate and classify words in a corpus. Unlike other types of taggers, memory-based taggers interweave processing and learning stages. Whenever a new language item + classifier is encountered in the training data, it leaves a memory trace that guides subsequent processing. This means that when a new instance is found and needs to be classified, a set of relevant instances is selected from memory with a number of useful features, and the new token is classified by analogy to that set. This therefore yields robust results especially when POS tagging languages with orthographical variation and rich morphology (see Zavrel and Daelemans 1997, and Daelemans and Van den Bosch 2005 for instructions, background and further functionalities).⁹

3.2.1 Automatic POS tagging

Once the Middle Welsh tagger is generated, the settings file of the tagger is used to assign POS tags to a new part of the corpus (presented as a tokenised text file). Based on the training set, the MBT divides the new text in need of annotation into ‘known’ and ‘unknown’ words. Depending on the exact

⁸ <https://languagemachines.github.io/mbt/>

⁹ For the new version of PARSHCWL, the results of this Memory-Based Tagger will be compared to a state-of-the-art BiLSTM-CNN-CRF tagger (see <https://github.com/achernodub/targer>) to see which of those yields better results and should form the basis of texts that need to be annotated and added to PARSHCWL in the future (see Meelen and Willis *forthc.*).

parameter settings and the features from the training set stored in memory, the tagger will then assign a tag to each word.

In Welsh the inflection appears as a suffix (on verbs or prepositions). When the tagger finds an unknown word like *ohonaf* ‘of me’, for example, it can compare the last three characters to known words with assigned tags in the training set. An example of this could be another inflected preposition, like *arnaf* ‘on me’ with the POS tag P-1SG (‘preposition + first person singular ending’). The exact same final characters (in combination with the other tags in the preceding and following context) lead the MBT to assign the same tag P-1SG to *ohonaf*, which would be the correct tag. Note that the context of inflected verbs and prepositions is often quite distinct (e.g. prepositions can appear after inflected verbs whereas inflected verbs do not). Since the automatic tagger is also sensitive to context, it will be able to distinguish verbs from prepositions and therefore verbs ending in *-naf*, e.g. *canaf* ‘I sing’, would be correctly tagged as first-person singular present verbs (VBPI-1SG) instead of inflected prepositions (P-1SG).¹⁰

Known words are easier if there are no homophones with different tags. If there are homophones, for example the above-mentioned Middle Welsh word *y*, the context in which they appear is crucial. In between an adverb (ADV) and an inflected verb (VB*/GT*/BE*/DO*), *y* is undoubtedly the preverbal particle following sentence-initial adjuncts, like in (1a). In front of verbal nouns, however, like at the end of (1b), *y* could be the preposition ‘to’ or a possessive pronoun (masculine, feminine or third-person plural), as earlier in (1b).

- (1) a. *Tranhoeth y deuthant y ’r llys.*
 next.day PTC come_{3PL.PRET} to DEF court
 ‘The next day they came to the court.’ (Bromwich and Evans 1992: line 595, [*Culhwch ac Olwen*])
- b. *a dyuot yn y uryt ac yn*
 and come_{VN} in 3SG_{MASC.POSS} mind and in
y uedwl uynet y hela
 3SG_{MASC.POSS} mind go_{VN} to hunt_{VN}
 ‘and he was minded to go and hunt’ (Williams 1951: lines 3–4 [*Pwyll Pendeuic Dyuet*])

¹⁰ Many thanks to an anonymous reviewer who suggested explaining this potentially ambiguous case.

The output file of the tagging process is a text file consisting of a word + TAG (shown in Figure 1) and an indication whether this word was known (signalled by a single forward slash “/”) or unknown (signalled by a double forward slash “//”) from the training set.

```
Kilyd//NPR mab/N Kyledon//NPR Wledic//NPR a/PCL uynnei/VBAI-
3SG wreic/N kynmwyd/ADJQ ac/P eġ/PRO ./PUNC <utt>
Sef/SEF gwreic/N a/PCL uynnwys/VBD-3SG ,/PUNC Goleudyt//NPR
merch/N Anlawd//NPR Wledic//NPR ./PUNC <utt>
```

Figure 1: Fragment of the output of the POS tagger for the text *Culhwch and Olwen*.

MBT allows for different settings according to the features of the words themselves or the context in which they appear. In order to obtain the maximally reliable tags, a wide range of parameter settings was tried, varying those features. The optimal settings for Middle Welsh known (-p) and unknown (-P) words are the following (see Meelen 2016: 331–332 and Daelemans et al. 2010 for further details on these specific settings):

- p dfa
- P sssdFawchn
- M 200 -n 5 -% 5 -O +vS -F Columns-G K: -a 0 U: -a 0 -m M -k 17 -d IL

For Middle Welsh, the corrected gold standard of one text was subsequently used to annotate other texts of the *Mabinogion* and the laws automatically with greater accuracy. Each of those texts was in turn manually corrected as well.

3.2.2 POS tagging results

In order to estimate the quality of the POS tagger and obtain optimal parameter settings, I evaluated the manually annotated data with a ten-fold cross-validation, i.e. taking 90% of the data, training the model on that subset and then testing it on the other 10%, repeating this procedure for ten 90%~10% splits. Since the ten percent that the model is tested on is manually checked, we can see how often the model assigns the correct tag to a word, as well as obtain insightful statistics about the over- and under-generalisations of some tags. The above-mentioned settings gave the following results for the 59,000-word Middle Welsh corpus (see Meelen 2016: 40–44 for a full overview and further discussion of the results):

- Global Accuracy: 90.4%
- Global Accuracy known words: 93.3%
- Global Accuracy unknown words: 63.3%

To give a better insight in the success of the tagger, I calculated the Precision (percentage of system-provided tags that were correct), Recall (percentage of tags in the input that were correctly identified by the system) and F-score (weighted harmonic mean of Recall and Precision). We find high results for simple tags like CONJ ‘conjunction’ or N ‘noun’ that occur extremely often. As expected, Precision and Recall for tags occurring only once or twice is extremely low. These tags are often forms of verbs that occur very infrequently with irregular endings. Precision and Recall give more insight in the degree to which the model over- or under-generalises certain tags for the individual categories. The genitive (possessive) pronoun category (PRO-G), for instance, is correct about 90% of the cases where it is applied (90% Precision), but out of all actual possessive pronouns, only 65% is recognised (Recall of 65%). This is understandable, because the possessive pronoun usually consists of only one letter that is homophonous with the object infixed pronoun. The model thus under-generalised that category in particular. If 95% of the actual conjunctions on the other hand are recognised as such, while the item is only classified as a conjunction correctly in 90% of the cases, this category would be slightly over-generalised. As expected, the F-score for frequently occurring tags is considerably higher than that for tags and tokens occurring only once or twice in the corpus. The extremely fine-grained tag set with over 200 morphosyntactic labels (see the Appendix of Meelen 2016 for a full overview) can thus only reach a good Global Accuracy in a large corpus. This first corpus is not very large, which is why a Global Accuracy of over 90% is an acceptable result.

To conclude this section, Middle Welsh presents a good test case for POS tagging a historical corpus of a language with rich verbal and prepositional inflection and non-standardised orthography. The MBT showed robust results and flexibility with the highly variable orthography of minimally pre-processed Welsh texts. The parameter settings of the MBT software allow for focus on the context and the last 3 letters of unknown words. Since Middle Welsh verbal endings usually consist of 2/3-letter suffixes (reflecting tense, mood, aspect, person and number combined), it is not difficult for the tagger to predict the right form (e.g. *gwel-eis* ‘I saw’ as VBD-1SG denoting ‘first-person singular preterite’). Other parameter settings like an additional focus on the first 3 letters of the word proved to be less helpful for a language like Welsh with initial consonant mutation. This might, however, improve the results for languages with a strong prefixing preference, like for example Navajo (Young and Morgan 1980: 103, 107).

4 Annotating syntax and information structure

In order to facilitate syntactic queries, the above-described morphosyntactic annotation was employed to develop hierarchical phrase structure as well. A full parse would require a detailed Context-Free Grammar or Dependency Grammar. Developing this was beyond the scope of the present study, however. Instead, I modified the rule-based chunk-parser available in the Natural Language Toolkit (NLTK) in such a way that not only phrasal chunks but also some theory-neutral hierarchical structure could be added.

4.1 Designing the rule-based grammar

The NLTK rule-based chunk-parser is a regular expression parser: it systematically combines POS tags as defined in a grammar that allows regular expressions to create more (specific) options. Frequently used regular expressions include:

? ⇒ for optional preceding items | ⇒ ‘or’

The combination of words with their POS tags into phrases, e.g. Noun Phrases (NPs), Determiner Phrases (DPs) or Prepositional Phrases (PPs), is achieved with the following sample pattern of commands:

- NP: {<N|NPL|NPR>}
- DP: {<D><ADJ|ADJPL>?<NP>}
- PP: {<P><NP|DP>}

According to the above set of rules, a noun phrase (NP) can be formed of words with one of three different POS tags: a noun (N) *or* a plural noun (NPL) *or* a proper noun (NPR). Similarly, a DP, in this grammar, is formed by a determiner (D) followed by a noun phrase (NP) with an optional adjective (singular ADJ or plural ADJPL) in between. The order in which this rule-based grammar operates is important. The DP-rule above must follow the NP-rule to find the label <NP>. In this way single-layered hierarchical structures (NPs within DPs) were created. Similarly, a further layer could be created resulting in a PP containing a DP containing an NP, as long as they are called in the right order.

This is all straightforward in a language with extremely simple noun phrases and/or with a very limited amount of POS tags. Middle Welsh noun phrases, however, present some problems in this respect. First of all, some adjectives either follow or precede the noun they modify, with different meanings in either of the two positions. In addition to this, possessive pronouns and quantifiers can be part of the noun phrase as well. Furthermore, demonstratives

must follow the noun (and its modifying adjectives) and they are also obligatorily accompanied by the definite article preceding the noun phrase, as in (2). Finally, phrases with numerals in Welsh come in many shapes and forms, as (3) shows. Welsh numerals above ten can be split to occur before and after the noun phrase. In addition to that, phrases with numerals can also employ the preposition *o* ‘of’.

- (2) a. *y cathod mawr*
 DEF cats big
 ‘the big cats’
 b. *yr hen lyfr mawr hwn*
 DEF old book big PROX
 ‘this big old book’
- (3) a. *tair o ferched*
 three_{FEM} of girls
 ‘three girls’
 b. *tri o bobl eraill/newydd*
 three_{MASC} of people other_{PL}/new
 ‘three other/new people’
 c. *dau hen lyfr lyfr*
 two_{MASC} old book book
 ‘two old books’
 d. *un mlynedd ar ddeg*
 one year on ten
 ‘eleven years’

Complex noun phrases can also consist of two juxtaposed nouns, as in (4). In these constructions, the definite article only appears before the second noun, but the whole construction is definite.

- (4) a. *dyn y siop*
 man DEF shop
 ‘the man of the shop’
 b. *pob yn ail fis*
 every PRED second month
 ‘every other month’
 c. *yr holl broblemau*
 DEF all problems
 ‘all the problems’

The above types of complex noun phrases require a very detailed rule-based grammar that includes all possible phrases, including some phrases with special labels to facilitate further syntactic queries, e.g. phrases with verbal nouns (which can function as infinitives or nouns). Such a rule-based grammar explicitly looks for sequences of, for example, numerals + adjectives + nouns such as *dau hen lyfr* ‘two old books’ in (3c). Similarly, when put in the right order, complex noun phrases with quantifiers can be combined with explicit searches for the sequence determiner + quantifier + noun (D>Q>N). The full rule-based grammar I designed can be found in Meelen (2016: Appendix). This is a flexible template that can easily be extended and adapted to achieve better results when more texts are added.

4.2 Manual correction

No automatic NLP task is 100% correct. The rule-based chunk-parsers performs very well with simple matrix clauses, but subordinate clauses and some complex DPs in particular need some correction. I manually corrected the entire corpus using CesaX. CesaX is a special software package developed by Erwin Komen to facilitate corpus-linguistic research (cf. Komen 2011). Another useful feature of CesaX is that it can automatically convert the chunk-parsed .psd-files to XML-files with a simplified TEI P5 header. These files can then be queried using CorpusSearch¹¹ or the XML-based XQuery language. Manual correction in CesaX is quick and easy, because of its graphic representation of the tree structures. Alternatively, the bracket representation shown below, can also be edited manually with any text editor if needed:

```
(S (DP (NP (N taryan))(ADJP (ADJ eur))(NP (N grwydyr)))
(VP (PCL a)(VBD-3PL dodasant))
(PP (P dan)(DP (PRO-G y)(NP (N penn)))) (, .))
```

Figure 2: Sample bracket representation.

¹¹ CorpusSearch is a query language that finds syntactic structures in a corpus of annotated sentence trees. It can be used as a development tool for building the corpus or as a research tool to find and collate results in a corpus (see Section 5.1).

The above output from the automatic chunk parser reflects the following example:

- (5) *Taryan eur grwydyr a dodassant dan y penn*
 shield gold enamelled PTC put_{3PL.PRET} under 3SG_{MASC.POSS} head
 They placed a gold enamelled shield under his head.’ (Williams 1908:
 lines 18–19 [*Breuddwyd Maxen*])

4.3 Annotating information structure

Information-structural features were added semi-automatically. With CorpusStudio software (see Komen 2009 and Section 5), various features can be automatically added. Information for these features can be derived from the detailed POS tags of the specific words, from the phrasal structure and/or from the context in which it occurs. For example, since personal pronominal subjects usually convey ‘old’ information, with some simple XQuery commands the referential status of these subject pronouns can be automatically labelled ‘Old’. Other specific features of the clause such as the tense, aspect or mood of the verb or the person-number inflection can be derived from the detailed set of POS tags in the same way.

Further information-structural notions such as topic or focus are not as easy to detect automatically. However, if special focus words, pronouns or particles are used, these were labelled as such by the detailed POS tagger and therefore the focus domain or articulation of the sentence can automatically be annotated accordingly. In addition to this, constituent focus in Middle Welsh could be indicated by a (reduced) cleft and a verb with default third-person singular inflection. Pronominal subjects in the first or second person or plural full DPs can be automatically detected as well. When it comes to labelling the exact type of topic (e.g. familiar, aboutness or contrastive) or focus, manual annotation is still required.

All additional features (including the information-structural ones discussed here) are added at the matrix clause level. In practice, this means a list of features with automatically derived values (by querying the POS tags) and open values (to be adjusted manually) is available for every matrix clause. These features include:¹²

¹² These features are chosen because combined they cover all relevant information-structural notions. See Meelen (2016: Chapter 2) for further motivation with detailed examples from Middle Welsh and other languages.

- Focus articulation, e.g. constituent focus
- Focus particle/word, e.g. *hefyd* ‘also’
- Point of departure, e.g. temporal clause ‘at that moment . . .’
- Information flow, e.g. unmarked
- Referential state subject, e.g. old information
- Referential state object, e.g. new information
- Diathesis, e.g. impersonal verb
- Tense/aspect, e.g. imperfect
- Mood, e.g. subjunctive
- Semantic roles (in order), e.g. agent-patient
- Animacy and definiteness subject, e.g. definite-animate
- Animacy and definiteness object, e.g. indefinite-inanimate

5 A brief note on possibilities to query the data

There are various online tools available for corpus research, e.g. the search interface for the British National Corpus.¹³ Search interfaces provide easy access to the data, because no prior knowledge of specific search algorithms is necessary to get any results. The types of searches are often limited to the level of individual words or simple part-of-speech labels, however, which is not sufficient for syntactic or more detailed linguistic analyses. If we want to gain a deeper insight in our linguistic data, we need a more thorough way of searching for the right information.

5.1 Extended querying with CorpusStudio and CesaX

For many historical syntacticians, CorpusSearch¹⁴ is a useful application that can retrieve the detailed linguistic data relevant to them. It enables queries in the treebank or labelled bracketing format (the .psd format described above). With CorpusSearch, these first POS-tagged and chunk-parsed files can thus be easily queried and, for example, compared to data from historical corpora of other languages. It should be noted, however, that this first partial corpus only contains an extended shallow parse without syntactic empty categories yet.

¹³ www.natcorp.ox.ac.uk/

¹⁴ <http://corpussearch.sourceforge.net/>

These will, however, be added to PARSHCWL in the future (see Meelen and Willis *forthc.*).

Another way to retrieve detailed syntactic information is by converting the (parsed) files to XML format and query them with the usual search function for XML-databases: XQuery and XPath. Erwin Komen developed a wrapper around CorpusSearch2 (Randall, Taylor and Kroch 2005) and XQuery to facilitate these searches: CorpusStudio (Komen 2009). CorpusStudio not only simplifies the task of formulating search queries, it also provides easy ways to organise them along with the corpus data and research logs documenting your goals, subqueries, definition files and any emendations while gathering the right data.

5.2 Textual markup

For the metadata markup, I chose a simplified version of the TEI P5 (TEI Consortium 2009) header that is suitable for philological data, translations and linguistic annotation in XML format. This simplified TEI P5 header was selected because the full header with all its details was unnecessary and inefficient to work with. In addition, parsed files can be converted to this simplified TEI P5 header automatically by the CesaX software (see Section 4 and Komen 2011). Any information about the philological background of the text can be stored in this header and easily retrieved for future online usage. In the textual markup, any changes to the annotation can be indicated as well to trace the history of the annotated text and corpus as a whole. Finally, it would ultimately be possible to combine different versions of the texts (i.e. diplomatic and critical editions) into one XML file to make sure invaluable philological information is not lost. Its systematic but flexible nature would allow future conversion to JSON and RDF format as well.

6 Conclusion

This chapter presents the first steps towards the creation of a fully annotated corpus of historical Welsh. The above description of the proposed procedure is meant as a blueprint for the development of a fully parsed historical Welsh treebank (PARSHCWL) in the future (see Meelen and Willis *forthc.*). I described how a combination of minimal pre-processing, a systematic extension of current tag sets for historical corpora and a hierarchical way of chunk-parsing can yield important information needed to address questions about the syntax and

information structure of the Welsh language that have hitherto been unanswered. At the current stage, the annotation of the corpus was done in such a way as to optimise the search queries specific to the change of word order in the early Middle Welsh period (see Meelen 2016). However, the same annotated partial corpus of Middle Welsh was also already used for studies on adjectival agreement in native and translated Welsh prose (see Meelen and Nurmio 2020; Parina and Poppe *forthc.*).

The flexible XML-based nature (compatible with the .psd file structure) of the corpus means that any further philological or linguistic annotation can be added at a later stage as well. At various stages in the process of creating the corpus, manual correction was necessary. For this pilot, there was only one annotator available to do the manual correction of the limited pre-processing and more-detailed POS tagging and parsing tasks. Therefore, checking cross-annotator agreement, which is needed to verify the results, was not an option. In future, when making the annotated files accessible for everyone online, a final check will be done to filter out any possible mistakes and/or inconsistencies.

This chapter presents a good test case for annotating a partial historical corpus of a language with rich verbal and prepositional inflection. The main challenges in building annotated corpora like these lie in the availability of good digitised diplomatic or critical text editions. Further collaboration with scholars specialised in the philological background producing these editions can help linguists to make the right decisions, both in terms of selecting the right texts and editions for the corpus, but also in pre-processing and tokenisation in particular. More elaborate pre-processing of the texts, including the development of a good stemmer to do normalisation etc. and expanding the training set takes time, but will yield better results for the automatic POS tagging and parsing tasks in the end. A relatively large set of over >200 morpho-syntactic tags was developed and presented here, because those details give important new research opportunities for both Welsh philologists and linguists. A standardised way of expanding the tag set for rich inflectional languages is called for and the proposed extensions outlined in Section 3 above aims to be a good starting point for future extension and refinements when more texts are added to the corpus (see Meelen and Willis *forthc.* for this and subsequent steps towards building a fully parsed Welsh treebank).

Theodorus Fransen

3 Automatic morphological analysis and interlinking of historical Irish cognate verb forms

1 Introduction

The main aim of the author's research project is to use computational approaches to gain more insight into the historical development of Irish verbs. One of the objectives is to investigate how a link between the electronic Dictionary of the Irish language (eDIL),¹ covering the period c. 700–c. 1700, but focussing on Early Irish (7th–12th centuries), and the nascent *Foclóir Stairiúil na Gaeilge* 'The Historical Dictionary of Irish',² covering the period 1600–2000, could be implemented. Such a link will be hugely beneficial for scholars operating at the intersection of the medieval and modern period (see Table 1), who currently lack a comprehensive lexical resource for the "intermediate" early modern period.

The above-mentioned lexicographical discontinuity is problematic, and needs to be remedied, especially in the light of the pervasive changes in the verbal system between Early and Modern Irish. The author's motivation for focussing on the verbal system in Early Irish resonates with the following observation by McCone in his authoritative monograph on the Early Irish verb:

Concentration upon the verb was dictated by its generally conceded status as the most difficult and interesting area of Old and Middle Irish morphology and few would deny that an understanding of the Old Irish system's workings and development into and through Middle Irish is a prerequisite for being able to deal with the abundance of Old and Middle Irish texts effectively. (McCone 1997: xviii)

During the author's research it was found that eDIL does not provide full verb paradigms for many verb entries. It was felt that additional language technology is necessary to deal with the complex Early Irish verbal system. Such technology will also facilitate more systematic and comprehensive interlinking of verb forms in lexicographical resources. The main contribution to this end by the author is the development of a morphological analyser for Old Irish, which is also the focus of this paper.

1 Available at: <http://dil.ie/> [accessed 7 February 2019].

2 Available at: <https://www.ria.ie/research-projects/focloir-stairiuil-na-gaeilge> [accessed 7 February 2019].

In order to make this contribution accessible to (computational) linguists whose research area is not Old Irish, a brief overview of the Irish language periods (section 2) and the basics of the Old Irish verbal system (section 3) is provided. The latter aims to show how phonology imposes itself on verb morphology, resulting in an often complex relationship between an underlying verb root and a verb's multiple surface shapes – an insight crucial for the computational implementation. Section 4 sums up important changes in the verbal system in Middle Irish and beyond. In the second half of the paper, the focus is on digital resources for historical Irish and Natural Language Processing methods. Section 5 surveys important existing digital resources and computational methods used to deal with historical texts. The proposed methodological framework of the paper is the topic of section 6. Section 7 introduces finite-state morphology and presents some highlights, as well as challenges, in the development of a morphological analyser for Old Irish verbs. The formulation of clear-cut verb stem entities constitutes a key feature in the implementation. Suggestions for automatically linking cognate verb forms are presented in section 8. A synthesis of matters discussed in this paper follows in section 9, which also outlines some research prospects.

2 A historical sketch of the Irish language

The historical period of Irish can be divided into the language stages shown in Table 1 below. Greene (1966) provides a succinct overview of the history of the Irish language. Early Irish represents the language from the early medieval period up until about 1200. After that we speak of Modern Irish. Old Irish, like the modern standardised language, can be treated as a normative phase in the

Table 1: Medieval and Modern Irish language periods.

	Language stage	Time period
Early Irish	Old Irish	7th–9th centuries A.D.
	Middle Irish	10th–12th centuries
Modern Irish	Early Modern Irish (including Classical Modern Irish)	13th –mid-17th centuries
	Post-Classical Modern Irish	mid-17th–mid-19th centuries
	Irish of the Revival period	late 19th–early 20th centuries
	contemporary standardised Modern Irish	1958–present

history of the language. Indeed, “classical” Old Irish, the language as witnessed predominantly in Old Irish glosses in Latin manuscripts, is the basis for many grammars and handbooks, including *A Grammar of Old Irish* by Rudolf Thurneysen (1946) (GOI). While representing a stable and normative phase in the language’s history (McCone 1997: 166), Old Irish shows diachronic as well as synchronic linguistic variation (for a discussion of the latter see McCone 1985).³ However, the linguistic variation in Old Irish is negligible compared to the unstable and highly variable language seen in Middle Irish texts. As McCone (1997: 166–167) has pointed out, Middle Irish comprises standard Old Irish forms and forms anticipating Modern Irish usage, as well as forms that are consonant with neither. The end of the Middle Irish period sees the production of the great medieval Irish manuscripts.⁴

The subsequent Early Modern Irish period (13th–mid-17th centuries) is dominated by a literary genre of praise poetry in syllabic verse composed by court poets, referred to as Classical Modern Irish (McManus 1994). In contrast to the highly regulated grammar of this bardic poetry, however, hugely varying registers can be observed with prose texts of this period, ranging from archaic language to registers that are not far removed from 19th-century Irish (Ó hUiginn 2013: 87–89).

Post-Classical Modern Irish refers to the literary period between the downfall of the Irish-speaking aristocracy in the early 17th century and the Great Famine (1845–1849), which is characterised by – amongst other developments – a more regional orientation in writing (Ó Háinle 2006). The classical literary standard that had emerged in the early modern period gradually gives way to writing conventions that more closely reflect the contemporary spoken language, resulting in the coming to the fore of the Irish dialects in texts of this period (Williams 1994).

The period between the Great Famine and the creation of the Free State (1922) is known as the Gaelic Revival, which witnessed an increased production of original work, facilitated by institutions such as the *Conradh na Gaeilge* [Gaelic League], established in 1893 (Mahon 2006). After independence, plans were made for a standardisation of Irish grammar and spelling, ultimately codified in a 1958 booklet published by the Irish government’s Translation Department (with further revisions in 2012 and 2016).⁵

3 Ó Cróinín (2001) discusses diachronic orthographical developments in the earliest Old Irish glosses. Two conventions, the “Irish” and “British” system, seem to have competed with each other; the latter ultimately became the standard for all subsequent Irish literature.

4 *Lebor na hUidre*, Rawlinson B 502 and The Book of Leinster (*An Leabhar Laighneach*) (Breatnach 1994a: 222–225).

5 Available at: https://data.oireachtas.ie/ie/oireachtas/caighdeanOifigiuil/2017/2017-08-03_an-caighdean-oifigiuil-2017_en.pdf [accessed 7 February 2019].

3 The Old Irish verb: The morphology-phonology interface

3.1 The main skeleton of the verbal complex

In general, Old Irish is a VSO language (Russell 2005: 430). However, additional variant structures are found (Mac Coisdealbha 1998), especially regarding the subject position (Lash 2014b). Both the verb, subject and object may be contained within the “verbal complex” (see McCone 1997: 1–19), comprising everything that falls within the accentual domain of the verb (Stifter 2009: 84), as such potentially constituting a highly synthetic “word”. Leaving aside copula constructions, Old Irish inflected verb forms incorporate the subject; no independent subject pronouns exist. Third person forms – from the viewpoint of word-based parsing – are inherently ambiguous in that there might or might not be an independent subject. In the present work, as is customary, third person verb forms are not glossed with a pronominal subject in the English translation.

There is a distinction between “simple” and “compound” verbs. Verbs with the verb root as their sole lexical element, as in (1) and (2), containing root *ber*, are called simple. A compound verb additionally takes one or more preceding lexical preverbs (PV), originating in prepositions,⁶ modifying the meaning of the verb root. In (3) and (4), the preverb is underlying/historical *to* combined with the root *ber*. As a rule, the first preverbal element within the accentual domain of the verb is realised as a proclitic, resulting in a juncture between, put simply, a prefix and the stressed part of the verbal complex, as in (2)–(4). This juncture is denoted by a mid-high dot to facilitate grammatical analysis; it is not present in manuscripts.

- (1) *beirid*
 carry_{3SG.PRES}
 ‘carries’
- (2) *ní·beir*
 NEG·carry_{3SG.PRES}
 ‘does not carry’
- (3) *do·beir (to·ber-)*
 PV·bring_{3SG.PRES}
 ‘brings’

⁶ Called thus by Thurneysen (GOI §§ 819–821).

- (4) *ní·tabair (to-ber-)*
 NEG·bring_{3SG.PRES}
 ‘does not bring’

Some commonly used grammatical notions relating to stem and ending formation are key to understanding the workings of the verbal complex. First, there are two ending sets, “absolute” and “conjunct”. Only simple verbs can take absolute endings, and only when occurring in clause-initial position. The conjunct ending set applies when a verb is conjoined with a preverbal element; compound verbs therefore invariably carry conjunct endings, while simple verbs take this set of endings when preceded by the preverbal “conjunct particles” (C), e.g. the negative particle *ní* ‘not’, as in (2). In (1), *-id* is the third singular present indicative absolute ending. The corresponding conjunct ending is seen in (2)–(4), where palatalisation of the root-final *r* (orthographically encoded by preceding *i*) is the only marker of inflection.

A verb preceded by a conjunct particle is said to be dependent, and independent otherwise. The distinction between independent and dependent has major repercussions for the surface shape of especially compound verbs. Generally speaking, an independent compound verb appears in its “deuterotonic” form as the first preverb is realised as a proclitic, causing the stress to fall on the second element (the verb root in [3]). When the proclitic “slot” in the verbal complex is occupied by a conjunct particle, as in the (dependent) compound form in (4), the stress is on the verb’s first preverb; this stem alternant is accordingly called “prototonic”.⁷ As (3) and (4) illustrate, the stress system of Old Irish may result in “complex synchronic morphophonemic alternations” (Stifter 2009: 90) and, consequently, a system of “double stem formation” (Russell 2005: 431). The abundant allomorphic variation seen in the Old Irish verbal system raises a question crucial for implementational purposes: what exactly is a verb stem in Old Irish? Section 7.2.2 will detail how this question has been tackled from a computational point of view.

⁷ There are some exceptions to the rules laid out here in relation to dependency, the distinction deuterotonic/prototonic and the set of endings that is demanded. In the imperative only conjunct endings exist, and compounds in the imperative appear in their prototonic form regardless of dependency. A further anomaly exists with compounds whose first preverb is either *to*, *fo* or *ro*, which may equally assume their prototonic form in independent position if the following element starts with a vowel, causing vowel elision (McCone 1997: 3). Proclitic conjunctions such as *co* ‘that’ (GOI § 896) may be found with either independent or dependent verb forms, i.e. they sometimes assume the status of conjunct particle.

Compound verbs may take up to four preverbs, each of which adhere to a positional hierarchy tentatively formulated by McCone (1997: 89–90). Verb roots cannot be arbitrarily compounded with any preverb. However, most verbs are liable to being (further) compounded with an “augment”. While a lexical preverb in origin, the augment has developed a “modificatory function that belongs to the grammar of Old Irish and not to its lexicon” (McCone 1997: 91). This preverbal particle supplies either a resultative or potential meaning, depending on the tense and/or mood of the verb form that it occurs with, illustrated with (5) and (6), respectively.

(5) *ro·léic*
 AUG·let_{3SG.PRET}
 ‘has let’

(6) *as·robair (ess-ro-ber-)*
 PV·say_{AUG.3SG.PRES}
 ‘can say’

The augment is most commonly *ro* (position 4), while the augments *ad* (position 3) and *cum* (position 4) are more restricted – i.e. the latter two co-occur with a limited set of (lexical) preverbs.⁸ For a discussion on the preverbal particle *ro* and other augmentation strategies see GOI (§§ 526–537). Simple verbs (which do not have a preverb) almost always take *ro*, rather than *ad* or *cum*. The augment adds to the already abundant allomorphic variation seen in stem formation and its position is subject to change during the Early Irish period, in parallel with other processes of reorganisation and simplification of the verbal system,⁹ most importantly the univerbation¹⁰ of old compound verbs.

The morphosyntax or morphotactics of the verbal complex, i.e. the legal combination of morphemes (Beesley and Karttunen 2003: 26–27), with optional morphemes in brackets, is schematically summarised in (7) (C = conjunct particle, * = zero or more, with the provision that the total of preverbs does not exceed four, E = ending). Table 2 shows the schematic structure of the verbal complex with examples of preterite formations (with the conjunct third person singular

⁸ *ad* and *cum* are underlying forms, subject to a substantial amount of allomorphic variation depending on whether they are stressed or not.

⁹ Already in the Old Irish period, and during the Middle Irish period, *ro* is gradually adopting the status of conjunct particle, mitigating its effects in stressed position. The positional behaviour of *ro* and its semantics is outside the scope of the present paper; the reader should refer to McCone (1997: 127–161) for a detailed description of this preverbal particle.

¹⁰ A lexicalisation process involving the “unification . . . of a syntactic phrase or construction into a single word” (Brinton and Traugott 2005: 48).

Table 2: Schematic structure, including the position of the augment *ro*, of the Old Irish verbal complex, adapted from McCone (1997), with third person singular examples of unaugmented and augmented preterite forms with root *lēc*, illustrating combinatorial possibilities and allomorphic variation in stem formation (C = conjunct particle, * = zero or more, with the provision that the total of preverbs is not more than four). For Old Irish phonemes and their graphemic representation see Stifter (2006: 377–379).

Lemma	Structure (<i>ro</i> = augment <i>ro</i>)	Dependency	Ending	Example, pret. 3sg. (bold = lexical element, italics = stressed syllable)
<i>léicid</i> (simplex)	i. VROOT E	indep.	abs.	léicis /l'e:g'əs'/ let _{3SG.PRET}
	ii. C · VROOT E	depend.		ní- léic /n'i: 'l'e:g'/ NEG·let _{3SG.PRET}
	iii. ro · VROOT E	indep.		ro-léic /ro 'l'e:g'/ AUG·let _{3SG.PRET}
	iv. C · ro VROOT E	depend.		ní- reilic /n'i: 'r'el'əg'/ (ro-lēc-) NEG·let _{AUG.3SG.PRET}
<i>do-léici</i> (compound)	v. PV1 · PV* VROOT E	indep. (deut.)	conj.	do-léic /do 'l'e:g'/ (to-lēc-) PV·cast _{3SG.PRET}
	vi. C · PV1 PV* VROOT E	depend. (protot.)		ní- teilic /n'i: 't'el'əg'/ (to-lēc-) NEG·cast _{3SG.PRET}
	vii. PV1 · PV* ro PV* VROOT E	indep. (deut.)		do-reilic /do 'r'el'əg'/ (to-ro-lēc-) PV·cast _{AUG.3SG.PRET}
	viii. C · PV1 PV* ro PV* VROOT E	depend. (protot.)		ní- tarlaic /n'i: 'tarləg'/ (to-ro-lēc-) NEG·cast _{AUG.3SG.PRET}

form having a zero ending); it illustrates how the stress pattern (phonology) of Old Irish impacts on the verb morphology. The situation is slightly simplified in that *ro* represents the augment; *ro* is the particle's most common allomorph and the one found with *léicid* 'lets' and *do-léici* 'lets go, releases, casts'.

(7) (C) PV* (AUG) PV* VROOT E

3.2 Adjuncts and *notae augentes*

The skeleton of the verbal complex, outlined in (7), allows for incorporation of unstressed, clitic “adjuncts” (McCone 1997: 9) and *notae augentes*, illustrated in this subsection with various examples.¹¹ Independent simple verbs allow a pronominal object, e.g. *-us* in (8), illustrated with the verb *benaid*.¹²

(8) *bentus*
 strike_{3SG.PRES}-3SG_{FEM}
 ‘strikes her’

The alternative strategy is to employ infixation, as in (9), with the pronoun *-m* attaching itself to the available proclitic (here *ní*). Infixed pronouns directly precede the proclitic boundary and come in three classes (GOI §§ 409–427); the choice between class A and B is phonologically conditioned, whereas the choice for class C is conditioned based on syntactic grounds. Simple verbs without a preceding proclitic element acquire the “meaningless” preverbal particle *no* for purposes including infixation of pronouns, illustrated in (10). Infixed pronouns are often accompanied by following initial mutations (which are often not orthographically marked due to the underspecified nature of the Old Irish spelling system).¹³

¹¹ Unless referenced explicitly, examples are either hypothetical or sourced from eDIL.

¹² The derivation is *ben(a)ith + us* with subsequent syncope (for which see 3.4) and delentition of *th* after *n*.

¹³ For the initial mutations see GOI (§§ 229–244). Lenition is the pronunciation of consonants with less acoustic energy. As Thurneysen has pointed out, scribal evidence of lenition in Old Irish is initially confined to the letters *p*, *t*, *c* which turn into fricatives, marked by a following *h* (*ph* /f/, *th* /θ/, *ch* /x/). Lenition of *f* and *s* is not indicated in the earlier glosses. Lenited *f* is silent and may be omitted altogether in the spelling; lenited *s* represents /h/. In the course of Old Irish, lenition is also marked on *f* and *s* by employing a *punctum delens* (*ḟ*, *ṡ*). Nasalisation refers to the prefixing of *n* to an initial vowel and the homorganic nasal to *b* and *g* (*mb* /mb/, *ng* /ŋg/), voicing of *p*, *t*, *c* and *f* (hardly ever expressed in the spelling) and gemination with *s*, *r*, *l*, *m*, *n* when preceded by a proclitic vowel (not always marked in the spelling).

- (9) *ním·beir*
 NEG-1SG·^{LEN}carry_{3SG.PRES}
 ‘does not carry me’

- (10) *nom·beir*
 PV-1SG·^{LEN}carry_{3SG.PRES}
 ‘carries me’

Special relative endings exist for independent simple verbs for the absolute third person singular and first and third person plural, e.g. (11). In other cases, relativity is marked by an initial mutation following the proclitic preverb, as in (12), or suffixing *-e/-a*, followed by lenition, in case of the preverbs *im(m)* (exemplified in [13]) and *ar*.¹⁴

- (11) *léices*
 let_{3SG.PRES.REL}
 ‘who lets, which (s)he lets’

- (12) *do·léici*
 PV·^{LEN}cast_{3SG.PRES}
 ‘who casts, which (s)he casts’

- (13) *imme·thét (imbi-tēg-)*
 PV-REL·^{LEN}go.about_{3SG.PRES}
 ‘who goes about, which (s)he goes about’

The enclitic *notae augentes* occur in final position in the verbal complex and reinforce an already present subject or object, as in (14).¹⁵

- (14) *at·beir=som (ess-ber-)*
 PV-3SG_{NEUT}·say_{3SG.PRES} =3SG_{MASC/NEUT}
 ‘he says it’

We arrive at the schematic overview in (15), loosely based on McCone (1997: 17) (C = conjunct particle, * = zero or more, with the provision that the total of

¹⁴ Occasionally *a/e* appears with other preverbs (GOI § 493.4): *reme·* (for *remi·*), *iarma* (for *iarmi·*, *iarmu·*) and *assa·* (instead of *as·*).

¹⁵ Examples of this form in the glosses are cited in Griffith (2008: 59).

preverbs does not exceed four, E = ending, A = adjunct, N = nota augens; A and N cannot occur together in an independent simple verb).

- (15) indep. simple: VROOT E_{ABS} (A) (N)
 indep. simple augm.: AUG (A) · VROOT E_{CONJ} (N)
 depend. simple: C (A) · (AUG) VROOT E_{CONJ} (N)
 indep. compound: PV1 (A) · PV* (AUG) PV* VROOT E_{CONJ} (N)
 depend. compound: C (A) · PV1 PV* (AUG) PV* VROOT E_{CONJ} (N)

Taking together all inflectional forms across the tense/mood paradigms, we arrive at about one hundred and twenty inflected forms per verb. If we include affixed adjuncts, augments and *notae augentes*, we are talking about several orders of magnitude more. This “combinatorial” problem is compounded by the fact that scribal practice was often to present the composite elements in the verbal complex as a concatenative string. This results in segmentation challenges, which will be addressed in 7.2.3. Essentially, a computational framework should be able to identify the verb root and all its surrounding elements in strings without mid-high dots, spaces and hyphens, as in *nondobmolorsa* in (16), found in the Würzburg Glosses (Wb.) (*Thes.* 1: 593). This example contains the first singular present indicative of the deponent verb (see 3.3) *molaitir* ‘praises’ with a *nota augens*, preceded by the meaningless preverbal particle *no*. In the indicative and subjunctive tenses, *no* is used to infix relative *n*,¹⁶ signalling a nasalising relative clause (GOI §§ 497–504). This is what we have in (16), enforced by the conjunction *hore* (*óre*, [*h*]úare) ‘because’ and realised by nasalisation of initial *d* of the infixed pronoun.¹⁷

- (16) *hore no -n -dob mol -or -sa*
 PV ^{-NAS} -2PL · praise ^{-1SG.PRES} =1SG
 ‘because I praise you’ (Wb. 14^c18)

¹⁶ <http://dil.ie/33202>.

¹⁷ For the initial mutation called *nasalisation* see fn. 13.

3.3 A brief description of stem and ending formation

Apart from the rather small class of hiatus verbs, with roots ending in a vowel, Old Irish exhibits an opposition of weak (W1/W2) and strong verbs (S1–S3), which are classified according to present stem formation.¹⁸ Verbs have five stems: present, subjunctive, future, preterite and preterite passive. Stem formation with weak verbs is through largely regular and hence predictable suffixation. Strong verbs show a combination of suffixation, vowel alternations (ablaut) and reduplication, which are largely unpredictable unless one knows the underlying abstract root shape (Stifter 2009: 96). For instance, *crenaid* ‘buys’ has a future 1sg. conj. *·cíur*, which can be explained by reduplication of the abstract root *cri* → *ci-cr* ... and subsequent lengthening of *i* to compensate for the disappearance of lenited (fricativised) postvocalic *c* before *r* (GOI §§ 71, 691). While Old Irish verb morphology abounds in complex allomorphic stem alternation, further complicated by analogy (for an example see 7.2.2), the term “irregular” is arguably best reserved for suppletion, i.e. usage of different roots across a verb’s paradigm.

There are six groups of ending sets which are not arbitrarily combinable with the five stems (Stifter 2009: 96). Apart from the imperative and “secondary” endings (used with the imperfect, past subjunctive and conditional), all ending sets come in two series, i.e. absolute and conjunct, albeit only relevant for simple verbs (see 3.1). Both suffixation and stem-internal modifications are employed in ending formation. The latter comprise alternation of the root-vowel, the change of quality ([non-]palatalisation) of the root-final consonant and the insertion of *u* into the stem (“*u*-infection”, Stifter 2009: 67).

There are separate inflectional endings known as deponent, used with a limited set of verbs. While appearing as passives due to endings in *-r*, deponent verbs convey active meaning; the deponent property is therefore “merely lexical”, and, consequently, “has to be known for each verb separately” (Stifter 2009: 87).

18 The classification system is the one used in McCone (1997). GOI employs A for weak (and hiatus) verbs, and B for strong verbs (with further subclassifications using Roman numerals). McCone’s classification is used here as the letters W and S are more obvious designators for verb type, and a third letter H is reserved for hiatus verbs. Furthermore, McCone’s classification reflects a re-examination of inflectional patterns, more clearly showing similarities between inflectional classes (using a subclassification systems of Arabic numerals followed by (optionally) the letters *a*, *b* and *c*). A conversion table is found in Stifter (2006: 381), who also adopted McCone’s classification system.

3.4 Syncope

Syncope is the deletion of vowels in even-numbered, non-final syllables in words with more than two syllables (Stifter 2006: 49). In verb forms, the syncope rule operates from the first stressed syllable onwards, that is, the one immediately following the proclitic juncture of the verbal complex. The addition of a *nota augens* (see 3.2) does not cause syncope. The effects of syncope are most pronounced in compound verbs (GOI § 107), where alternation in stress causes much allomorphic variation in the verb stem and the preverbs (see 3.1), e.g. (viii.) *ní-tar[†]laic*¹⁹ (*to-ro-lēc-*) in Table 2, with deletion of *o* in *ro*. A syncopated front vowel (*e*, *i*) results in a palatalised consonant cluster, while a syncopated back vowel results in a non-palatal cluster. The latter explains the surface form *·tarlaic*, where the consonant cluster *·rl-* becomes non-palatalised because the syncopated vowel was a back vowel, with verb root *lēc* surfacing as *laic* /læg^j/. There are many attested instances of irregularly applied syncope; an in-depth discussion of some irregular patterns is provided in Ó Cruaíoch (1999); see also 7.2.2 in the present chapter.

4 The verb in Middle Irish, and beyond

The Old Irish verbal system undergoes major changes in Middle Irish, eventually resulting in a much-simplified inflectional system in Modern Irish. The key Middle Irish developments are documented in detail in Breatnach (1994a: 278–325) and McCone (1997: 163–241). The changes between Early and Modern Irish are summarised in a–c below.

- a. Development of an immutable root shape and transparent stem formation, i.e. univerbation of compound verbs and, as mentioned in 3.1., the gradual development of *ro* as a conjunct particle (Breatnach 1994a; McCone 1997).
- b. Replacement of affixed pronominal objects by independent object pronouns (Breatnach 1994a; McCone 1997);
- c. Homogenisation (and later renewal) of personal endings, the gradual emergence of independent subject pronouns (outside copula constructions) and, in conjunction with this, analytic verb forms (Breatnach 1994a; Greene 1958; Greene 1973; McCone 1997; McManus 1994).

Developments a. and b. reach completion in Middle Irish, while the developments in c., apart from the streamlining of present and preterite endings, are present in

¹⁹ The dagger denotes a syncopated vowel.

embryonic form (subject pronouns) or take place, for the most part, in Early Modern Irish (the development of analytic verb forms). A comprehensive discussion of these pervasive changes is outside the scope of this paper, but some important references have been provided.

The opposition of deuterotonic and prototonic and associated morphophonemic variation was largely done away with by creating new (generally weak) simple verbs based on mainly old prototonic compound bases (McCone 1997: 192–193). This can be illustrated with *do-léici*, prototonic *-teilci*, developing into the simple verb *teilcid* on the basis of analogy with the simplex: *léicid: léici*, *x: -teilci*, *x = teilcid*. A more extreme example of stem simplification is the Old Irish compound verb *do-sluindi*, *·diltai* (*dī-slond-*) ‘denies’, developing into the Middle Irish simple verb *diltaid* (McCone 1997: 207–209), which is the basis for the modern stem *diúltaigh*. These examples illustrate how a verb stem or lemma can change beyond recognition between Old and Modern Irish.

5 Survey of digital resources and computational methods

5.1 Overview

This section gives a survey of resources and tools to be incorporated in – or potentially useful for – the author’s research. The main goal is to illustrate the under-resourced status of historical Irish. The introduction to the present volume already documents the available lexical resources and corpora for Early Irish. This section, therefore, focusses on resources for Modern Irish. The most important digital resources are plotted on a timeline in Figure 1. A distinction is made between lexicons and corpora, which are discussed separately in 5.2 and 5.3, respectively. The picture that emerges is one of fragmentation and, especially in the case of lexicons, discontinuity. We are faced with a “lexicographical gap” in the middle, roughly corresponding to the Early Modern Irish period (13th–mid-17th centuries). Discussing modern scholarship and bardic poetry, Mac Cárthaigh (2018: 28) observes that “we still lack such basic infrastructure as a dedicated dictionary for the [Classical Modern Irish] period”. Similar observations have been made in Griffith, Stifter, and Toner (2018), who provide a comprehensive research survey on Early Irish lexicography. Subsection 5.4 provides a short excursion into Natural Language Processing for historical texts, and efforts made so far in this area in the Irish context, paving the way for the author’s proposed methodology in section 6.

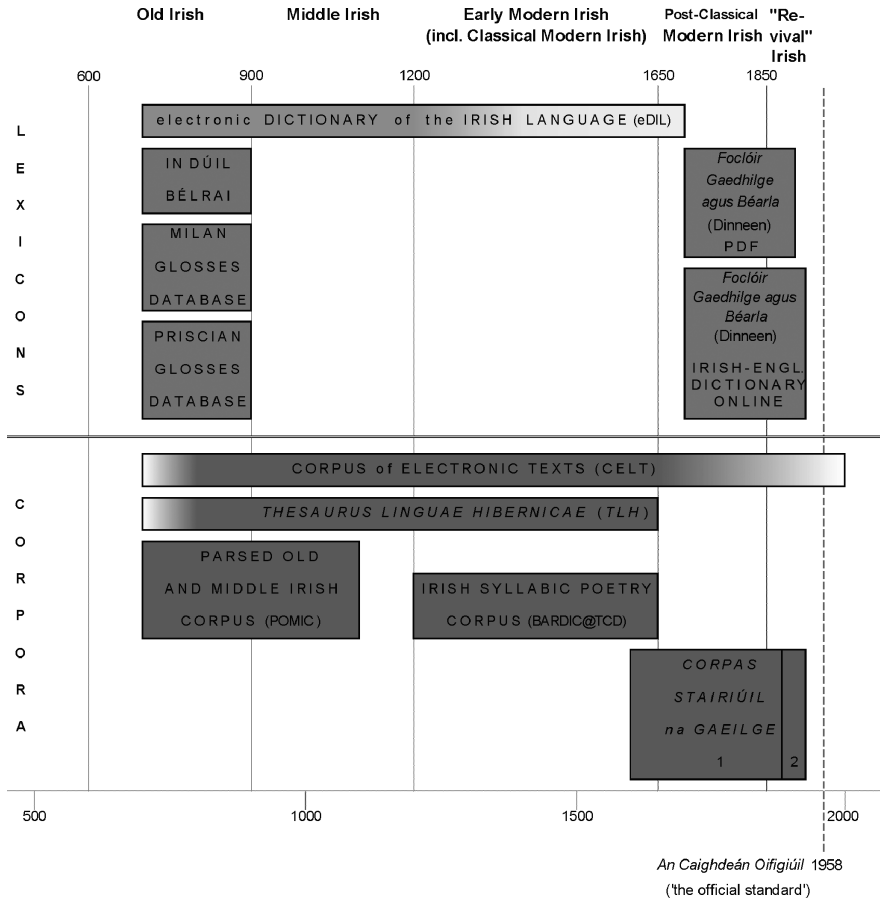


Figure 1: Visual representation of available digital linguistic support for historical periods of Irish. Lighter shades denote lesser support.

5.2 Lexicons

The Dictionary of the Irish Language (DIL) is the only dictionary that bridges the Early and Modern Irish period and “its publication as an electronic resource has been a great boon” (Stifter 2009: 59). However, the resource is not an ideal starting point for an Old Irish morphological parser due to aspects of structure and contents, inherited from the original hard copy. For example, the dictionary is far from exhaustive in listing inflected forms. Other limitations, some of which have meanwhile been remediated by the publication of the electronic

version, are discussed in Nyhan (2006). It should be added that the original objective of the eDIL project was not to revise the original hard-copy dictionary, but to open up the wealth of information contained in it and to make it accessible to a variety of users (Fomin and Toner 2006).

The most important dictionary for Post-Classical Modern Irish is Dinneen (1927), digitised versions of which were prepared in the context of a few independent projects. Publicly available resources include a PDF version of the first edition of the dictionary²⁰ as well as the online Irish-English dictionary,²¹ the latter allowing both English and Irish searches, including the option to be directed to the relevant scanned page of the 1927 edition. The research goals of another project, Digital Dinneen, bear resemblance to the goals of the present work. The aim of this unfinished and dormant project was to create an edition that could be integrated with (mainly) Early Irish resources, including an XML-encoded electronic Lexicon of Medieval Irish (Nyhan 2006),²² eDIL and CELT. The resulting infrastructure was envisaged to allow a user to follow a Modern Irish form back to its earlier forms (Nyhan 2008). No tools were implemented, but the Digital Dinneen project has produced a (not publicly available) XML-encoded version of Dinneen (1927).²³

5.3 Corpora

The Irish Syllabic Poetry (or “Bardic Poetry”) corpus (c. 1200–c. 1650) consists of approximately 2000 poems from the Classical Modern Irish period, including 500 previously unpublished ones edited in McManus and Ó Raghallaigh (2010). Corpus preparation and annotation is a joint effort by the Irish Department in

20 Available at: <https://celt.ucc.ie/Dinneen1sted.html> [accessed 30 January 2019].

21 Available at <http://glg.csisdmsz.ul.ie/index.php> [accessed 30 January 2019].

22 While extensively documented in a Ph.D. thesis, this resource is, unfortunately, not available. The following link to a sample of the Lexicon was kindly provided to me by Peter Flynn (email dated 4 November 2014), former manager of the Academic and Collaborative Technologies Group (ACTS), University College Cork IT Services: <http://research.ucc.ie/lexicon/sample> [accessed 4 May 2019].

23 For online information about this project, see <https://celt.ucc.ie/digineen.html> [accessed 7 February 2019]. Further information was obtained by means of email contact with Beatrix Färber (30/10/2014 and 03/11/2014), who had the initial idea for *Digital Dinneen*, and Julianne Nyhan (23/2/2012), who informed the present author that neither a lookup mechanism nor a search interface has been implemented.

Trinity College Dublin, the School of Celtic Studies (Dublin Institute for Advanced Studies) and Dr Katharine Simms of the History Department in Trinity College Dublin, who indexed the poems and has created a database, which is currently being updated.²⁴ As part of the new project BARDIC@TCD (Eoin Mac Cárthaigh and Elaine Uí Dhonnchadha), a POS-tagged corpus currently consisting of 500 syllabic poems has been made freely available and it will be updated regularly.²⁵

The tagging of the above-mentioned Bardic Poetry corpus employs automatic standardisation techniques which had already been developed in the context of *Corpas Stairiúil na Gaeilge* ‘Historical Irish Corpus’ (envisaged to comprise 90+ million words), constituting the basis for the Royal Irish Academy’s ongoing project *Foclóir Stairiúil na Gaeilge* ‘The Historical Dictionary of Irish’ 1600–2000.²⁶ Uí Dhonnchadha et al. (2014) report on the adaptation of the “modern” tagging tools for the second segment of this corpus (1882–1926), containing seven million words, many of which in a pre-standard orthography (before 1958; see also 5.4).

5.4 Natural Language Processing methods

Natural Language Processing (NLP)²⁷ is concerned with the ability of computers to process human language (Jurafsky and Martin 2009: 35). The NLP pipeline involves the following activities (in this order): tokenisation,²⁸ lemmatisation,²⁹ part-of-speech (POS) tagging³⁰ and syntactic parsing.³¹ A crucial activity in the case of historical texts (and non-standard language in general) is spelling normalisation, influencing all further language processing tasks

²⁴ Available at: <https://bardic.celt.dias.ie/> [accessed 7 February 2019].

²⁵ The project website with a link to the corpus is found at <https://www.tcd.ie/slscs/research/areas/corpora/bardic.php> [accessed 17 July 2020].

²⁶ Available at: <https://www.ria.ie/research-projects/focloir-stairiuil-na-gaeilge>. The corpus is found at <http://corpas.ria.ie/> [accessed 7 February 2019].

²⁷ Alternative names for the field are Speech and Language Processing, Computational Linguistics and Human Language Technology.

²⁸ Separating punctuation marks and other non-alphabetic characters from words (Jurafsky and Martin 2009: 67).

²⁹ Grouping inflected forms of a word under its base form, i.e. its lemma (Mitkov 2003: 744).

³⁰ Assigning a syntactic class marker (e.g. verb, noun) to each word in a corpus (Jurafsky and Martin 2009: 167). POS taggers may be rule-based or trained on annotated data (e.g. statistical), or both.

³¹ Parsing is a broadly defined concept in Speech and Language Processing that involves taking an input form and produce a structured linguistic representation. Parsing can be done on the morphological, syntactic, semantic and discourse level (Jurafsky and Martin 2009: 79).

(Piotrowski 2012: 11). Standardisation of historical forms to arrive at modern forms is best described as spelling modernisation (Piotrowski 2012: 69–70). The term “canonical cognate” is used by Jurish (2010) to refer to the mapping of an extant equivalent of a historical word that preserves the latter’s morphological root and morphosyntactic features. However, sometimes the aim is not to map a historical form to a modern form, but instead to a normalised or canonical historical spelling. This typically involves dealing with both diachronic and synchronic variation.

Using NLP to deal with language variation in historical texts is far from straightforward:

[T]here is no underlying computational model that describes how synchronic and diachronic variants relate to each other and – possibly – to some shared meaning or some kind of prototype that represents the relatedness of the variants (Piotrowski 2012: 9)

Piotrowski (2012) has pointed out that historical language is inconsistent and highly variable, which hinders POS tagging. The same author mentions various way of tackling this problem. Two common methods, often used in conjunction with each other, are:

- a. Bringing an older language variety in line with a standardised or normative – typically modern – variety (either by using rule-based or statistical methods) and use a “modern” POS tagger, if it exists.
- b. Employing already existing lexical resources, and create mappings across resources, i.e. through lemmas, dictionary headwords, etc.

In the *Foclóir Stairiúil na Gaeilge* ‘The Historical Dictionary of Irish’ project (1600–2000), a morphological analyser and POS tagger for the standard language (Uí Dhonnchadha and van Genabith 2006) are conjoined with a standardiser (*An Caighdeánaitheoir* [Scannell 2009, 2017]),³² employing rule-based and statistical methods and a lexical database of historical and modern word pairs, created by the project’s language experts (Uí Dhonnchadha et al. 2014). Initial evaluation of the POS tagging of the 1882–1926 segment of the corpus pointed to F-scores³³ ranging from 91–96% (Uí Dhonnchadha et al. 2014).

Dereza (2018), who discusses lemmatisation approaches for ancient and morphologically complex languages, reports that neither rule-based approaches

³² Code available at: <https://github.com/kscanne/caighdean/> [accessed 10 March 2020].

³³ A measure of a test’s accuracy that incorporates “precision” (e.g. what percentage of the items subjected to standardisation were correctly standardised?) and “recall” (e.g. what percentage of items which should have been standardised were actually standardised?) (Jurafsky and Martin 2009: 479).

(using stems and affixes) nor statistical machine learning methods are useful for Early Irish due to morphophonological complexity, non-transparent orthographical features and scarcity of data. She has developed an Early Irish lemmatiser using *form:lemma* mappings extracted from eDIL and compared two methods: 1) an approximate matching approach using a lemma predictor based on the Damerau-Levenshtein distance, checking for all possible strings of the forms on edit distance 1 and 2,³⁴ and 2) a neural network approach learning character-level sequences.³⁵ The first implementation of the lemmatiser shows 45.2% accuracy (i.e. the percentage of correctly generated lemmas) with unknown words and 71.6% with known words, while the neural network metrics are 64.9% and 99.2%, respectively; the neural network approach thus greatly outperforms the one based on edit distance.

6 Proposed methodological framework

In this section, the author will briefly point out how the interlinking of cognate verb forms is envisaged (see section 8), and how some of the resources described in section 5, together with a morphological analyser for Old Irish (section 7), will be employed to this end. The project's methodological framework is schematically represented in Figure 2.

Two morphological finite-state transducers (FSTs, see 7.1), located at the opposite ends of the chronological spectrum, play a pivotal role in the envisaged mapping of cognate historical (verb) forms. Both Old Irish and contemporary standardised Modern Irish reflect stable and normative phases in the language's history and are (relatively) well resourced. For the modern standard language, a morphological FST and a POS tagger are available (Uí Dhonnchadha and van Genabith 2006). As illustrated in Figure 2, standardisation methods are formulated relative to Old Irish and contemporary standardised Modern Irish. Advanced computational standardisation methods are already successfully being used for tagging *Corpas Stairiúil na Gaeilge* 1600–1926 (Uí Dhonnchadha et al. 2014) and the Bardic Poetry corpus, as discussed in 5.3 and 5.4.

³⁴ Minimum edit distance, an approximate matching technique widely used in Natural Language Processing, calculates how similar two strings are by calculating the minimum number of editing operations (insertion, deletion, substitution, transposition) needed to transform one string into another. In one of the most well-known variants, the Levenshtein distance, particular costs are assigned to each of these operations (Jurafsky and Martin 2009: 74).

³⁵ Available at: <https://github.com/ancatmara/early-irish-lemmatizer> [accessed 13 February 2019].

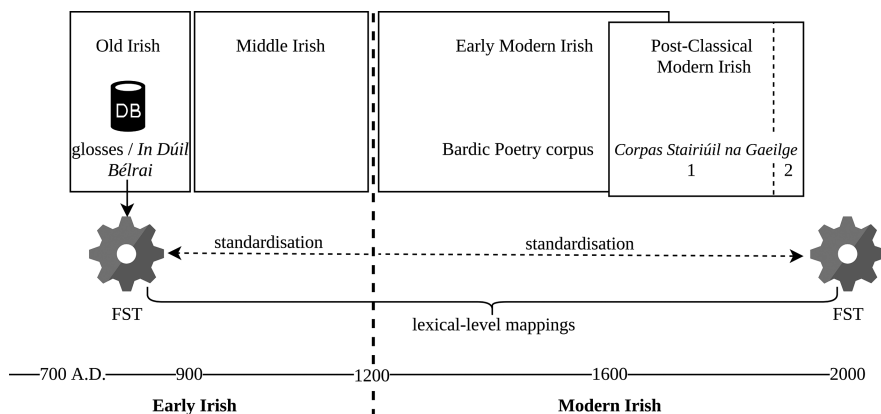


Figure 2: Framework for automatic identification and linking of cognate Irish (verb) forms. FST = finite-state transducer, see 7.1.

Seeing that much progress has already been made in tagging increasingly earlier historical Modern Irish forms, the present author is concentrating on the Early Irish side of the timeline. The following three tasks constitute the framework of the author’s project:

- Building a morphological finite-state transducer (FST) for Old Irish, which can assist in future work on a POS tagger for this period.
- Creating lexical-level mappings between the Old Irish morphological analyser and the available tagging tools for Modern Irish.
- Employing standardisation methods and potential analyser/tagger adaptation, in conjunction with digital corpora, to cover the language periods between Old and Modern Irish.

Task a. reflects the most novel approach in the author’s research project. The finite-state transducer can be augmented with manually parsed data from the databases of the Old Irish glosses (currently being streamlined in CorPH, see the introduction to this volume) and partial lemmatisation tables for verbs as present in *In Dúil Bétraí* (King, Lash, and Gabay 2006). It should be noted that the present work deals with morphological parsing rather than POS tagging. The task of automatic morphological analysis is to present all the grammatical possibilities on the word level. POS tagging is a subsequent task that aims at disambiguating between morphological parses (e.g. is Old Irish *ben* a verb or noun?) based on commonly the phrasal context. Due to the highly synthetic nature of the Old Irish verb, fine-grained morphological analysis is an essential prerequisite for POS

tagging as well as other subsequent tasks in the NLP pipeline for Old Irish. Morphological parsing of the Old Irish verb is the topic of the next section.

7 Automatic morphological analysis and generation of Old Irish verb forms

7.1 Finite-state morphology

Finite-state morphology is based on the mathematical notion of a finite-state automaton, a machine that recognises a particular set of symbol sequences (strings) as defined by a regular expression (a language for specifying text search strings, Jurafsky and Martin 2009: 17–18). Automata can be conceptualised as networks with transitions through a finite number of states. A finite-state transducer (FST) is an extension of this concept and contains two-level symbol correspondences for each path in the network. Figure 3 shows an FST with a mapping between a lexical-level and surface-level string representing present indicative third person singular absolute *léicid*. One of the advantageous features of this two-level formalism is that the relations encoded are inherently bidirectional: an FST can be used in recognition mode to analyse (parse) orthographical words in a text, but it may also be used to generate, say, a specified set of inflected forms (listing, for instance, complete paradigms of verbs with root *lēc*). Jurafsky and Martin (2009: 80) describe an FST as a “key algorithm for morphological parsing . . . and crucial technology throughout speech and language processing”.

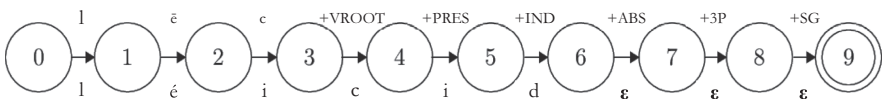


Figure 3: A finite-state transducer accepting, at final state 9, the surface string *léicid* (lower level) and lexical string *lēc+VROOT+PRES+IND+ABS+3P+SG* (upper level), constituting a two-level mapping. The epsilon (ϵ) denotes a so-called “empty transition”: a mapping where there is no accompanying symbol on the opposite level, i.e. when the upper and lower strings are of unequal length. The term analysis is used for upward mapping, which translates into morphological parsing. Downward mapping equals generation of (most commonly) orthographical strings.

Beesley and Karttunen (2003) is an important reference work accompanied by a toolkit called Xerox Finite State Tool (*xfst*).³⁶ This tool provides an extended set of regular expression operators, including the conditional rewrite rule format used in phonology, to intuitively model morphological and morphophonemic processes. The lexicon compiler (*lexc*) program (Beesley and Karttunen 2003: 203–278) facilitates and simplifies the creation of morphological grammars and can be used in conjunction with *xfst*. The finite-state toolkit *foma* (Hulden 2009) – which is freely available³⁷ and compatible with *xfst* – is used by the author to develop a morphological FST for Old Irish. The development of this tool is extensively documented in the author’s Ph.D. thesis (Fransen 2019), which, together with the code, is available online.³⁸

7.2 Implementation: Some highlights and challenges

7.2.1 Lexical and surface-level description

The two-level morphology paradigm is a fitting choice for the often daunting discrepancy between underlying and surface forms in Old Irish (verb) morphology, as detailed in section 3. The observation below is a suitable precursor to the computational challenges faced and choices made, as detailed in the remainder of this section:

The bewildering complexities . . . become transparent only when viewed from a diachronic position, and in order to understand allomorphic variation correctly it is essential to work with underlying forms and their often quite dissimilar surface representations
(Stifter 2009: 60)

The two-level formalism does not prescribe which linguistic entities are to be assigned to the upper level, although the latter is commonly reserved for synchronically motivated underlying morphemes. The (final) surface-level forms, however, should obviously match against the (commonly) orthographical forms as found in a text corpus.

Typically, the lexical level starts off with a lemma, and the surface level with a stem. In many languages, the latter bears an obvious relation to the former. This relation, however, is far from trivial in Old Irish, as pointed out by

³⁶ The accompanying website is <http://www.fsmbook.com> [accessed 7 February 2019].

³⁷ Available at: <https://fomafst.github.io/> [accessed 25 January 2019].

³⁸ Available at: <https://github.com/ThFransen84/OIfst> [accessed 10 March 2020].

Stifter (2009: 60). The full range of surface (inflected) forms often cannot be deduced from a single inflected form across a verb's paradigm, which means that the conventional citation form provided in lexicographical resources and grammatical descriptions of the language, the independent present indicative third person singular (e.g. eDIL),³⁹ is of little use when it comes to formulating a surface-level stem entry. This hurdle will be tackled in the next subsection.

The lexical level in the author's FST consists of diachronically motivated underlying morphemes rather than a citation form. In other words, a verb form's upper-level parse includes the abstract root shape of the verb and (with compound verbs) the underlying form of the preverb(s). In addition to linguistic motivations echoing Stifter (2009: 60), there are two practical reasons for the author's *modus operandi*. First, the use of diachronically motivated verb roots enables one to generate (surface) verb forms which have the same historical verb root. Secondly, employing "diachronic tags" allows for – and facilitates – interoperability with projects dealing with other historical Indo-European languages, or, indeed, Proto-Indo-European.⁴⁰

Example (20) illustrates the two-level encoding of the verb form *as-oilgi* 'opens' (L = lexical level, S = Surface level), based on the derivation provided in Stifter (2006: 364). The lexical-level tag *+PROCL_JUNCT* denotes the separation between the proclitic(s) and the stressed part of the verbal complex. The upper-level tag *W2a* can be added to enable extraction of verbs with this specific stem type. Consecutively numbering the preverbs is also expected to facilitate in-depth linguistic analysis; for example, it allows for a systematic investigation of the positional hierarchy of preverbs.

- (17) L *uss+PV1+PROCL_JUNCT+od+PV2+lēc+VROOT+W2a+PRES+IND+CONJ+3P+SG*
 S *as-oilgi*

Even though Old Irish can be treated as a normative phase within the medieval period, the language is far from orthographically stable. In the current

³⁹ It should also be noted that headwords in eDIL are not consistently provided in a form representative of Old Irish. An example is classical Old Irish deponent *molaitir* 'praises', which is represented by "generic" Early Irish *molaid* in the dictionary. (a new entry, albeit solely containing a reference to *molaid*, has been introduced in the revised 2019 version of the dictionary; s.v. *molaitir* or *dil.ie/50393*).

⁴⁰ See, for example, *Proto-Indo-European Lexicon*, a generative etymological dictionary of Indo-European languages, also implemented with the finite-state toolkit *foma*. Available at: <http://pielexicon.hum.helsinki.fi/> [accessed 7 February 2019].

implementation, the surface or lower level adheres, as closely as possible, to classical Old Irish grammar and orthography. Orthographical variation in Early Irish texts is expected to be successfully handled by one of Dereza’s (2018) lemmatiser implementations (see 5.4) used as a standardiser (see 7.4).

7.2.2 Monolithic stems

Section 3 has illustrated that a significant amount of allomorphic variation can be seen with verb stem formation, with syncope often causing truncation of the verb root, as in (17) above. This variation is challenging for a finite-state rule-based system, in which one typically starts off with a list of stems and affixation rules. Recall the morphotactics of the verbal complex (see example 7 in 3.1), repeated here as (18).

(18) (C) PV* (AUG) PV* VROOT E

If we blindly applied the morphological concatenations without regard to phonology, we would get, for instance, *ní-to-ro-lēc-* (C-PV1-AUG-VROOT), where the morphological derivation is quite far removed from the surface or orthographical form *ní-tarlaic* (see [viii.] in Table 2 on page 55, and 3.4). Employing the above concatenation schema to model Old Irish verb morphology was therefore not considered a feasible starting point – even when equipped with knowledge about the positional hierarchy of preverbs (McCone 1997: 89–90).

Allomorphic alternations are essentially a product of the morphology-phonology interface in Old Irish, as has been demonstrated in section 3. In other words, “unpredictable” stem formation is largely due to stress patterns, including syncope. Looking at Table 2 in section 3.1, the examples that do not show allomorphic stem variation are simple verb forms (i.)–(iii.) and the deuterotonic compound with one preverb in (v.), exactly those forms that have a stressed verb root. In all the other examples given in Table 2, where the root is unstressed ([iv.] and [vi.]–[viii.]), stem formation is less trivial, at least from a computational viewpoint and if operating with a set of clear-cut, synchronic rules.

The opposition of stressed versus unstressed verb root was found to be of major significance in formulating verb stems.⁴¹ An additional base form is required for any combination of a preverb or augment and an unstressed verb

⁴¹ Note that this distinction does not fully coincide with the traditional binary oppositions of simple versus compound and independent versus dependent.

root. A simple weak verb often requires an additional entry for the dependent augmented form, such as *reilic* in the case of *léicid*. A weak compound verb requires more stems; *do-léici*, for example, can be said to have four stems: (*do*) *léic*, (*do*)*reilic*, *teilic* and *tarlaic* (see Table 2). While stems of weak verbs are generally unmodified in the different tenses/moods, strong verbs may show root-internal stem modifications in each of the five tense/mood stems; in other words, the above-mentioned numbers for weak verbs should be multiplied by a factor of up to five for strong verbs.⁴²

Stem entries such as *reilic*, *teilic* and *tarlaic* are called “monolithic stems” in the author’s computational framework. These bases represent synchronically motivated multi-morpheme strings not trivially segmentable on the surface. Accordingly, they are not produced by diachronic phonological rules in the author’s FST rule framework, but keyed in as invariant stems in the *lexc* grammar. Monolithic stems subsequently enable the encoding of straightforward inflectional endings.⁴³ While initially born out of programming considerations, the concept of a monolithic stem is perhaps also theoretically insightful. When these bases have been determined and encoded for a large amount of verb lemmas, the minimum or average amount of stems necessary for operating with simple morphological rules can be calculated, which could be an interesting linguistic diagnostic for the level of complexity of the Old Irish verbal system.⁴⁴

The formulation of monolithic stems partly remedies the problem of synchronically opaque stem formation and alternation. However, dealing with syncope remains a complicated aspect in the implementation. For example, syncope may cause secondary palatalisation/non-palatalisation. Consider the independent (deuterotonic) and dependent (prototonic) present indicative first person plural forms

42 The concatenation of preverbal proclitics results in less allomorphic variation and can be modelled using separately defined surface morphemes (more on the programmatic treatment of proclitics versus (stressed) stems in 7.2.3).

43 But see the discussion on complications associated with syncope below.

44 I am thankful to Prof. David Stifter for bringing this additional insight to my attention. One could ask the question how the complexity in Old Irish verb stem formation compares to other languages. Such a cross-linguistic examination is, unfortunately, outside the scope of this paper. It is not unlikely, however, that the perceived complexity of the verbal system is at least in part due to the absence of a comprehensive synchronic description of Old Irish, and, consequently, a framework employing transparent morphological rules. So far, scholars of Old Irish have been mainly relying on historically oriented grammars such as GOI. A related issue, not unique to Old Irish, is the fact that the description of a historical language is typically based on a closed and often relatively small corpus; many forms across inflectional paradigms are not attested, which may impede a full synchronic description of the morphological rules at play.

of the strong verb *as·beir* (*ess-ber-*) ‘says’ in (19) and (20), respectively. The root vowel *e* in *ber* has been subject to syncope in (20). The *e* in ending *-em* in (20), as opposed to *-am* in (19),⁴⁵ marks subsequent secondary palatalisation of the consonant cluster *-pr-* (*/b^hr^h/*). Example (21) is the dependent prototonic equivalent of *as·beir*, with palatalisation of root-final *r* (*/eb^hər^h/*) to mark the personal ending. This form is not liable to syncope as it only consists of two syllables. The mechanisms behind stem and ending variation of this kind occur throughout paradigms of compound (and augmented simple) verb formations.

(19) *as·beram*

PV·say_{1PL.PRES}

‘we say’

(20) *ní·eprem*

NEG·say_{1PL.PRES}

‘we do not say’

(21) *ní·epir*

NEG·say_{3SG.PRES}

‘does not say’

In the current implementation, syncope is incorporated in the framework of regular expression rules; a conditional rewrite rule targets vowels in even-numbered syllables (but not final ones), which are liable to syncope. A monolithic stem such as *tarlaic* should therefore be encoded as *tarolaic* (even though this form never surfaces) to make sure vowel syncope is correctly applied in subsequent even-numbered syllables. Monolithic stems are therefore perhaps best described as semi-surface forms.

Unavoidably, “mechanical” treatment of syncope results in cases where the resulting consonant cluster violates the phonotactics of Old Irish. While this can be (and partly has been) counteracted by changing the conditional rewrite rule, irregularly applied syncope is very hard to cater for. For example, the augmented preterite third person plural surface form *reilciset* (*·reil^tciset*) generated by the FST does not match attested *·reil^tc^tset*,⁴⁶ with syncope of the vowel in the second as well as third syllable (the difference between the vowel sequence

⁴⁵ Apart from the different quality of the preceding consonant(s), both endings represent */əṽ/*.

⁴⁶ This form is cited and discussed, alongside other examples of compounds with root *lēc*, by Ó Cruaíoch (1999: 97–98).

e and *ei* in the first syllable is purely orthographical). Perhaps we should operate with the stem *reilc* instead of *reilic* to arrive at what would then be regular syncope of the vowel in the second syllable; however, the question in that case is how to derive forms without syncope, such as expected and attested⁴⁷ preterite third person singular *·reilic*. Intra-paradigmatic analogy further complicates a rule-based approach to syncope, as can be illustrated with dependent present third person plural passive *·epertar* (expected **·ep[†]retar*) of *as·beir* ‘says’ (*essber-*), modelled on the present third person singular passive *·eperr*.⁴⁸

The complexities relative to syncope and analogy (operating both within and across paradigms) raise the question whether rule-based stem-and-ending generation using monolithic stems is invariably more economical than manually encoding (“hard-coding”) an entire verb paradigm. Strong verbs such as *beirid* and *as·beir* are very frequent, and therefore more liable to irregularity and analogical processes. For many other verbs, the distinction between “regular” and “irregular” (or, perhaps better, “predictable” versus “unpredictable”) is not as clear-cut, which deters deciding *a priori* whether an automatic generation or manual encoding approach is most feasible. Establishing a good balance between automatic and manual methods (based on expert knowledge) is further complicated by the fact that no exhaustive list of Old Irish verbs or verb roots exists – let alone a comprehensive overview of stem classification and stem formation processes that could inform the formulation of monolithic stems.⁴⁹

In the author’s project, the focus is initially on weak verbs; compared to the group of strong verbs, weak verbs show transparent tense/mood stem formation

⁴⁷ eDIL s.v. *léicid* or dil.ie/29766.

⁴⁸ For an overview of the entire inflectional paradigm of *as·beir* see Strachan (1929: 68–71).

⁴⁹ Another problem is that some works deal with roots, and others with lemmas. Pedersen (1909–1913, vol. 2) lists 204 roots (based on the dedicated number of paragraphs, 650–854). However, the focus is on primary verbs (mainly verbs with Proto-Indo-European roots), which are mainly strong verbs. A more up-to-date work on primary verbs is Schumacher (2004), who lists 197 reconstructed Celtic verb roots, 166 of which are found in Irish verbs. However, this work excludes causatives. Le Mair (2011) discusses weak verbs in the Old Irish glosses, giving a total number of 365. McCone (1997) lists a good number of inflections in his *index verborum* but does not include stem class and mainly considered material from the Old Irish glosses. The online eDIL contains 4,127 verb headwords but does not systematically provide a stem classification. This number includes duplicates as some verbs have a separate Old and Middle Irish headword. Moreover, some (e)DIL headwords are more indicative of Middle Irish than Old Irish (e.g. *molaid*, rather than Old Irish deponent *molait^hir* ‘praises’). Rossiter (2004) applied McCone’s stem classification (the one adhered to in the present work) to verbs in DIL, but only dealt with compound verbs. The vocabulary section in Stifter (2006) is far from exhaustive but does systematically provide the stem class and roots for verbs.

by means of suffixation only (see 3.3); in other words, one does not have to cater for non-trivial stem-internal (non-concatenative) modifications based on an abstract root. However, as the above has shown, most verbs need more than one monolithic stem regardless if we want to cater for augmented simple verbs and compound verbs.

7.2.3 “Word” segmentation and separated dependencies

Morphological parsing operates on the word level, and words are defined as strings surrounded by space. A form such as *beirid*, with the ambiguous ending *-id*, will receive three grammatical analyses during morphological parsing, as it may occur in three grammatical contexts, illustrated in (22)–(24). The presence or absence of a conjunct particle (here negative or negative imperative), if separated by space, is a disambiguating feature in the subsequent task of building a POS tagger (not part of the present paper), which operates beyond the word level, even if merely typographical.

(22) *beirid*
 carry_{3SG.PRES}
 ‘carries’

(23) *ní beirid*
 NEG·carry_{2PL.PRES}
 ‘you do not carry’

(24) *(ná) beirid (. . .)!*
 (NEG·) carry_{2PL.IMPV}
 ‘(do not) carry (. . .)!’

Morphological boundary markers, including spaces, are absent in faithfully transcribed texts and diplomatic editions. More commonly, text editions (which might make their way into a digital corpus) are subject to editorial choices and policy, according to which typographical morpheme boundary markers might be employed. The current finite-state implementation anticipates instances of a potentially highly synthetic verbal complex written as one “word” (consecutive string), optionally with a mid-high dot (for the proclitic juncture) or hyphens; it accepts forms of the type *nondobmolorsa*, but also, for instance, *nondob·molorsa* (example (16) discussed in section 3.2) (a different yet interesting approach focused on pre-processing is provided in Doyle, McCrae, and Downey (2019),

who explore the possibilities of automatic tokenisation for Old Irish using a neural-network-based approach.). This choice in the implementation facilitates recognition but obviously also results in a vast amount of combinations to be considered of which only a limited amount are morphotactically valid. The restrictions are generally separated dependencies (co-occurrence of non-consecutive morphemes) and most of these have been successfully encoded. The generation of exclusively morphotactically valid forms prevents wrong parses due to ambiguity at the surface level (e.g. identical absolute and conjunct endings).

The interaction between monolithic stems (see 7.2.2) and separated dependencies is illustrated in Table 3, using the same verbs as in Table 2.⁵⁰ By their very nature, simplexes such as *léicid* cannot be preceded by a lexical preverb. Compound stems of the type *teilic* or *tarlaic* preceded by the proclitic augment *ro* or the proclitic preverb *do* are impossible as one or both of these elements are already incorporated in the (monolithic) verb stem.⁵¹ However, *ní* (and conjunct particles in general) can precede any stem except deuterotonic stems.

The author's *lexc* implementation contains separate lexicons for proclitics (preverbs and particles, optionally with infixed pronoun and relative marker) and verb stems with endings, which may occur typographically as strings separated by space, and are recognised as such. The lexicons can be optionally concatenated. In the author's implementation, separated dependencies are partly encoded with "flag diacritics" (Beesley and Karttunen 2003: 339–373), special regular expression symbols accompanying morphemes (*lexc* entries) that either allow or block paths in the network. Flag diacritics are not visible during analysis/generation, apply at run-time, and can, together with the blocked morphotactically illegal strings, be deleted from the network. For example, if *do* is marked as "preverb *do* seen", and prototonic *teilc* as "preverb disallowed", we will never get, for instance, **do-teilic*. A simple verb such as *marbaid* 'kills', also accompanied by "preverb disallowed", will equally never be prefixed with *do* (or any other preverb). Deuterotonic stems,

50 It should be noted that there might be overlap in monolithic stems across verb lemmas (e.g. *reilic*). In the current implementation, formulation of monolithic stems is on a per-verb (lemma) basis. An approach whereby monolithic stems are used for multiple lemmas, while not impossible, fails to make a distinction between, for instance, simple and compound verbs, which are subject to different constraints. "Recycling" monolithic compound stems might be of use, however, with verbs liable to preverb alternation (e.g. *in-fét*, *ad-fét* 'relates'), secondary composition (Stifter 2006: 254) and the employment of "dummy" preverbs in Middle Irish (McCone 1997: 194–197).

51 *ro-teilic* and *níro-teilic* for *do-reilic* and *ní-tarlaic*, respectively, reflect a Middle Irish development whereby *ro* gradually assumes the status of conjunct particle (see 3.1 and Breatnach 1994a: 279). The prefixation or infixation of proclitic *ro* with prototonic stems is blocked in the current version of the FST; systematic encoding of Middle Irish features (such as the relaxation of grammatical rules relating to proclitic *ro*) is envisaged as a subsequent adaptation stage of the FST.

Table 3: Schematic overview of separated dependencies with a selection of proclitics and monolithic stems in Old Irish, exemplified with *léicid* and *do-léici*.

Lemma	LEXICON 1				LEXICON 2		
	Proclitic (“prefix”)				Monolithic stem	Ending	
	Conj. part.	Pre-verb	Aug-ment	Optional infix rel. marker and/or pronoun class		Simple/ proto- tonic	Deutero- tonic
<i>léicid</i> (simplex)			<i>ro</i>	A or rel.(+C)	<i>léic-</i>		abs.
				A	<i>reilic-</i>		conj.
<i>do-léici</i> (com- pound)	<i>ní</i>			A	<i>teilic-</i>		
		<i>do</i>		A or rel.(+C)	<i>tarolaic-</i>	<i>léic-</i>	
					<i>reilic-</i>		

however, have a flag diacritic of the type “requires preverb *X*”, which only allows the specified preverb, and nothing else, not even a proclitic that is not a preverb. Prototonic/simple stems, on the other hand, allow anything *but* a proclitic preverb, and are (mostly correctly) preceded by proclitics that are not preverbs (e.g. *ní*).

Flag diacritics have proven to be convenient for the encoding of (essentially arbitrary) combinations of proclitic preverb and deuterotonic stem. The disadvantage of flag diacritics, from a programmatic point of view, is the fact that one needs to think carefully about separated dependencies *in advance* when laying down the morphological concatenation architecture. Consequently, “adding Flag Diacritics *post hoc* to an existing system can require non-trivial re-editing of your source files” (Beesley and Karttunen 2003: 340). A sometimes more convenient way of restricting the generation of ill-formed words is the use of upper-level filters (Beesley and Karttunen 2003: 247–255), i.e. specifying incompatible upper-level tags for an initially over-generating *lexc* grammar, and filtering all the illegally formed strings out of the network.

7.3 Preliminary test results

The morphological FST was tested on the Old Irish text *Táin Bó Fraích* [Cattle-raid of Fróech], using the digital version available on CELT,⁵² taken from the edition by Meid (1974). The FST was augmented by personal names occurring in the story, a limited set of function words, and the extremely frequent defective verb *ol* ‘said’. It turned out that 9.6% of word types (unique forms) were morphologically parsed, with an average, comparable score of 10% for four other Early Irish narrative texts edited by Greene (1955).⁵³ While the consistency of these scores is a promising result, the main goal of this exercise was to see how the FST would cope with weak verb inflection, which was concentrated on during implementation. It should be noted that weak verbs were found to be rather infrequent; in terms of tokens, W1 and W2a verbs constitute only 8.3% of the total amount of verb forms (excluding verbal nouns) in *Táin Bó Fraích*.⁵⁴

Out of the 50 W1 and W2a inflected forms (types) in *Táin Bó Fraích*, 36 (72%) were found to be correctly parsed. Most of the 14 non-recognised forms either deviate from a “canonical” spelling or show idiosyncratic features that are difficult to capture in general rules. Two verb forms in *Táin Bó Fraích* show grammatical variation that perhaps legitimises a rule; present subjunctive third person singular *forruma* (*fo-ruimi*, *fo-rumai* ‘puts’)⁵⁵ and preterite passive third person singular relative *arrálad*⁵⁶ (*ar-áili* ‘arranges’) show fluctuation in stem-final consonant quality, which is a feature of W2 verbs (McCone 1997: 27–28).

Spelling variation in and across texts such as the ones considered here are often seemingly trivial, as can be illustrated with verbs with stem *léic-*, which, first of all, may equally be spelled *léc-*. Another feature with the no apparent grammatical implication is the occurrence of the digraph *ll* in all three instances

⁵² Available at: <https://celt.ucc.ie//published/G301006/> [accessed 7 February 2019].

⁵³ Also available on CELT. The individual stories are *Fingal Rónáin* (available at: <https://celt.ucc.ie//published/G302011/>), *Orgain Denna Ríg* (available at: <https://celt.ucc.ie//published/G302012/>), *Esnada Tige Buchet* (available at: <https://celt.ucc.ie//published/G302013/>) and *Orgguin trí mac Diarmata Mic Cerbaill* (available at: <https://celt.ucc.ie//published/G100037/>) [all accessed 7 February 2019].

⁵⁴ Middle Irish simple weak formations from original compound verbs were excluded from this count, as the FST does not (yet) deal with Middle Irish, e.g. *fácbaid* and *oslaigid*, from Old Irish *fo-ácaib* ‘leaves’ and *as-oilgi* ‘opens’, respectively.

⁵⁵ The headword *fo-ruimi* is given in Meid (1974) and eDIL (s.v. *fo-ruimi*, *-fuirmi* or *dil.ie/24043*, under which attested third person singular present indicative *forrumai* is listed).

⁵⁶ The nasalising relative marker appears on the final consonant of the preverb (*-rr*), rather than on *ál-* (*nál-*), as expected. We would also expect the preverb to appear as *ara-*. Variation of this type would have prohibited successful analysis by the FST in the first place.

of sentence-initial present indicative third person singular *do-léici* in *Táin Bó Fraích*, for what is more commonly a single *l* in this context. Frequent alternations of this kind beg the question if – and to what degree – spelling variation should be encoded as a final module in the FST framework. This and related possibilities will be briefly discussed in the next subsection.

7.4 Standardisation for Early Irish

Beesley and Karttunen (2003: 287–293) recommend building different transducers for different tasks so as to make the parsing pipeline as modular and flexible as possible. Such a pipeline typically includes a “standard” FST, where the surface level represents normative grammatical or orthographical forms. A separate transducer could be devised which can be secondarily invoked when a form does not conform to a standard grammar or spelling. The latter may take care of variation in spelling, as with, for instance, *léicid* for “standard” *léicid*.

The lemmatisation tools being developed by Dereza (2018) are also expected to be of benefit for standardisation purposes. As stated in 5.4, Dereza (2018) has implemented an Early Irish Lemmatiser using two approaches: one method is based on approximate matching using string similarity, the other uses neural machine learning. The first implementation predicts a mapping between an unknown inflected form in a text to a known variant, based on a dictionary of *form: lemma* mappings originally retrieved from eDIL. The second, more start-of-the-art and better performing version employs the latter mappings to learn character sequences in order to produce lemmas it has never seen. The morphological FST for Old Irish, currently being developed by the author, will, in time, surpass the amount of inflected verb forms listed in the Dereza’s Lemmatiser dictionary. Moreover, the inflected forms generated by the FST adhere to a large extent to classical Old Irish inflection and spelling. By adding these canonical or “standard” Old Irish forms to the known mappings in the Lemmatiser’s dictionary, we not only increase the power of the Lemmatiser enormously, but we can also use this resource as a spelling standardiser, namely, by mapping an unknown variant in a text to a “standard” form from the FST, and retrieve the morphological parse of the latter. Lemmatisation and standardisation methods – in conjunction with the author’s morphological FST – have only been tested to a very limited extent.

The author has taken the liberty to use the term “standardisation” in his framework (see section 6), to show the similarity with the approach taken in the *Foclóir Stairiúil na Gaeilge* project. The terms “canonicalisation” or “normalisation”

are perhaps more fitting terms as an absolute standard did not exist in Early Irish, at least not in orthographical terms, not even in the otherwise reasonably homogeneous language of Old Irish.

8 Suggestions for linking cognate verb forms

The author's most fundamental envisaged approach is to create operability between his own Old Irish morphological FST and the one for Modern Irish (Uí Dhonnchadha and van Genabith 2006). Such an infrastructure could incorporate mappings between lexical-level tags, i.e. between Old Irish preverbs and verb roots and modern verb lemmas of the type *lēc+VROOT:lig+Verb+VTI* and *to+PV1+lēc+VROOT:teilg+Verb+VTI*. Additional tag mappings between inflectional categories could be devised, with the provision that there is often a discrepancy between Old and Modern Irish. For example, there is no straightforward modern grammatical category that matches the Old Irish augment. Although the modern past tense in many cases etymologically derives from a perfect construction with the preverbal particle *do* (for earlier *ro*, the resultative augment), it does not inherit either “perfectivity” or “perfect” as a grammatical feature.

Nonetheless, tag mappings of this kind facilitate juxtaposition of Old and Modern Irish paradigms, facilitating research into historical roots and grammatical developments such as innovatory processes in stem and ending formation. The historical connection between lemmas such as *lig* ‘let’ and *teilg* ‘cast’ is not present in the modern-language morphological FST. However, this connection can be established by means of *lēc+VROOT*, the “common denominator” for all verbs with this root in Old Irish, for which individual paradigms can be generated. The “modern” analysis additionally tells us that both *lig* and *teilg* can be used transitively (+*VTI*), a feature which is expected, in many cases, to transfer back to Old Irish.

Another “linking route” is through lemmatisation using *droichead* (Scannell 2018), a digitised version of the mappings between standardised contemporary Modern Irish lemmas from *Foclóir Gaeilge-Béarla* ‘Irish-English dictionary’ (Ó Dónaill 1977) and eDIL headwords, originally prepared by de Bhaldraithe (1981). Scannell (2018) added POS tags and used the imperative second person singular (as in Ó Dónaill 1977) as the modern lemma rather than the third person present indicative (matching the eDIL headword) in the original list. Since the FST for Modern Irish (Uí Dhonnchadha and van Genabith 2006) employs the lemmas in Ó Dónaill (1977) on the lexical/upper level, and *droichead* provides the corresponding eDIL headword, mappings between any modern standard inflected

form (as well as many pre-standard forms, see 5.4) and the Early Irish eDIL headword can be facilitated. Mappings between eDIL and (increasingly earlier) Modern Irish inflected forms or headwords would be of great benefit to scholars working on texts produced at various stages during the medieval period, who are currently confronted with a vast range of grammatical and orthographical variants while operating with limited lexicographical resources, especially for Early Modern Irish.

9 Synthesis and future work

The aim of the work is to link up lexical resources for the Early and the Modern Irish period. This chapter has identified a lack of digital linguistic resources for the historical Irish period, with a fragmentation and discontinuity in terms of lexicographical support, which makes the aim of the research, the interlinking of cognate verb forms, a far from straightforward process. This challenge is compounded by significant linguistic developments between Old and Modern Irish, mainly in the verbal system, and especially in Middle Irish.

The computational methodology proposed employs two finite-state transducers (FSTs) at the opposite end of the historical spectrum – Old Irish and contemporary standardised Irish – as these language stages represent normative and/or standardised varieties and are well resourced. Advanced methods are being employed in the context of the Royal Irish Academy’s *Foclóir Stairiúil na Gaeilge*, using a standardiser in conjunction with a modern-language POS tagger (based on an FST), greatly increasing recognition of increasingly earlier historical variants and connecting the latter with the modern lemma.

The current focus in the author’s project is on creating a morphological FST (and, subsequently, POS tagging tools) for Old Irish using the software *foma* (Hulden 2009). The FST is planned to be used in conjunction with a lemmatiser for Early Irish based on eDIL (Dereza 2018), which could be employed to predict canonical Old Irish inflected forms generated by the Old Irish morphological FST for orthographical variants in Early Irish texts, as such functioning as a standardiser.

The challenges relating to a rule-based FST include morphophonemically complex verb stem formation. Allomorphic stem variation and truncation of the verb root, especially prevalent with compound verbs, have been tackled computationally by devising multi-morpheme, non-derived units called “monolithic stems” in the author’s work; these bases consist of the verb root and, if present, preverbs and augment following the proclitic juncture. While the formulation of

monolithic stems is time- and knowledge-intensive, the resulting stem entries reduce complex, non-concatenative stem and ending formation to relatively straightforward morphological rules, which can be largely automated. Separated dependencies have been successfully handled with instruments of the finite-state paradigm.

Unpredictable inflectional patterns resulting from irregular syncope and analogy in inflectional patterns challenge a linguistically motivated, rule-based approach. A further issue is the absence of an exhaustive list of Old Irish verbs and information about stem type and stem formation. These conditions make it difficult to exactly establish the balance between automatic methods and manual efforts and expert knowledge needed. However, the concept of a monolithic stem strikes an interesting balance between “automatic” and “manual” and may well be a leap forward in establishing this balance. The incorporation of lexical resources such as the databases produced as part of the *Chronologicon Hibernicum* project and the lemmatised verb tables as part of *In Dúil Béirai* will likely speed up the development of the author’s FST.

Test results are promising but incorporation of more verbs and verb classes as well as catering for inflectional variation and non-standard forms is an important prerequisite in the context of establishing the feasibility of the implementational choices and further development of the FST for Old Irish.

Linking cognate verb forms across the entire historical period is very much future work. However, two methods have been proposed in this work. The first one involves mappings on the lexical level of the FSTs for Old and Modern Irish, facilitating the juxtaposition of entire historical paradigms based on Old Irish roots as well as systematic investigation of linguistic change. Alternatively, mappings between eDIL headwords and modern lemmas from Ó Dónaill (1977) can be established by integrating the tagger (and standardisation tools, Scannell 2009, 2017) for Modern Irish (Uí Dhonnchadha and van Genabith 2006) and the mappings as part of *droichead* (Scannell 2018).

Standardisation methods in conjunction with the Old Irish and Modern Irish morphological analysis/tagging tools will result in increasingly better coverage rates of intermediate variants. With the modern-language tagger “stretching back” and the one for Old Irish “reaching forward” we can metaphorically describe the adaptation process as a “two-pronged attack”. It should be stressed that, in catering for variation throughout the medieval period, adaptation processes are likely to move beyond the realm of orthography.

The substantial linguistic variation and change seen in the Middle Irish period in particular will be an interesting challenge for either the “old” and “modern” FST/tagger. Dereza’s (2018) Early Irish Lemmatiser will definitely have a role to play here, as it incorporates Middle Irish inflections given in eDIL; in

other words, we will (hopefully) arrive at the Early Irish eDIL headword. Adaptation of the Old Irish morphological FST to deal with “Middle Irishisms” is also a possibility. To properly deal with the verbal system of Middle Irish, a list of unverbated verb stems is necessary. Alternatively, or as a complementary approach, a list of prototonic compound stems from the Old Irish morphological FST can be extracted and combined with weak simple verb inflection. The latter will result in the generation of many non-existing or unattested “new” simple verb formations. Overgeneration, however, is not a problem from an analysis perspective (the grammatical analysis of an unattested form will never come up) and will enhance recognition scores.

Further possibilities include incorporating tags for language variety or linguistic features (such as for Middle Irish) on the lexical level of the (adapted version of the) Old Irish FST. Encoding this information will provide us with a way to augment morphological analysis with automatic textual dating. An application could be to establish the proportion of Middle Irish, as opposed to Old Irish, forms in an Early Irish text.

A more distant research prospect is the integration of POS taggers, databases, corpora and dictionaries into one lexical resource. Such a resource will hugely benefit scholars operating at the intersection of the Early and Modern Irish period, who now rely mainly on eDIL and Dinneen (1927), with no lexicographical facility that comprehensively spans the entire historical period. The author hopes to establish academic collaborations in the future to get a better grip on both the computational and linguistic challenges of his project.

Acknowledgement: This paper stems from research carried out during a Government of Ireland Postgraduate Scholarship (GOIPG/2017/1808) funded by the Irish Research Council. The author would also like to acknowledge the anonymous reviewer for helpful feedback and the editors for seeing this publication through.

Christopher Guy Yocum

4 Text clustering and methods in the *Book of Leinster*

Most investigations of the *Book of Leinster* (hereafter LL) have used close reading, historical, and philological techniques to identify authors within LL (for instance, see Mac Gearailt 1993; Bhreathnach 2002; Mac Gearailt 1997–1998; Ó Lochlainn 1941–1942; Ó Lochlainn 1943–1944; Mac Eoin 1982: 113–114). While this has met with some success, the methods used are by their nature idiosyncratic and prone to individual scholarly opinion. One notable exception is Derick Thomson’s paper *The Poetry of Niall MacMhuirich* which attempts to use statistical methods to attribute authorship of poems to Niall MacMhuirich (Thomson 1970). This paper will use methods of anonymous authorship attribution, which has been developed within the discipline of machine learning and statistical analysis to accomplish two goals: first, to demonstrate the means and methods of unsupervised machine learning techniques in early Irish literature and second, to discuss the implications of the application of this methodology to LL with a view towards a larger research project.

The paper will proceed in four stages. First, some scholarly literature concerning LL is reviewed. Second, the methods of data gathering, along with certain related problems, as well as the algorithms used in the analysis are commented upon. Third, the outcome of the analysis is summarised. Fourth, the paper concludes with an examination of the contribution the analysis makes to the debate surrounding the authorship of LL.

1 The context of LL

LL, next to *Lebor na hUidre* (hereafter, LU, Best and Bergin 1929), is one of the great monuments of Irish literary culture, which was written between 1151 to around 1201 (Schlüter 2010: 24; see also Duncan 2012: 45–56). Most of the manuscript is in the hand of Áed mac Crimthainn (usually cited as A); however, there are five other discernible hands: F, T1–4, M, U, and S (Duncan 2012; see also Schlüter 2010: 27). Overall, there are 164 texts, which have 189,472 words in total. The provenance of the manuscript is the subject of much debate. According to Schlüter, LL was the product of the monastery at Cloneagh in Leinster and it may have been moved to Núachongbáil for safety during the wars of the 12th century,

where it gained its medieval name. In addition, it was probably written for the Loígis and celebrates their ancestor the Ulsterman Conall Cernach, which accounts for much of the Ulster material in an otherwise Leinster book (Schlüter 2010: 30–35; see also Duncan 2012: 45–49).

A modern diplomatic edition of the entire manuscript was produced by Best, Bergin, and others (Best et al. 1954–1983). Moreover, many of the texts found in the manuscript have been published separately as critical editions (for instance, see O’Rahilly 1967). Currently, the entirety of the LL diplomatic edition is available in TEI (TEI Consortium 2009) XML format at CELT. The CELT version of the diplomatic edition of LL is the basis for all work considered in this chapter.

In an attempt to ascribe authorship to scribes of the versions of *Táin Bó Cúalnge*, *Cath Ruis na Ríg*, and *Mesca Ulad* found in LL, Mac Gearailt breaks up *Táin Bó Cúalnge* into “regions”, then performs some statistical analysis on the orthography and language of each region (Mac Gearailt 1993: 172–178). He concludes that scribes A and T intervened in the text and made their own contribution (Mac Gearailt 1993: 205). In a later study, Mac Gearailt (1997–1998: 405) attempts to date the *Táin Bó Cúalnge* by counting and cataloguing infixed pronoun usage and style and concludes that:

Finally, it may be noted that while non-historical infixed forms in the LL Tain can be assigned to the period when CRR-LL was composed, many others which conform fully to OIr. or Mid. Ir. rules are survivals from a much earlier stage of Recension II . . .

Mac Eoin (1982) discusses authorship in terms of unreliability of dating. In this he laments the situation but holds out hope that the separation of prose and poetry could assist in distinguishing between the original and any additions:

But how are we to judge the validity of attributions to poets who fall within the Middle Irish period? It is often assumed that ascriptions to Middle Irish poets in Middle Irish manuscripts like LU, LL, and Rawl. are reliable. Some would certainly seem trustworthy, but their reliability is not enhanced by ascriptions to Cormac mac Airt, Medb Lethderg, and Ailill Óluimm on the adjoining pages. (Mac Eoin 1982: 124)

Ó Lochlainn (1941–1942) and (1943–1944) attempts to use textual sources to secure attribution of authorship to poems traditionally ascribed to Mac Coise. This is done by using various features of Middle Irish to date the poem and the date of Mac Coise’s death in the early Irish annal to demonstrate that Mac Coise could not have written the poems under consideration. This position was re-evaluated by O’Leary (1999) who argued that there were three different people named Mac Coise whose poems can be securely ascribed: Airbertach mac Cosse Dobráin, Iorard mac Coisi, and mac Coisi (O’Leary 1999: 69–71).

As the examples above show, with few exceptions the main method of modern scholarly authorship attribution uses stylistic dating and dating with reference to annals or other sources of textual evidence to assign or at least question the authorship attributions made by the scribes of various early Irish manuscripts.

With this in mind, the next step is to begin with the mechanics of how to use unsupervised machine learning techniques on early Irish literature using LL as an example.

2 LL as a list of vectors

While the foregoing has set the scene for the current state of the scholarly debate, this section will explore the various methods of data gathering and analysis. This section is by necessity highly technical in nature and will require close attention to the means by which a written text can be transformed into a mathematical object.

Many of the texts which appear in LL, (e.g. *Lebor Gabála*), also appear in other Irish manuscripts, which would suggest that other manuscript versions should also be included in the analysis. While this is an area for future research, at the moment for the ease of analysis and modelling, other manuscript versions which are available on the CELT website have not been included. Thus, LL and its texts are the only ones under analysis in this paper.

2.1 Dividing LL into texts

There are several legitimate ways of viewing LL: as an indivisible complete work, or as the work of a group of scribes, or as a conglomeration of separate texts, or as the six-volume set as prepared by Best et al. (1954–1983). If LL is viewed in the first way, the analysis in this paper would not be possible as there would only be one text to analyse and the method proposed would not work. If LL is taken as the second, texts would need to be split by hand rather than by title. A slightly modified form of this analysis is attempted in the course of this chapter. If LL is viewed as the modern six-volume set, it would, much like the indivisible complete work, contain too few texts to analyse using the proposed method. Thus, for the purposes of this paper, LL will be viewed as the third type, a collection of

separate texts collected into one whole work by a group of scribes. This policy accords well with Duncan's (2012: 28) argument on the composition of LL:

To regard *Lebor na Nuchongbála* as a single manuscript does not allow for its complexity either physically, palaeographically, or textually, and gives the impression that it was written at the same time and place in a straight run.

The CELT XML edition is composed of TEI XML files which correspond to each published volume of Best et al. (1954–1983). Within these files each text is divided by a <div1> XML tag. Each of these were extracted by use of two XSLT documents in succession using the SAXON XSLT engine. From there, a further use of XPath is made to extract the textual information and place it in files given names based on the <div1> tags which eased further analysis. All computer code to accomplish this is available at https://github.com/cyocum/bol_project.

Not all texts as extracted by the method above were used in the subsequent analysis. The texts that were not included, are:

- *Haec sunt nomina virorum componentium lapides*
- *Lebor Gabála Érenn*
- *Togáil Troí*
- Prose *Dindsenchas*
- Metrical *Dindsenchas*
- All Genealogies
- All King Lists

As the reader will undoubtedly notice, most of the texts are king lists or other kinds of lists. These texts do not give enough of the kinds of information necessary to create a good representation of the texts that are of interest to this kind of analysis. In addition, as will be shown below, they may distort the outcome of this exercise. Moreover, *Lebor Gabála* has been excluded because it has a complex textual history of its own and may confuse the analysis (Scowcroft 1987: 81–89). The consequence of this policy is that texts which are identified by author within *Lebor Gabála* are not included. For instance, a text of Flann Mainistrech's poem *Éstid a eolchu cen ón* and Gilla Cóemáin's *Góedel Glas ó tat Goídil* and *Tigernmas mac Follaig aird* are not included in the analysis. As Peter Smith (2007: 27) states: "Both *Góedel Glas ó tat Goídil* and *Tigernmas mac Follaig aird* appear to have formed an intrinsic part of *Lebor Gabála* since their composition." The inclusion of the texts with identified authors but considered within *Lebor Gabála* would distort the analysis as the texts would be mixed with whomever transmitted the LL recension of *Lebor Gabála*. Similarly, *Togáil Troí* was excluded due to its own complicated textual history as described in Mac Gearailt (2016). The exclusion of the Prose and Metrical *Dindsenchas* is because these texts are similar in structure to *Lebor*

Gabála in that Prose and Metrical Dindsenchas contain component texts. Additionally, the prose and metrical versions also appear together in most versions of the text (Theuerkauf 2017: 49–50). Attempting to extract and examine these texts would overburden the methodology to the detriment of the illustrative purposes of this chapter. A coherent method of extracting these kinds of texts from their surrounding textual context both in a single manuscript and in many versions across manuscripts deserves a far more thorough examination than can be accomplished here. Having such a methodology would allow an unsupervised machine learning method to be applied while remaining faithful to their history and context. Additionally, there are three texts which were not included in the CELT XML of the *Book of Leinster* but were used in the analysis, namely: *Táin Bó Cúlinge*, *Fingal Rónáin*, and *Esnada Tige Buchet*. These were supplemented from additional CELT files.

2.2 From texts to vectors

There are two generally accepted forms of text tagging for anonymous author attribution (Juola 2008: 262–266). First is part-of-speech tagging (hereafter, POS). This form of analysis uses a set of tags which mark the text for parts of speech. Using a method called Maximum Entropy, other untagged texts of the same language can be POS tagged (Jaynes 1957a, 1957b). Middle Irish does not have a POS tagger at the moment. POS tagging without an automatic POS tagger is extremely time-consuming and would be impossible in this instance. Lash (2014a) has constructed a corpus of POS tagged texts in Old/Middle Irish, which could form the basis for a POS tagger in the future. Additionally, there is the new *Corpus Palaehibernicum* (CorPH) (see the introduction to this volume), which could also help in this regard. However, the accuracy of automatic POS tagging can cause errors in itself:

. . . especially for POS taggers, is the introduction of errors in the processing itself; a system that cannot distinguish between contraction apostrophes and closing single quotes or that can only tag with 95% accuracy will conflate entirely different syntactic constructs, muddying the inferential waters. (Juola 2008: 265)

The second type is function word tagging. Famously, function word tagging was used in identifying the authorship of the Federalist Papers (Mosteller and Wallace 1963). The Federalist Papers were a set of anonymously written essays to promote the ratification of the Constitution of the United States of America. Function word tagging is less time-consuming and a proven way of identifying anonymous authorship and was therefore the chosen method for this exercise.

Function words are words which have no lexical meaning in a sentence and serve only to structure the sentence grammatically. In English, this includes words like *of*, *and*, *a*, *an*, etc. In early Irish, this includes but is not limited to pronouns *sí* ‘she’, prepositional pronouns *duit* ‘to you’, conjunctions (7), and definite articles *in*, *int* ‘the’. Difficulties arise when infix pronouns are encountered in the various texts. Infix pronouns were left out of this analysis on the grounds that it would be difficult, but not impossible, to add them cleanly while not disturbing the creation of the document vector, more on which below.

In LL, there are 1,125 different categories of function words. The number of categories is large because there is no normalisation done during the counting of the words. This means variant spellings and initial mutations of words are left unnormalised. Thus *n-uile* ‘all’ is counted separately from plain *uile* ‘all’. The consequences of this choice will be explored later as it has bearing upon the mathematics involved. Of the 189,472 words, as mentioned above, there are a total of 56,513 function words which means that there are an average of 344.59 function words per text. As there are so many function words, to list them here would be impractical so the total raw frequency can be found online.¹

Once the tagging is finished, the $tf*idf$, which stands for term frequency times inverse document frequency, is calculated as shown below. Term frequency in the formula $tf*idf$ means that the frequency of function words in a document is a major factor in determining the author of the document, which according to Zhong and Ghosh (2005) gives the best results for this kind of analysis.

Let D be the set of all documents under consideration and N be the number of documents in the set. For LL, $N = 164$. The normalised frequency $tf(t, d)$ of a term t in a document d is computed thus:

$$tf(t, d) = \frac{f(t, d)}{\max\{f(w, d): w \in d\}}$$

In other words, the term frequency of a term in a document is the number of times that term appears in that document, denoted $f(t, d)$, divided by the maximum raw frequency of any term in that document, denoted $\max\{f(w, d): w \in d\}$. This includes non-functional terms (in other words, terms that have semantic meaning: nouns, verbs, etc.).

The *inverse document frequency*, idf , is then computed thus:

$$idf(t, D) = \log\left(\frac{N}{|\{d \in D: t \in d\}| + 1}\right)$$

¹ https://github.com/cyocum/bol_project/blob/master/raw_frequency.txt

In other words, the logarithm of the number of documents in the corpus divided by the number of documents where the term t appears. It is common to adjust for the fact that term t may not appear and thus one is added to it, which avoids divide by zero situations.

Finally, the tf^*idf is calculated thus:

$$tfidf(t, d, D) = tf(t, d) \times idf(t, D)$$

This is calculated for each possible function word in a text. The tf^*idf is well-defined for all words in the corpus but only function words are of interest here. Not all function words occur in all texts; if a term t does not occur in a document d , then $tf(t, d)$ is 0 and hence $tfidf(t, d, D)$ is also 0.

We pick an ordering, which, while random, must be fixed as discussed below, t_0, t_1, \dots, t_n of function words, and an ordering d_0, \dots, d_{N-1} of documents, and for each document d_j we form a *document vector*:

$$(w_{1,j}, w_{2,j}, \dots, w_{k,j})$$

where $w_{i,j} = tfidf(t_i, d_j, D)$. The document vector is a list of the tf^*idf values as defined above.

The fixed ordering of words allows comparisons across documents. For instance, if γ (*ocus* ‘and’) is first in the list, then the tf^*idf for γ would be the first component of any document vector. Putting this all together, LL gives rise to a list of vectors (or a list of lists) of tf^*idf values for each document. A list of vectors is called a matrix. It is an interesting feature of the LL that this matrix is sparse, as many of the entries in the vectors are 0.

2.3 From matrix to clusters: *k-medoids*

Once the matrix of the tf^*idf of each function word which appears in a particular text has been calculated, the entirety of LL is ready for the next stage in its transformation. There are many different means of taking the digitised corpus and determining the possible clusters. The most common of these and the one that will be used here is called *k-medoids*, which was first introduced in Kaufman and Rousseeuw (1987) (see also MacQueen 1967). Additionally, while there are numerous statistical packages available to complete the last leg of the journey from text to mathematical object, the technical computing programming language Julia was chosen to compute and graph the final results (Bezanson et al. 2012).

The *k-medoids* algorithm uses a *distance metric* to partition the matrix (Park and Jun 2009). In this case, the cosine distance is used (see Tan, Steinbach, and

Kumar 2005: 500 and Singhal 2001). The optimal partition is then found and all texts are placed in the optimal clusters based on their cosine distance from each other and the possible number of clusters. In other words, texts are placed together in one cluster when the algorithm determines that their vectors are close to each other. The output of the algorithm is called a *clustering solution*. The clustering solution must then be interpreted, which means that the results may demonstrate author attribution, genre, or scribal activity (Stamatatos 2009: 23). All these possibilities will be explored below.

One drawback of the *k-medoids* algorithm is that it does not estimate the number of clusters and thus, the number of clusters must be supplied. There is research into estimating the number of clusters; however, the research has not yet reached a point at which researchers are comfortable with the accuracy of the results (Maitra and Ramler 2010: 380; see also Solka 2008: 103).

What this means in practice is that the scholarship on early Irish literature drawn upon above plays a role in considering accuracy of the results. Results which approximate what other scholars have determined using traditional methods are taken as being more credible than those which do not. However, this does not mean that a slavish attitude toward either the computational results or the previous scholarship should be taken. If the results do not reflect what is expected, this could mean that continuing investigations are warranted to determine the exact reason for the differences and what these mean for the LL text clustering question.

2.4 Normalisation and Principle Component Analysis

Consider the matrix of LL: the frequency of each function word is often 0 and the number of possible function words is 1,125. In this context, each function word represents a dimension along which the vector resides. In other words, the vector space within which LL sits has 1,125 dimensions. When the *k-medoids* algorithm is applied to such a space, the number of dimensions contributes to the difficulty of finding optimal clusters.

This argues for the use of orthographic normalisation to reduce the number of potential function words and hence dimensions with the intention of reducing the size of the space and obtaining more optimal clusters. However, normalisation is not a simple operation in the context of early Irish orthography. First removing initial mutations would reduce the number of dimensions but not by much. After this was done, a more difficult task would present itself: which one of the different orthographic variations would be appropriate? In addition, if one from a particular linguistic period was chosen, would this force other texts

which may have been written earlier or later to act like the normalised version? For the purposes of this study, this is a problem that is bearable. The complications of normalisation are high and deserve their own discussion, which is outside the scope of the current examination.

In any case, a far more sophisticated and illuminating method than normalisation is known as Principal Component Analysis (Abdi and Williams 2010) which attempts to capture as much of the variance as possible in as few dimensions as possible. Additionally, Principal Component Analysis allows visualisation of *k-medoids*.

As one can see from the graph below (Figure 1), a majority of texts in LL fall in a fairly small range. This means that the variation between the texts is fairly small or the texts in LL were either picked carefully to be very similar or they were edited in the process of copying to be more similar to each other. The farthest outliers in Figure 1 are: on the x-axis, *Dindgnai Temrach* and, on the y-axis, *Cóica epscop dodeochatar dochum Moedoc Ferna do Bretnaib Cille Muine* and *Drochcomaithech robaí i n-ocus dosom*. The Principal Component Analysis ultimately shows that the texts have fairly low variance, which means that the use of function words tends to be uniform across the texts included in the analysis. One interpretation of this result is that the scribes could have chosen texts which used function words in a regular fashion. Another, more likely interpretation, and commonly understood by modern scholarship, is that the scribes were involved in not only copying texts but also changing those texts and by those changes, used function words in a consistent fashion.

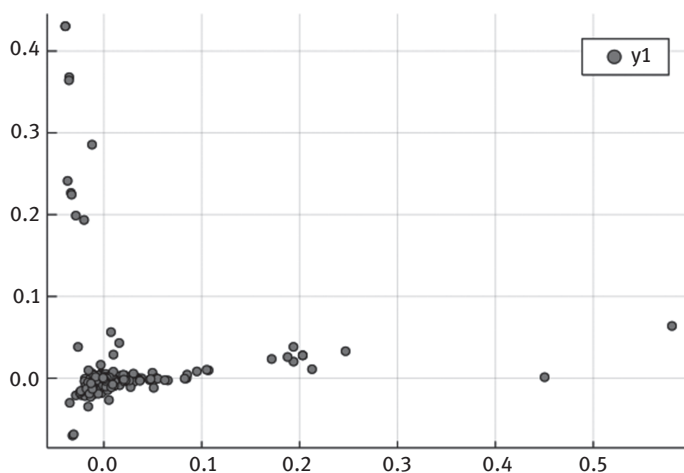


Figure 1: Principle Component Analysis.

3 Analysis

As shown above in the discussion of Principal Component Analysis, the variation between the texts is constrained. Without more texts and texts which contain more variation in the use of function words, this makes any clustering solution rather weak. However, pressing on with the final calculations will demonstrate how to interpret and analyse the results of machine learning techniques on early Irish material. Therefore, three differing classes of further analysis will be presented. The first uses known authors as the basis upon which to understand the clustering solutions provided. The second will investigate the scribal hands as the main basis to understand the clusters provided. The third will use genre identification as the main method to understand the clustering solution provided by the *k-medoids* algorithm.

3.1 Authors

Once all the calculations from above are complete, the texts are ready for the final step. In this section, two *k-medoid* analyses will be presented. As stated above, estimating the number of clusters is still an area of active research. Thus, attempting to estimate the number of clusters is a subjective process. A method is to use a set of texts which have a known author, then attempt to fit that set to the output of *k-medoids*. For illustrating the analytic techniques involved in unsupervised machine learning analysis, two known authors who may assist in evaluating the accuracy of any particular cluster solution are used: Flann Mainistrech and Gilla Cóemáin (Smith 2007), who both wrote historical poems which are included in LL and are fairly well known. There are other known authors in LL. For instance, Fothad na Canone, Ailill Ólomm, and Flan Fína but they will not be included in the analysis to keep the points being illustrated here clear. For instance, In the case of Flann Mainistrech, poems attributed to him in LL are listed below as presented by Póðör (1999):

- *Éstid a eolchu cen ón* [Listen, scholars, without flaw]; as stated above in 2.1, this is not included.
- *Ríg Temra dia tesbann tnú* [The kings of Tara, without envy]
- *In éol dib in senchus sen* [Do you know the old tradition . . . ?]
- *Mide maigen Chlainne Cuinn* [Mide, homestead of the descendants of Conn] (Smith 2001: 108–144)
- *Cia triallaid nech ainsnis* [Whoever attempts to tell the story] (Gwynn 1991 [1903–1935], 4: 100–107)

- *Cind cethri n-dini iar Frigrind* [At the end of four generations after Frigriu] (MacNeil 1913: 48–54)
- *Ascnam ní seól sadail* [Let us proceed – it is no easy undertaking] (MacNeill 1913: 48–58)
- *Aní do ronsat do chalma* [What Eogan’s race have done of valiant deeds] (MacNeill 1913: 59–70)
- *An gluind, a n-echta* [Their deeds, their death-dealings] (MacNeill 1913: 70–82)
- *Mugain ingen Chonchraid chain* [Mugain, daughter of righteous Concrad]
- *Sil Aeda Slaine na sleg* [The race of Aed Sláne of the spears] (MacNeill 1913: 92–99)

While not included in Póðör’s list, the following is also added:

- *Rig Themra toebaige iar tain* [The kings of many sided Tara, after that]
- *A Gillu gairm n ilgrada* [O lads of the names of great rank]

The list of poems ascribed to Gilla Cóemáin are listed below (Smith 2007: 25–7):

- *Hériu ard inish na rrig* [Lofty Ireland, island of the kings] (Smith 2007: 104–169)
- *At-tá sund forba fessa* [Herein is the apex of knowledge] (Smith 2007: 170–187)
- *Annálad anall uile* [All the annal-writing heretofore] (Smith 2007: 188–211)

With the above in mind, it is time to consider the k in k -medoids. The variable k represents the number of clusters to which a solution is found by the algorithm.² The clustering solution places all texts in LL into the number of clusters as signified by k . A cluster could mean an author so that k could equal the number of authors of LL. The possibility that clusters do not represent authors is explored below. As previously mentioned, there is no reliable way yet to estimate the number of clusters in a clustering solution. This leaves k as arbitrary, although not wholly so. In particular, the method chosen for this chapter is that k is adjusted iteratively until the texts begin to coalesce into clusters which look like the above list of texts. This began to happen when $k \rightarrow 15$. Using a technique called a *Silhouette*, which is described below in section 3.3, the choice is further refined until it was decided to use $k = 27$, which gave the best clustering solution and thus 27 authors of LL.³ As

² The full clustering solutions in CSV format for each value of k used in this paper are available at: https://github.com/cyocum/bol_project/tree/master/clustering_solutions.

³ https://github.com/cyocum/bol_project/blob/master/clustering_solutions/clustered-27.csv

mentioned above previously, known authors can constrain k to a number in which their texts begin to cluster together. As explained above, to test this hypothesis, we use known authors with known lists of texts in LL. First we will examine Flann Mainistrech and where the texts which are attributed to him fall within this list.

Table 1 is subset of the clustering solution for all the texts of LL for which Flann Mainistrech is the attributed author. The Cluster ID is the number of the cluster which is assigned by the algorithm. All texts assigned the same Cluster ID are within the same cluster. As the reader will notice, not all of Flann's texts are clustered together and a few are spread out among different clusters. Moreover, one will also notice that all the poems in cluster 10 are of the same genre, in particular, they are all historical poems which have many personal names in them, which reduces the overall frequency of function words. The reason for this may be that the style and meter are tightly constrained and thus, the function word use is similar across all the poems. This might mean that there are fewer function words and that the same ones are used frequently. However, authorship may still be preserved. These results may just be a consequence of the choice of known authors, who are both noted for writing mostly historical poems. A fuller analysis of this phenomenon will be discussed below.

Table 1: Flann Mainistrech where $k = 27$.

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
None	<i>Éstid a eolchu cen ón</i>	1	None	None
10	<i>Ríg Temra dia tesbann tnú</i>	3	U	U
10	<i>Mide magen clainne Cuind</i>	4	U	U
10	<i>Cind cethri ndíni iar Frigrind</i>	4	U	U
10	<i>Síl Aeda Sláne na sleg</i>	4	U	U
10	<i>Aní doronsat do chalmu clanna Eogain</i>	4	U	U
10	<i>Ascnam ní seol sadal</i>		U	U
3	<i>Angluind a n-echta a n-orgni batar infhir</i>	4	U	U
3	<i>Inn eól dúib in senchas sen</i>	3	U	U
6	<i>Mugain ingen Chonchraid chain</i>	3	U	U
6	<i>A Gillu gairm n ilgrada</i>	1	T	T2

Moving to Gilla Cóemáin, the situation is even more pronounced:

Table 2: Gilla Cóemáin where $k = 27$.

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
10	<i>Hériu ard inis na rríg</i>	3	A	A
11	<i>Attá sund forba fessa</i>	3	U	U
17	<i>Annalad anall uile</i>	3	U	U

None of Gilla Cóemáin’s historical works falls into the same clusters. This surprising result shows that our assumptions regarding ascription of the texts to Gilla Cóemáin may be incorrect in this instance. On the one hand, that *Hériu ard inis na rríg* clustered with historical texts by Flann Mainistrech does in cluster 10 may mean that there is enough information to cluster it with other historical poems. On the other hand, that *Analad anall uile* clusters with other historical poems in cluster 17 marginally supports an argument that the cluster algorithm is identifying genre rather than authorship. As for *Attá sund forba fessa* in cluster 11, this poem is in the same cluster as another didactic poem, *Sluindfet duib dagaisti in dana* (Thurneysen 1912: 73–77) and a poem about a quarrel between an old woman and a retainer of the king of Leinster, *A bairgen atá i ngábud* (Ua Nualláin 1904). However, this does not negate the interpretation of clustering solutions as authorship. Ultimately, it means that Gilla Cóemáin may not have written these texts and these texts are being placed with their anonymous authors who may have also written historical poems. In addition, the limitations of the present methodology may be interfering with the placements of the texts.

3.2 $k = 20$

The only change between this clustering solution and the previous one is that k was set to 20 rather than 27.⁴ All other parameters were kept the same for the sake of consistency and comparison. As would be expected, the change in the clustering solution is small. Most texts stay in the same clusters even if the cluster numbers have shifted. However, some texts like *Cia triallaid nech aisnis*, which was in the same cluster as those identified with Flann Mainistrech in the case of $k = 27$

⁴ https://github.com/cyocum/bol_project/blob/master/clustering_solutions/clustered-20.csv

where $k = 20$ (see Table 3 below), have moved. This means that the text was probably near the boundary between two clusters, and thus two authors, and was assigned to the correct author by *k-medoids* but when the number of clusters changed, the text was assigned to a new cluster based on the new boundaries calculated. This situation is why a quality check as described below using Silhouettes and previous scholarship on early Irish literature are consulted to detect these situations and determine the best clustering solution.

Table 3: Flann Mainistrech where $k = 20$.

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
None	<i>Éstid a eolchu cen ón</i>	1	None	None
2	<i>Aní doronsat do chalmu clanna Eogain</i>	4	U	U
9	<i>Mugain ingen Chonchraid Chain</i>	3	U	U
11	<i>Cia triallaid nech aisnis</i>	4	U	U
14	<i>A Gillu gairm n ilgrada</i>	1	T	T2
19	<i>Mide magen clainne Cuind</i>	4	U	U
19	<i>Cind cethri ndáni iar Frigrind</i>	4	U	U
19	<i>Síl Aeda Sláne na sleg</i>	4	U	U
19	<i>Ascnam ní seól sadail</i>	4	U	U
19	<i>Angluind a n-echta a n-orgni batar infhir</i>	4	U	U
19	<i>Inn eól dúib in senchas sen</i>	3	U	U
19	<i>Ríg Themra dia tesband tnú</i>	3	U	
19	<i>Rig Themra toebaige iar tain</i>	3	U	U

For Gilla Cóemáin, the clustering solution where $k = 20$ (see Table 4 below) is much the same. There is no overlap between his known historical poems and the cluster solution.

While the focus has been on Flann Mainistrech and Gilla Cóemáin, there are other indications that the clustering solution is registering authorship rather than style. If the analysis is done with $k = 5$, one will notice that Fothad na Canone, Ailill Ólomm, and Flan Fína's texts will cluster together. Thus, it

Table 4: Gilla Cóemáin where $k = 20$.

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
4	<i>Annalad anall uile</i>	3	U	U
15	<i>Attá sund forba fessa</i>	3	U	U
19	<i>Hériu ard inis na rríg</i>	3	A	A

seems given the amount of clustering in both Flann and other cases that this method gives clues as to the underlying authorship of various texts in LL.

3.3 Silhouettes

While the above may, at first glance, look like an open and shut case for the case in which the clustering and the scholarship coincide, there is more to the story than Flann and Gilla. Silhouettes are used to validate and interpret clustering solutions (Rousseeuw 1987). Silhouettes are a measure of how well each text resides in its cluster and thus the quality of the clustering solution.

The Silhouette histogram (Figures 2 and 3) is a graph which has an x-axis that is bounded between -1.0 and 1.0. The closer to 1.0 a text falls, the further away it is from its cluster. The closer to 0 a text falls, the closer the text is on the boundary between clusters. The closer to -1 a text falls, the closer the text is to the wrong cluster. The y-axis counts how many texts have the same value on the x-axis. Ideally, there should be more bars on the positive side of 0.0 on the x-axis, meaning the text is in the correct cluster, for a good clustering solution.

As one can see in both the case where $k = 20$ (Figure 2) and where $k = 27$ (Figure 3), a large number of texts fall on the border between two clusters, around zero on the x-axis on the graph. This means that they fall on the edges of clusters and many others fall in the negative area on the x-axis. In turn, this means that they are probably in the wrong cluster. This is true for most clustering solutions used in the analysis. Increasing the number of clusters (the size of k) should provide a solution for this but, as stated above, texts which are known to be composed by different authors start to be placed into the same clusters so given: the constraints of known authorship, the use of function words rather than POS tagging, since $k = 27$ has numerically more texts on the positive side of 0.0 on the x-axis than $k = 20$, $k = 27$ is the best clustering solution presently available. In fact, none of the clustering solutions are entirely satisfactory. Once an accurate POS tagger for Old and Middle Irish is created, the results from applying

k-medoids on the vectors created from the POS tagger can then be investigated and checked against the current analysis to see how well this method worked and how well, overall, these statistical methods can work with material like the Old and Middle Irish corpus.

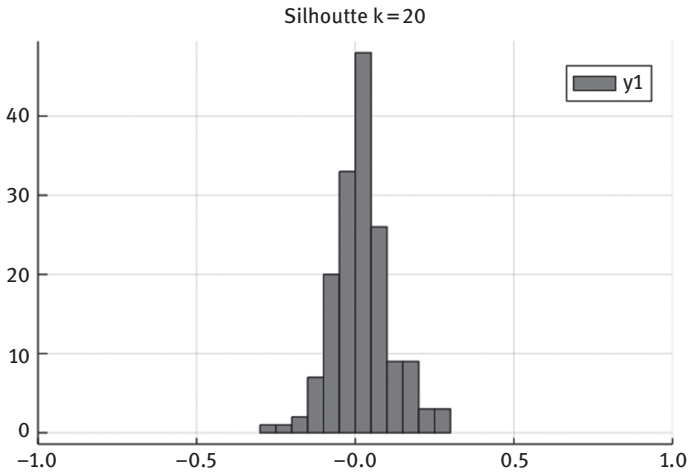


Figure 2: Silhouette where $k = 20$.

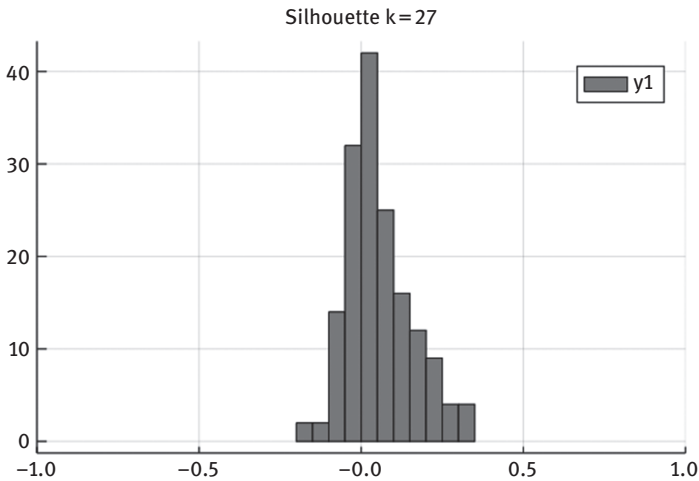


Figure 3: Silhouette where $k = 27$.

3.4 Scribes

As is well-known, scribes in medieval Ireland were not above making changes to their source material. This included updating the language in various ways: modernising the spelling, moving sentences, and, most importantly for this analysis, changing function words in various texts (Boyle and Hayden 2014: xxxvii–xlvi). Thus, while an argument can be made that *k-medoids* analysis can identify authors, there is an equally strong chance that the texts cluster because of this kind of scribal “editorial” activity regardless of the ascribed authorship.

Elizabeth Duncan has identified and catalogued the various folios to the nine scribes of LL (Duncan 2012). She does not, however, give a table of folios to texts. Duncan’s tables can be supplemented by the appendix supplied by Schlüter (2010: 226–243), which does give the scribal hand to text. Correlating the two tables gives a good indication of what each scribe wrote. In the case of overlapping scribal activity, the text is awarded to the scribe who contributed the most number of folios to a particular text. This is, of course, arbitrary and open to criticism but it suits the purposes of this analysis.

Using *k-medoids* analysis on the scribes of LL is much simpler than attempting the analysis for authors because much more is known about the hands involved in the manuscript and the fact that scribes sometimes claim, by, for instance adding a note above the beginning of a text, that a text was written by an incorrect author is not as much of a problem. The physical activity of the scribes means that there is less uncertainty surrounding which scribe wrote what section of text than attempting a more speculative analysis of authorship.

The results for the clustering solution where $k = 9$ (see Appendix) show that the scribes’ works admix freely and do not necessary cluster together as one would expect if scribes were strongly represented in the texts. This indicates that the scribes were not strongly influencing the texts themselves.

4 Genre

Defining different genres in early Irish literature is a problem which has exercised both the early Irish themselves and modern scholars. The problem is one of principles of categorisation. While modern scholars tend to organise the tales into a series of “cycles”, the medieval Irish organised the tales by the main action of the story (for discussion on medieval Irish literary theory, see Coileáin 1974; Backhaus 1990; Poppe 1999, 2008; Stam 2010: 66–68). Mac Cana (1980: 41–73) compiled the canonical list of medieval Irish genres in *The Learned Tales of*

Medieval Ireland from manuscript source, a summary of which is given below. There are twenty genres according to Mac Cana's list. In other words, does $k = 20$ cause the texts to fall into the same order of genres as proposed by Mac Cana (1980: 73–81)? If this is the case, then *k-medoids* could be clustering by genre as defined in the tale lists rather than by author (Juola and Baayen 2005).

- *Aided* 'death-tale, violent death'
- *Aithed* 'elopement'
- *Baile/buile* 'vision, frenzy'
- *Cath* 'battle'
- *Compert* 'conception, begetting, procreation'
- *Echtra(e)* 'expedition, journey (to the otherworld), adventure'
- *Fess/feis* 'feast'
- *Fis* 'vision'
- *Forfess/forbais* 'beleaguering, seige, night-watch'
- *Im(m)ram* 'sea-voyage'
- *Orcrain/orcun* 'murdering, ravaging'
- *Serc* 'love'
- *Slúagad/slógad* 'a hosting, a military expedition'
- *Táin* 'driving off, cattle raid'
- *Tochmarc* 'wooing, courting'
- *Togail* 'attack, destruction; attacking destroying'
- *Tomaidm* 'bursting forth of lake or river'
- *Úath* 'terror, horror' (although it seems to be a late genre) (Mac Cana 1980: 81)

A simple example will suffice for an answer. In the clustering solution for $k = 20$,⁵ four *Aided* tales (*Aided Cheltchair meic Uthechair*, *Aided Cuanach meic Ailchini*, *Aided Derb Forgaill*, *Aided Meidbe*) appear in cluster 11; however, this is misleading as the cluster is the largest with forty entries, which will not be shown here as it would be impractical (see rather the data referenced in footnote 5), and contains many other kinds of tales which are obviously not related. Two other *Aided* texts appear in cluster 2 but cluster 2 contains twenty-two other texts. While this is one example genre, it holds true for the other tale types. Especially since some of them cluster together with each other rather than in their own clusters. For instance, cluster 11 contains two *táin* texts (*Táin Bó Flidais* and *Táin Bó Fraích*) and two *cath* texts (*Cath Carn Chonail* and *Cath Maige Mucrima*), which are among many texts that are decidedly not *aided* texts. This argues strongly against the clustering solution reflecting the early Irish tale lists.

⁵ See https://github.com/cyocum/bol_project/blob/master/clustering_solutions/clustered-20.csv

5 Conclusion

The fundamental question asked at the beginning of this chapter was: how can unsupervised machine learning and statistical techniques be used in assessing authorship attribution in early Irish texts? An attempt was made to answer this question from a machine learning perspective using *tf*idf* and spherical *k-medoids* analysis to create a methodology from which a clustering solution was created using the Julia programming language. This methodology was then applied to LL with a number of texts removed for various reasons. Then a number of ways of understanding the clustering solutions was attempted: author, genre, and scribe. For authorship, the attributed texts in LL for Flann Mainistrech and Gilla Cóemáin were used to understand the clustering solutions $k=20$ and $k=27$. In this particular case, it seems that, while Flann Mainistrech's texts tend to cluster together, Gilla Cóemáin's do not, which would suggest that the texts ascribed to Gilla Cóemáin could possibly not been written by him but were written by others then subsequently ascribed to him. From the foregoing analysis, it would seem that there are 27 authors in LL. These clustering solutions were quality-assessed using Silhouettes, which showed some difficulties with the clustering solution. Two further clustering solutions were created: one that attempted to match known early Irish genres to clustered texts, and one that attempted to match clusters of texts to known scribal hands. In both cases, no correlation was found. This result suggests that scribal hands and genre are not useful when attempting to attribute authorship and should be avoided.

As identified in this paper, the methodology is strict; however, there is ample room for further improvement and research. For instance, the use of POS tagging rather than function words and the reduction of the dimensionality of the resulting matrix by using orthographic or other normalisation techniques should be investigated. This research may increase the accuracy and quality of the clustering solutions as identified using Silhouettes and bring more scholarly interest to this style of analysis. The method is not necessary conclusive, but rather suggestive and can help guide future research into the issue. In a broader sense, once the methodological difficulties are overcome, this methodology is possibly applicable to all of early Irish literature for which we have electronic versions.

Additionally, there is also room for more texts. For instance, *Lebor Gabála* was left out because of its complex textual and scholarly history. Moreover, the way in which the texts were separated in the electronic versions caused some known texts to be excluded. These exclusions directly influence the accuracy and reliability of the clustering solution. Further research into creating methods for extracting these texts in a coherent way, given that they are intimately

bound into the textual history of their enclosing text, and presenting them for analysis is also necessary.

Using unsupervised machine learning techniques and methods as presented here to answer questions of authorship in early Irish opens up new avenues of research and discovery, not just for LL, but for the whole of early Irish literature.

Appendix

The clustering solution where $k = 9$ is presented below in Table 5.

Table 5: $k = 9$.

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
1	<i>A Gillu gairm n ilgrada</i>	1	T	T1
1	<i>Cinti crábuid gnathaignthe scoile Sinchil</i>	6	A	A
1	<i>Cormac mac Culennain larfaiged nech acaib dam</i>	1	T	T1
1	<i>Diarmait mac Cerbaill maing thocheas rí clerchib ceil</i>	3	U	U
1	<i>Dublítir hua Uathgaile rédig dam a Dé do nim</i>	3	U	U
1	<i>Dubthach hua Lugair Andsu immarbáig ri Lagnib</i>	1	A	A
1	<i>Dubthach hua Lugair Crimthan clothrí cóicid Hérend</i>	1	A	A
1	<i>Feidlimid athair Echach</i>	1	A(?)	A
1	<i>Fland Fina in rigan ecanaid óg fjal</i>	3	U	A + U
1	<i>Fland Fina Ro ddet a hlnis find Fáil</i>	1	A	A
1	<i>Fland Mugain ingen Chonchraid chain</i>	3	U	U
1	<i>Fothad na Canone Cert cech ríg co rréil</i>	3	U	U
1	<i>Fothad na Canone Eclais Dé bí</i>	3	U	U
1	<i>Gilla Cóemáin Annalad anall uile</i>	3	U	U
1	<i>Gilla Mo Dutu Ádam óenathair na ndórene</i>	3	U	U

Table 5 (continued)

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
1	<i>Gilla in Chomded úa Cormaic a rí g rí chid reidig dam</i>	3	U	U
1	<i>Gilla na Naem Hua Duind Cuiced Lagen na lecht rí g</i>	1	A	A
1	<i>Mac Cosse of Ros Ailithir Rofessa i curp domuin dúir</i>	3	U	U
1	<i>Teist Chathail meic Finguine</i>	3	U	U
1	<i>Trí Fothaid Elgga cen chron</i>	3	U	U
2	<i>Angluind a n-echta a n-orgni batar infhir</i>	4	U	U
2	<i>Ascnam ni seol sadal</i>	4	U	U
2	<i>Brisleach Mór Maige Muirthemni</i>	2	U	U
2	<i>Cind cethri ndíni iar Frigrind</i>	4	U	U
2	<i>Cináed húa Artacáin Fíanna Batar i nEmain</i>	1	A	A
2	<i>Echta Lagen for Leth Cuind</i>	1	A	A
2	<i>Fland Mainistrech Rí g Themra dia tesband tnú</i>	3	U	U
2	<i>Fland Manistrech Inn eól dúib in senchas sen</i>	3	U	U
2	<i>Gilla Cóemáin Hériu ard inis na rrí g</i>	3	A	A
2	<i>Guidim Comdid cumachtach</i>	1	A	A
2	<i>Inis Dornglais ro gab Crimthann</i>	4	A	A
2	<i>Mael Muru Othna Can a mbunadas na nGaedel</i>	3	U	U
2	<i>Marb Cairpre Músc co n-áne</i>	1	T	T2
2	<i>Mide magen clainne Cuind</i>	4	U	U
2	<i>Mugdorn ingen Moga Duib Conan gilla Find</i>	4	A	A
2	<i>Ossin Ogum i llia lia úas lecht</i>	3	U	U
2	<i>Síl Aeda Sláne na sleg</i>	4	U	U
2	<i>Turim Tí gí Temrach</i>	1	T	T2
3	<i>A Maccáin ná cí</i>	3	U	U

Table 5 (continued)

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
3	<i>Ailill Ólomm beir mo sciath fri úath</i>	3	U	U
3	<i>Aní doronsat do chalmu clanna Eogain</i>	4	U	U
3	<i>Cinaed úa hArtacain a cholch thall for elaid úair</i>	3	U	U
3	<i>Dallán mac Móre Cerball Currig cáemLife</i>	1	A	A
3	<i>Dían airing</i>	1	T	T1
3	<i>Orthanach húa Cáellama A Chóicid chóem Chairpri chrúaid</i>	1	A	A
4	<i>A ben bennacht fort na raid</i>	5	F	F
4	<i>Aided Derb Forgaill</i>	2	U	U
4	<i>Aided Meidbe</i>	2	U	U
4	<i>Audacht Morainn</i>	5	S	S
4	<i>Clanna Ailella Uluim uill</i>	3	U	U
4	<i>Cathcharpat serda</i>	4	A	A
4	<i>De Gabail intSida</i>	5	F	F
4	<i>De dúlib feda na fored</i>	1	T	T1
4	<i>Dá brón flatha nime</i>	5	F	F
4	<i>Días macclerech</i>	5	F	F
4	<i>Echtra Laegaire meic Crimthainn</i>	5	F	F
4	<i>Fechtas aile do MLing is Toídin</i>	5	(F?)	F
4	<i>Mo Lling Luachra dalta do Maehóc Ferna</i>	5	(F?)	F
4	<i>Mo Lling Rochúala la nech légas libru</i>	3	U	U
4	<i>Nuallguba Emire</i>	2	U	U
4	<i>Senchán Torpéist Apair ri sil nEogain Móir</i>	3	U	U
4	<i>Slan seiss a Brigit co mbuaid</i>	1	A	A
4	<i>Túathal Techtmar ba rí Temrach</i>	1	A(?)	A
4	<i>Esnada Tige Buchet</i>	5	F	F

Table 5 (continued)

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
5	<i>Cu Chulaind atbert. De aduentu Christi</i>	2	U	U
5	<i>Maiccni Echach ard a ñgle</i>	3	U	U
6	<i>A Chormaic coisc do maicni</i>	3	U	U
6	<i>Audacht Moraind</i>	6	A	A
6	<i>Augaine Mår mac ríg Hérend</i>	1	A(?)	A
6	<i>Ciaran cecinit</i>	6	A	A
6	<i>Dialogue between Brendan and Moínenn</i>	6	A	A
6	<i>Fiacail Patric</i>	6	A	A
6	<i>Fochond Loingse Fergusa meic Roig</i>	5	F	F
6	<i>Macclerech do muntir Ferna móire</i>	5	(F?)	F
6	<i>Messe bad rí réil</i>	3	U	U
6	<i>Senbriathra Fithail</i>	6	A	A
6	<i>Tecosca Cormaic</i>	6	A	A
6	<i>Táin Bó Flidais</i>	5	F	F
7	<i>Brandub mac Echach</i>	1	A	A
7	<i>Cethri srotha déc éicsi</i>	1	T	T1
7	<i>Fithal 7 Cormac Níba mé linfes do neoch dara thráth</i>	3	U	U
7	<i>Orthanach húa Cáelláma Masu de chlaind Echdach aird</i>	1	A	T1 (over A)
7	<i>Secht mbémmed Brandub for Brega</i>	1	A	A
8	<i>Birth of Brendan</i>	6	A	A
8	<i>Broccán Craibdech Lecht Cormaic meic Culennáin</i>	1	A	A
8	<i>Bórama</i>	5	S	S
8	<i>Cellach Húa Rúanada sluindfet dúib dagaisti in dana</i>	1	T	T1

Table 5 (continued)

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
8	<i>Cethrur ar fichet nosfail</i>	6	A	A
8	<i>Clanna Falge Ruis in ríg</i>	1	A	A
8	<i>Colum Cille cecinit</i>	6	A	A
8	<i>Connachta cid dia tá in t-ainm</i>	1	T	T1
8	<i>Cuan Hua Lothchain Temair breg bale na fian</i>	1	A	A
8	<i>Epscop Ibar</i>	6	A	A
8	<i>Gilla Cóemáin attá sund forba fessa</i>	3	U	U
8	<i>Rig Themra toebaige iar tain</i>	3	U	U
8	<i>Sarbili anim Mo Ninni</i>	6	A	A
8	<i>Scrín Adomnáin</i>	6	A	A
8	<i>Trea ropo maith in ben</i>	6	A	A
8	<i>Táin Bó Cúalnge</i>	2	T (+ F)	T1 + F
8	<i>Étsecht Luin Garad</i>	6	A	A
8	<i>Úar in lathe do Lum Luine</i>	3	U	U
9	<i>A bairgen ataí i ngábud</i>	1	A	A
9	<i>Aided Cheltchair meic Uthechair</i>	2	U	U
9	<i>Aided Choncobuir</i>	2	U	U
9	<i>Aided Cuanach meic Ailchini</i>	5	F + A	F
9	<i>Aided Guill meic Carbada 7 Aided Gairb Glinne Rígi</i>	2	U	U
9	<i>Aigidecht Aithirne</i>	2	U	U
9	<i>Baí rí amra de Grécaib Salemón a ainm</i>	5	F	F
9	<i>Beochobra Con Culaind isind ló fúair bás</i>	2	F	U
9	<i>Buí siur Mo Lassi Lethglinni oc lé Gund i fail Mo Lasse</i>	5	F	F
9	<i>Caillech dorat a mac dósum do Mling</i>	5	(F?)	F
9	<i>Cath Carn Chonaill</i>	5	F	F

Table 5 (continued)

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
9	<i>Cath Maige Mucrima</i>	5	F (+ S)	F + S
9	<i>Cethrur macclerech</i>	5	F	F
9	<i>Cogad Gaedel re Gallaib</i>	5	T	T2
9	<i>Cormac mac Cuilennain cecinit</i>	6	T	T1
9	<i>Cummíne Fota mac Fiachnai di Eoganacht Chassil</i>	5	A	F
9	<i>Cóica epscop dodeochatar dochum Moedoc Ferna do Bretnaib Cille Muine</i>	5	F	F
9	<i>De Chophur in da Muccida</i>	5	F	F
9	<i>Dindgnai Temrach</i>	1	T	T1
9	<i>Do fallsigud Tána Bó Cualnge</i>	5	F	F
9	<i>Drochcomaithech ro baí i n-ocus dosom</i>	5	(F?)	F
9	<i>Epscop do Gaedelaib dochoid do Róim</i>	5	F	F
9	<i>Fechtas do Mling is Tóidin co n-acca Mael Doborchon</i>	5	(F?)	F
9	<i>Fechtas dósom oc ernaighi ina eclais</i>	5	(F?)	F
9	<i>Fland Manistrech Cia triallaid nech ainsis</i>	4	U	U
9	<i>Fothart for trebaib Con Corbb</i>	1	A(?)	A
9	<i>Gormlaith ingen Flain cia dír do chlérchib na cell</i>	1	A	A
9	<i>Gormlaith ingen Flain tanic ar debaid ó Cherball mac Murician</i>	1	A + T + 2	T2
9	<i>Iartaige na hingine colaige do Grécaib</i>	5	F	F
9	<i>Immacallam in dá Thúarad</i>	4	U	U
9	<i>Incipit Cath Ruis na Ríg</i>	4	A	A
9	<i>Incipit de maccaib Conaire</i>	5	S	S
9	<i>Longes Chonaill Chuiric</i>	5	F	F
9	<i>Longes mac nUsnig</i>	5	U (+ M)	U (+ M)
9	<i>Luid Feidilmid Rechtaid ó Themair do sáerchuard for Laighniu</i>	1	A + T + 2	T1 (over A)

Table 5 (continued)

Cluster ID	Title	Volume	Scribe (Schlüter)	Scribe (Duncan)
9	<i>Luid Mael Ruain Tamlachta fechtas dia airge</i>	5	F	F
9	<i>Medb Lethderg Macc Moga Corbb celas clú</i>	1	A	A
9	<i>Mesca Ulad</i>	5	M	M
9	<i>Na Trí Fothaid</i>	4	A	A
9	<i>Noenden Ulad 7 Emuín Macha</i>	2	U	U
9	<i>Orgain Dind Ríg</i>	5	F	F
9	<i>Rí iriscech ro boí do Grecaib</i>	5	F	F
9	<i>Scél Niall Frossach</i>	5	F + A	A
9	<i>Scél mucci Meic Da Thó</i>	2	U	U
9	<i>Scéla Chonchobuір</i>	2	F	F
9	<i>Senchas Ailíúin Chobthaig</i>	5	F	F
9	<i>Sloiged mār rucsat Gréic co Hebríb fechtas n-aile</i>	5	F	F
9	<i>Story of Athirne Ailgessach and Midir of Brí Leith</i>	2	U	U
9	<i>Story of Athirne and Amairgen son of Ecet Salach and Aigidecht Aithirne</i>	2	U	U
9	<i>Talland Etair</i>	2	U	U
9	<i>Tech Midchúarda</i>	1	T	T1
9	<i>Temaile fáid Miled Espáin</i>	4	A	A
9	<i>Tochmarc Ferbae</i>	5	U	U
9	<i>Trefocul</i>	1	T	T2
9	<i>Triar macclerech</i>	5	F	F
9	<i>Tréide Cétna Labratar Iarna Genemain</i>	2	U	U
9	<i>Trí Dé Donand</i>	1	T	T1
9	<i>Táin Bó Fraích</i>	5	F	F
9	<i>Túarastla Rosa Failgi</i>	1	A + T + 2	A
9	<i>Fingal Ronain</i>	5	F (+ A)	F
9	<i>Óenach Talten</i>	5	F + A	A

Acknowledgement: The author would like to thank Miles Gould, Miles Osborne, and the anonymous reviewers for their insightful and helpful comments on earlier drafts of this paper. All errors are, of course, the author's own.

**Part 2: Morphosyntactic variation and change
in medieval Celtic languages**

Liam Breatnach

5 The demonstrative pronouns in Old and Middle Irish

1 Introduction

The distinction between stressed and enclitic demonstratives is fundamental. In modern editions enclitic forms are usually printed with a preceding hyphen, this convention being most frequently observed in the case of the *notae augentes*, which are always enclitic, e.g. *baitsim-se* ‘I baptise’, *ad-cobra-som* ‘he desires’.¹ Unfortunately, however, the distinction is hardly ever observed in the case of the demonstratives, which, unlike the *notae augentes*, have both stressed and enclitic forms, which are not consistently differentiated in writing. As Old and Middle orthography does not, for example, regularly mark the long vowels in the stressed forms *só* and *sé* with a length-mark, or separate stressed *sin* from the preceding word by a space, as opposed to writing enclitic *-sin* as part of the preceding word,² the most reliable criterion left to us is metrical evidence. Accordingly, most of the examples which follow are taken from Old and Middle Irish verse.³ Much of what is established here regarding the earlier language applies also to Classical Modern Irish, the rules for which are set out by McManus in McManus (1994: 431–432, §§ 9.4–9.5), although the system there is further complicated by the approximation in form of the third singular masculine and third plural *nota augens*, *-s(e)an*, with the enclitic demonstrative, *-s(a)in*.⁴

1 The examples are taken from GOI § 403. See also Griffith (2008).

2 For word-division see GOI § 34. In Breatnach (2003) I showed that the stressed forms of the demonstrative meaning ‘this’, previously believed to be *so*, *se*, with short vowels, are in fact *só* and *sé*, with long vowels, and thus more differentiated from the enclitic forms than had been thought.

3 Where necessary, I silently introduce hyphens before enclitic forms, and separate stressed forms from what precedes.

4 They differ of course in the quality of the final *-n*; nevertheless the superficial resemblance of these two forms in Classical Modern Irish may have contributed to uncertainty as to whether a particular case of *sin* in an Old or Middle Irish text was enclitic or stressed. The replacement of the *-m* in *-som* by *-n* had begun in the late Middle Irish period; a few examples from the Book of Leinster are given in Breatnach (1994a: 264 § 10.2), where *dóib-sin* (LL line 8367), is a misinterpretation of the diplomatic edition; this should be read as *dóib sin*, with the stressed demonstrative.

Taking the enclitic demonstratives first, a few metrical examples will suffice to establish their prosodic status. They are all taken from the *Féilire Óengusso*, a text which can be closely dated to c.800 AD.⁵ The metre of this substantial text of 591 quatrains is *Rinnard*, with obligatory rhyme between the finals of the second and fourth lines, six syllables in every line, and each line ending in a disyllable. This last requirement guarantees the enclitic status of *-sin* in the example cited (rhyming parts are bolded):⁶

- (1) *Paiss Eutaicc la Fintan*
*Maeldub, mór a **ngáir-sin**,*
*caíngrían ocont **sléib-sin**,*
*dend Eoganacht **áin-sin**.*

‘The passion of Euty chius, with Fintan Maeldub—great is that shout!—the fair sun at that mountain, of those splendid Children of Eogan.’ (Stokes 1905: verse for 20 Oct.)

The enclitic status of *-sa* ‘this’ is confirmed not only by the disyllabic ending in the third and fourth lines, but also by the rhyme of *slóg-sa*, with the demonstrative, and *tróg-sa*, with the *nota augens* in:⁷

- (2) *Dom-rorbae domm théti,*
*ol am triamain **tróg-sa**,*
*iar timnaib ind **ríg-sa***
*rith ro ráith in **slóg-sa**.*

‘May it profit me for my comfort, for I am a wretched weary one, the course which this host has run according to the commandments of this King!’ (Stokes 1905: Prologue 25)

Similarly in the case of *-se*, the variant after a palatal consonant, we have the demonstrative in the third and fourth lines, and the *nota augens* in the second line in:

⁵ Cf. Breatnach (1996: 74–75).

⁶ As well as further instances in the verses for 20 Jun., 2 Aug., 12 Oct., 16 Oct. and at Epilogue 29.

⁷ Note also the *aicill* rhyme between *in tráth-sa* ‘at this time’, with demonstrative, and *ro gád-sa* ‘I have prayed’, with *nota augens* (Stokes 1905: Epilogue 411–412).

- (3) *Á Dé móir not guidiu,*
cluinte mo chneit trúaig-se,
ro beó iarsin báig-se
i mbithgnáis int slúaig-se!
 ‘O great God, I entreat Thee, hear my wretched sigh! May I be after this battle in the everlasting company of this host.’ (Stokes 1905: Epilogue 313)

The only variation is phonological, either contextual, viz. assimilation of the *s-* to the quality of the preceding consonant, or historical, viz. *-so* > *-sa*. Otherwise, the same form can be attached to a noun in any case or number preceded by the article.

2 Stressed demonstratives and their flexion in Old and Middle Irish

In these, there is a degree of variation in the forms. I will take the forms and range of use of *sin*, and of *só/sé*, where the referent is inanimate, and then instances of both of these with animate referents.

As for the demonstrative pronoun for ‘that’, apart from whether or not *in* precedes it, and the rare variant *sen*, there is no variation in its form for case inflexion, that is, it is always spelled (*in*) *sin*, and the final *-n* is palatalised, as shown by the rhyming examples below. All the Old Irish examples of the rare variant *sen* cited in DIL (S 231.8) are singular, and are from the tract on the Mass in the Stowe Missal, viz. one instance as the subject of the copula, *in sen* ‘that’ (Thes. 2: 253.16) and two instances after prepositions, *for sen* ‘thereafter’ (Thes. 2: 252.14) and *hō sen sūas* ‘from that upwards’ (Thes. 2: 255.7). In Middle Irish, on the other hand, the form varies between *sin*, *sein*, and *sain*, with some rare instances of *sen*; see Breatnach (1994a: 275 § 10.24).⁸

⁸ A reader adds an example from the Southampton Psalter: *linn in sen oc Hiurusalem* ‘that [is] a pool at Jerusalem’ (Ó Néill 2012: LXIII no. 12). The same gloss also has enclitic *-sen* in *esin lind-sen* ‘in that pool’.

2.1 Nominative

While verbal endings and copula forms allow for distinction between nominative singular and plural to be expressed, nevertheless, examples of the plural are very rare in Old Irish, only becoming well attested in Middle Irish. Most of the examples I have of the nominative plural have animate referents, for which see further below in section 4.⁹

An example as the singular subject of a (passive) verb is (here and below, all relevant demonstratives are glossed in bold; if only part of the example is glossed, the translation of the glossed part is underlined):¹⁰

- (4) *Gabthae tí chorcrae imund ríg*
lasa senad co ndimbríg
ba do genuch fo-cres sin,
 be_{3SG.PRET} for mockery_{DAT} PV·put_{3SG.PRET.PASS} **that_{NOM}**
níbu dúthracht a chumtaig.

‘A purple cloak was put about the King by the ignoble assembly; in mockery that was put about him, not from a desire to cover him.’ (Blathm. verse 52)

A Late Old Irish/Early Middle Irish instance with an intransitive verb is:

- (5) *Lethbairgen 7 ordu ēisc 7 lind in topair do-rat Dīa dam.*
dom-fic sin cach dīa
 PV·1SG·come_{3SG.PRES} **that_{NOM}** every_{GEN.SG.MASC} day_{GEN}
ol sē trīa thimthirecht aingel

‘“Half a loaf and a morsel of fish and the liquor of the well, God has given me. That comes to me every day”, said he, “by the service of angels.”’ (LU line 1846 [hand H]; author’s trans.)¹¹

⁹ A reader notes two examples with an inanimate referent, viz. *it he riaglóri in sin adchomlatair fri epacta* ‘those are the regulars which are added to epacts’ (Thes. 2: 17 [Carlsruhe Bede 32^a8]), and, in *Scél Mongáin: Batar hé sin a imthechta* ‘These were his adventures’ (White 2006: 76, 82).

¹⁰ Further examples with a passive verb are in Blathm. (verse 245), and, in prose, Binchy (1962: 60 § 12), and Gwynn (1914: 166.13).

¹¹ From *Imnram Curaig Maíle Dúin*; the translation deviates slightly from that in Oskamp (1970: 139).

A Middle Irish instance with a transitive verb is:

- (6) *ar nis fil do plaig nó dunibad for bith*
nachus bera sin for culu.
 NEG-3PL·bring_{3SG.PRES.SUBJ} **that**_{NOM} upon back_{ACC.PL}
 ‘For there is no plague or mortality on earth which that would not repel.’
 (Stokes 1891: 430–431 § 21)

An Old Irish example with the copula is:¹²

- (7) *Is ed trā in sin amnin*
 COP_{3SG.PRES} it then the_{NOM.SG} **that**_{NOM} indeed
nī mēte nī thormassid
ēcosc nī-āimin airm hī tā
tegdassa ad-chondarc-sa
 ‘That then is indeed—no doubt you can solve it (viz. the riddle)—the lovely form, where it is, of the house which I have seen.’ (Thes. 2: 292 verse 8; author’s trans.)¹³

2.2 Accusative

When the demonstrative (whether *sin* or *sé*) is the object of a verb, the verb may be accompanied by an infixed pronoun; see GOI (§ 478) for examples. This is only attested in the singular, and with a neuter pronoun. While masculine or feminine singular, as well as plural infixed pronouns with *só* and *sin* might theoretically be conceivable, none are attested. The instances with a neuter pronoun may, then, be a special case.

Examples of the accusative are:¹⁴

- (8) *Cenél do-rigni in sin*
 race_{NOM} PV·do_{AUG.3SG.PRET} the_{ACC.SG} **that**_{ACC}

¹² Cf. also the example in verse in Thes. 2: 290.14.

¹³ For the second line see Breatnach (1983). The poem from which the example is taken has been re-edited with translation and notes by Ahlqvist (2018).

¹⁴ A further metrical example is in Thes. 2: 294.13. A reader notes also examples of the asg. of *sin* as the object of comparison after equatives in the Old Irish Glosses, viz. *sic bith suthainidir sin ainm Solmon*, ‘even so lasting will be the name of Solomon’ (ML. 90^b10), and the instances in ML. 36^c21, 57^c12, 75^b7 and 131^d12 (all with *sin*).

*atá foraiḃ orbbadail;
is ainces ngalair cen tráig
a mbith cen flaith fo bithphláig.*

‘The race who did that suffer dispersal of heritage; their being without a kingdom under eternal plague is a sickly undiminishing misery.’ (Blathm. verse 117)

- (9) *Cenid relcset Iudei sin,
though-NEG-3SG_{NEUT}·suffer_{AUG.3PL.PRET} Jew_{NOM.PL} that_{ACC}
coíniud Críst dia sainmuintir,
nem cona airbrib—trén dú!—
ro-coínset uili Ísu.*

‘Though the Jews did not suffer that Christ should be mourned by his own people, Heaven (strong place!) and its hosts, all mourned Jesus.’ (Blathm. verse 128)

Examples from Middle Irish texts are:

- (10) *Ro airigestar Marggíni gilla Óchinn sein
AUG-observe_{3SG.PRET} Margine_{NOM} servant_{NOM} Óchinn_{GEN} that_{ACC}
‘Margíne, the servant of Óchinn observed that.’ (LL line 21149 [*Prose Dindsenchas*]; author’s trans.)*

A probable instance of the plural is:¹⁵

- (11) *do-ratsat sain uile n-óg
PV·give_{3PL.PRET} that_{ACC} all_{NOM.SG.NEUT} ^{NAS}complete_{NOM.SG.NEUT}
buidni Banba cen bithbrón
‘The hosts of Banba, free from enduring sorrow, gave all these completely
[as pledges].’ (LL line 25233; trans. MD 3: 11)¹⁶*

¹⁵ The preceding two verses consist of a list of what was given in pledge, e.g. *Eich claidib . . . gaí scēith* ‘horses, swords, spears, shields’, and it seems unlikely that these are being referred to collectively by a singular *sain*.

¹⁶ I take *uile n-óg* as an adverb (lit. ‘completely and entirely’), which probably goes back to an Old Irish neuter substantive *uile* followed by the nasalised adjective (lit. ‘the complete whole’). A further Middle Irish example is *De sin ro ort uile n-óg. / ind ēnlaith olc ecalmór* (LL line 20219), translated ‘Thereupon he slew them all entirely, the evil formidable fowls’, in MD 3: 259.

2.3 Prepositions

GOI (§ 480) notes that

any of the pronouns of §§ 478, 479 may be used after a conjugated preposition which is introduced by the copula. Examples: *is dó in so* ‘it is for this’ (Wb. 27d20); *is airi in sin* ‘it is therefore’ (Sg. 213a1); and often *is samlid in sin* or *sin* ‘it is like that’ [. . .] But where there is no periphrasis, such combinations are still rare – e.g. *fuiri sídi* (instead of *for súidi*) (Sg. 199a5), *ant sin* (for *i-sin*) (Ml. 356a1) – although later they become common.¹⁷

There are, then, two types:

- (a) Simple preposition + stressed *sin*, e.g. *ar sin*, *fri sin*, *iar sin*.
- (b) Prepositional pronoun + *sin*, e.g. *and sin*.

A metrical example of (a) is:

- (12) *ba* *sruith* *gruad* *ro ruid* *i sin*,
 COP_{3SG.PRET} venerable_{NOM.SG.NEUT} cheek_{NOM.} AUG-redden_{3SG.PRET} in **that**_{DAT}
fri náimtea co n-aithisib.
 ‘Venerable was the cheek that reddened thereat, facing insulting enemies.’ (Blathm. verse 122)¹⁸

The fact that there are no distinctive plural forms for either *sin* or *só* means that type (a) can only be singular.

While plural forms of type (b), such as *díb sin*, are well attested in Middle Irish, I have so far found no example in any early Old Irish text. The earliest example of this type I have is from the late Old Irish text *Immram Curaig Máele Dúin*: *Bá leis trá búaid cech cluchi díb sin* (LU line 1671 [hand M]), ‘He then was the winner in every one of those games’, although even this is in a manuscript of the Middle Irish period, and is not confirmed metrically.

In this type also, the demonstrative was stressed. All the metrically confirmed examples I have are Middle Irish. It bears repeating, however, that modern editions are inconsistent in distinguishing stressed forms from enclitics.

Examples with the singular are (the second part of the rhyming pair in 14 and 15 is in square brackets here and elsewhere):¹⁹

¹⁷ A reader notes also the prepositionless dative *sin* as the object of comparison after comparatives, e.g. *nand máa sin a bríg* ‘that it is of no more account than that’, Sg. 150^b1, sim. 150^b5.

¹⁸ I take the word-division in the edition (*ro-ruidi sin*) to be a slip.

¹⁹ Both the position of *sin*, etc. at the end of a line, and *rind ocus airdrind* rhyme confirm that in every one of these examples we have to do with a separate stressed word.

- (13) *dena lágrib tuctha and sin.*
 from-the_{DAT.PL} spear_{DAT.PL} bring_{3PL.PRET.PASS} in_{3SG.NEUT.DAT} **that**_{DAT}
dē atát Lagin for Lagnib
 ‘From the spears that were brought in that time, hence the Lagin are so called.’
 (LL lines 21057–21058 [*Prose Dindsenchas*]; author’s trans.)

- (14) a. *and sin* [: *mōrneim*] (LL line 26995 [*Metrical Dindsenchas*])
 b. *and sin* [: *mīlid*] (LL line 25712 [*Metrical Dindsenchas*])
 c. *and sain* [: *Alpain*] (LL line 27837 [*Metrical Dindsenchas*])

- (15) *coistid riss sein* [: *Taltein*]
 listen_{2PL.IMPV} to_{3SG.NEUT.ACC} **that**_{ACC}
 ‘Listen to that!’ (LL line 27775 [*Metrical Dindsenchas*])

Examples with the plural are:²⁰

- (16) a. *Nochor bruthi bir dīb sein.*
 NEG-AUG-cook_{3SG.PRET.PASS} spit_{NOM} from_{3PL} **that**_{DAT}
in tráth tucait ón tenid
 ‘Not a spit of those was cooked, when they were taken from the fire.’
 (LL line 29245; author’s trans.)²¹
 b. *ní dīb sein* [: *tromneim*]
 something from_{3PL} **that**_{DAT}
 ‘any one of those things’ (LL line 26846 [*Metrical Dindsenchas*])

2.4 Genitive

Unlike the other cases of the demonstrative pronoun, the genitive will have a noun preceding it, and this noun is usually preceded by a possessive pronoun, coreferential with the demonstrative, i.e. of the type *a fius sin* (Wb. 10^b27), ‘knowledge of that’, and the demonstrative is stressed.²² A careful distinction

²⁰ Further examples, with animate referents, are cited below.

²¹ The edition prints *dībsein*, as one word.

²² A reader notes instances in the Milan Glosses without a possessive in the case of the nominal prepositions *i ndiad* and *i ndigaid* ‘after’ (GOI §§ 858, 859), viz. *indiadsin*, 65^a12 (glossing *proinde*), 75^c8 (glossing *proinde*), 96^b13 (glossing *hinc*), *indiadsin*, 20^b4 (glossing *sed etiam*; sic manuscript, but emended to *innadiadsin*, Thes. 2: 29), and *indigaidsin*, 71^b11 (glossing *proinde*).

must then be made between two syntagms in which a demonstrative follows a noun, viz. the type *in lebor-sin* ‘that book’, with preceding article and enclitic adjectival demonstrative, and the type *a lebor sin* ‘the book of that one’, with preceding possessive pronoun and following stressed demonstrative pronoun.

The only metrically confirmed example I have so far from an Old Irish text is with *sé*, cited below in (43). Nevertheless, another indication that the demonstrative is stressed is that the form with *in* can be found in this position, as in the following passage from the Old Irish Glossing of *Senchas Már*:²³

- (17) *Somuine bech .i. lestur lulaice, ian oil lān di mellit 7 dā thartine dec,*
a lleth in sin ar lestar colpthaige,
 her^{GEN} half^{NOM} the^{GEN.SG} **that**^{GEN} for hive^{ACC} two.year.old.heifer^{GEN}
a trian ar lestar ndairte
 ‘Interest on bees, i.e. for a milch-cow hive, a pail of an *ól*-measure full of hydromel, and twelve small loaves; half of that for a two-year-old-heifer hive; a third of it for a yearling-heifer hive.’ (CIH 920.32; author’s trans.)²⁴

Similarly, the interposition of *ám dóib-sium* and *dam-sa* shows that *sain* is a separate stressed word in these two Middle Irish examples:

- (18) *Fail a mórabba ám dóib-sium sain*
 be^{3SG.PRES} its great.cause^{ACC} indeed to^{3PL=3PL} **that**^{GEN}
 ‘They have indeed good cause for that.’ (LL line 12066 [*Táin Bó Cúailnge*]; author’s trans.)
- (19) *Fail a mōrabba dam-sa sain*
 be^{3SG.PRES} its great.cause^{ACC} to^{1SG=1SG} **that**^{GEN}
 ‘I have good cause for that.’ (LL line 22899 [*Cath Ruis na Ríg*]; author’s trans.)

Metrically confirmed examples, however, are plentiful in Middle Irish; cf.:

- (20) *At-chūala co ngili gné.*
dā dam Dile derscaigthe.

²³ For this text see Breatnach (2005: 338–346).

²⁴ For the units of measurement used here see Kelly (1997: 578–580), and for *mellit* ‘hydromel’ (1997: 113).

Fe 7 Men fria nǵairm sein
 Fe_{NOM} and Men_{NOM} to-their^{NAS} calling_{ACC} **that**_{GEN}
ō fail ainm ar Maig Fēmin

‘I have heard of the two oxen of Dil, radiant of beauty, conspicuous; Fe and Men are they called, whence Mag Fēmin gets its name.’ (LL lines 27259–27262; trans. MD 3: 199)

- (21) *rap ferr lēo nā [a] silliuð sain*
 AUG-COP_{3SG.PRET} LEN better with_{3PL} than her looking_{NOM} **that**_{GEN}
a tabairt bēo fon talmain
 their putting_{NOM} alive_{NOM.SG} under-the_{ACC.SG.MASC} earth_{ACC}
 ‘Sooner than look upon her they had chosen to be buried under earth alive.’ (LL line 29927; trans. MD 4: 141)²⁵

- (22) *a mac samla sain [: gēnair]*
 his son_{NOM} likeness_{GEN} **that**_{GEN}
 ‘his match’ (SR line 5367; author’s trans.)

- (23) *Is fō samla sain sunna*
 COP_{3SG.PRES} under-her likeness_{ACC} **that**_{GEN} here
 ‘It was after her likeness in this place.’ (LL line 21473; author’s trans.)²⁶

3 The demonstrative *só*, *sé*, in Old and Middle Irish

As in the case of *sin*, both *só* and *sé* can be preceded by *in* (GOI § 478), but unlike *sin*, one of the variants is correlated with case inflexion. For some comments on the apparently free variation between *sé* and *só*, see Stifter (2015: 93–94).²⁷ The form *síu*, however, is found only in the dative, either with prepositions or as an independent dative.

²⁵ I supply in brackets the *a* found in two other copies (cf. MD 4: 141).

²⁶ The diplomatic edition prints *samlasain*, while MD 1: 10.61 reads *fon samla-sin*, with the article rather than the possessive and enclitic *-sin*, in spite of the internal rhyme with *calma* in the following line.

²⁷ On *sé* in the poems of Blathmac, see also Uhlich (2018: 64–67).

3.1 Nominative

Examples of *in sé*, *sé* and *só*, respectively, as the subject of a (passive) verb are:

- (24) *Is cian do-raimngred in sē*
 COP_{3SG.PRES} long_{NOM.SG.NEUT} PV·prophesy_{AUG.3SG.PRET.PASS} the_{NOM.SG} this_{NOM}
no mbíthe int áugaire.
 ‘Long has this been prophesied: that the shepherd would be struck down.’
 (Blathm. verse 127)

- (25) *ro-comallnad uile sē*
 AUG·fulfill_{3SG.PRET.PASS} all_{NOM.SG.NEUT} **this**_{NOM}
inge mod a thuidechtae.
 ‘All this has been fulfilled save the act of his [second] coming.’
 (Blathm. verse 233)

Another example is found in the late Old Irish *Immram Curaig Máele Dúin*:

- (26) *In dún-ni fo-rrácbad sō*
 INT-COP_{3SG.PRES}· for_{1PL=1PL} PV·leave_{AUG.3SG.PRET.PASS} **this**_{NOM}
ol Máel Dúin frisín cat
 ‘“Is it for us that this was left?”, said Máel Dúin to the cat.’ (LU line 1714
 [hand M]; author’s trans.)

Early examples with the copula are:²⁸

- (27) *Ní réid la céill mbuirp*
 NEG-COP_{3SG.PRES} easy_{NOM.SG.NEUT} with sense_{ACC} ^{NAS}uncouth_{ACC.SG.FEM}
in sē.
 the_{NOM.SG} **this**_{NOM}
 ‘This is not easy to the uncouth intelligence.’ (Blathm. verse 159)²⁹

²⁸ A reader notes further examples in copula sentences in the OIr Glosses, viz. Ml. 24^c4, 61^b7, 70^c6, 104^a4, 122^c9, 130^a16; Sg. 203^a16 (all with *it hé in sō*); Ml. 86^c3, 115^c1 (both with *it hé in sē*); Sg. 4^b12 (with *it hé sē*); Carlsruhe Bede 32^c8 (Thes. 2: 19, with *it . . . in sō*); Sg. 104^b1, 148^b12 (both *in sō*, with zero copula).

²⁹ Further examples with *in sé* are in Blathm. verses 187 and 237.

- (28) *reic Críst, ba drohcundrad sē.*
 selling_{NOM} Christ_{GEN} COP_{3SG.PRET} bad.contract_{NOM.SG} **this**_{NOM}
 ‘selling Christ!—an evil bargain this’ (Blathm. verse 108)³⁰

- (29) *Níbu for talam a dú;
 anní as fíriu
 iss ed a ngein cruchae in sō*
 COP_{3SG.PRES} it the_{NOM.SG.NEUT} ^{NAS}birth_{NOM} cross_{GEN} the_{NOM.SG} **this**_{NOM}
ro buí re ndíliu.
 ‘The earth is not the proper place for him: rather is this the being destined
for the cross who has been before the Flood.’ (IrGospThomas verse. 33)³¹

Another example is found in the late Old Irish *Immram Curaig Máele Dúin*:

- (30) *immafoacht dó cía mulend sō*
 PV-3SG_{MASC}·ask_{3SG.PRET} to_{3SG.MASC} what mill_{NOM} **this**_{NOM}
 ‘He asked him “what mill is this?”’ (LU line 1757 [hand M]; author’s trans.)

Two examples of the plural are:

- (31) *Derb batar é gnímae sē*
 certain_{NOM.SG.NEUT} COP_{3PL.PRET} they deed_{NOM.PL} **this**_{NOM.PL}
do maic máir maiss, a Maire.
 ‘It is certain that these were the deeds of your great beautiful son, Mary.’
 (Blathm. verse 41)

- (32) *IT ē in sō danō freptai inna*
 COP_{3PL.PRES} they the_{NOM.PL} **this**_{NOM.PL} also remedy_{NOM.PL} the_{GEN.SG.FEM.}
santi . . .
 avarice_{GEN}
 ‘These again are the remedies against avarice . . .’, (Gwynn 1914: 154–155, § 1e)

³⁰ Further examples with *sé* are in Blathm. verses 20, 140 and 208.

³¹ A further example with *in só* is in Gwynn (1914: 166.7).

3.2 Accusative

Examples of the accusative as object of a verb are:

- (33) *In fer ad-chuäid in sē*
 the_{NOM.SG.MASC} man_{NOM.SG} PV·^{LEN}relate_{AUG.3SG.PRET} the_{ACC.SG} **this**_{ACC}
is oen a thecht torise.
 ‘The one who has related this is one of his faithful messengers.’ (Blathm. verse 225)

- (34) ‘*Már huath*’, *ol in tuath,*
‘do mac do-gní sē;
 your son_{NOM} PV·do_{3SG.PRES} **this**_{ACC}
nícon cualamar co sō
nach macán am-nē.
 “A great terror”, said the people, “is your son who does this thing; until now we never heard of any such little boy.” (IrGospThomas verse 18)³²

- (35) *as-ber nícon dergēnus in sō nó a*
 PV·say_{3SG.PRES} NEG·do_{AUG.1SG.PRET} the_{ACC.SG} **this**_{ACC} or the_{ACC.SG.NEUT}
n-í aill.
 one other_{ACC.SG.NEUT}
 ‘who says: “I did not do this or that.”’ (Gwynn 1914: 160–161 § 23)

- (36) Late Old Irish
is airi do-gnú-sa sō. . .
 COP_{3SG.PRES} for_{3SG.NEUT.ACC} PV·do_{1SG.PRES=1SG} **this**_{ACC}
 ‘the reason I do this is. . .’ (LU line 1930, hand H [*Immram Curaig Máele Dúin*]; author’s trans.)

3.3 Prepositions

The situation regarding the demonstrative meaning ‘this’ is somewhat different to that of *sin* (above in subsection 2.3). The second type, prepositional pronoun +

³² A further example with *sé* is in IrGospThomas verse 44.

stressed demonstrative, seems to be rare, and the only example I have from an Old Irish text is in the Old Irish Glossing of *Senchas Már*:

- (37) *Is airi sō*
 COP_{3SG.PRES} **for**_{3SG.NEUT.ACC} **this**_{ACC}
nī tiagat dāla huīne i n-aile acht fri dīthim. . .
 ‘It is for this reason that matters proper to distraint with a stay of one day do not merge with those proper to distraint with a stay of two days, except in the case of delay in pound.’ (CIH 885.5; author’s trans.)³³

Even in the plural, forms such as *dīb sō* (CIH 1662.36, 1701.20 [sic leg.]) ‘of these’ do not appear to be attested in manuscripts of the Middle Irish period.³⁴

As for the first type, viz. simple preposition + stressed demonstrative, a distinction is made between accusative *sé*, as in *ar sé*, *fri sé*, etc., and dative *sú*, as in *iar sú*, *de sú*. An early rhyming example of the accusative form is *co sē* [: *gnē*] (LU lines 4576–4577) ‘up to this’ (*Táin Bó Cúailnge*).³⁵ As for the dative form, the long diphthong *íu* is confirmed by *úaitne* ‘consonance’ with *céo* and *mbéo* in:³⁶

- (38) *Rom-śnádat de śú*
 AUG-1SG·^{LEN}protect_{3PL.PRES.SUBJ} **from**·^{LEN}**this**_{DAT}
ar demnaib na céo,
céili Maic ind Ríg
a tírib na mbéo.
 ‘From here may they protect me against the fog-surrounded demons, these companions of the King’s Son from the lands of the living.’ (Murphy 1956: 26–27 verse 16)³⁷

³³ For the legal procedure in question here see Kelly (1988: 177–179).

³⁴ That is, MSS written before 1200. The majority of the examples with a prepositional pronoun given in DIL (S 307.5–17) are from Early Modern Irish texts.

³⁵ Further examples are given in Breatnach (2003: 138). Contrast *co sō* in verse 18 of the *Irish Gospel of Thomas* cited just above.

³⁶ Cf. also *sú* ‘here’, without a preceding preposition, making *úaitne* with *dó* : *fó* (LL lines 4816–4819), in the poem *Fothairt for trebaib Con Corb*, as well as the spelling *de sú* (LU line 1731 [hand M]), ‘from this side’.

³⁷ In the citation I have removed the hyphen in Murphy’s *de-śú*, to emphasise that *sú* is not enclitic.

A further variation is that *sund* can be used in place of *síu* in the dative singular, as in the following selection of examples from *Senchas Már*:

- (39) *Is i sund*
 COP_{3SG.PRES} **in this**_{DAT}
con-árrachta in dá recht.
 ‘It is in this that the two laws have been bound together.’ (Breatnach 2017a: 32–33 § 30)
- (40) *Is for sund*
 COP_{3SG.PRES} **on this**_{DAT}
ro suidigthea bechbretha la Féniu
 ‘It is on this that bee-judgments have been established in Irish law.’
 (Charles-Edwards and Kelly 1983: 88–89 § 55)
- (41) *Is for sund*
 COP_{3SG.PRES} **on this**_{DAT}
ro suidiged coibnius uisci t[h]airidne la Féniu.
 ‘It is on the foregoing [rules] that the kinship of conducted water has been established in Irish law’ (Binchy 1955: 72–73 § 15).
- (42) *conid-n-oiscfe di sunn.*
 PV-3SG_{MASC}·^{NAS}alter_{3SG.FUT} **from this**_{DAT}
 ‘he who shall alter it from this’ (Binchy 1966: 46–47 § 37).

This type became rare in Middle Irish; thus, for example, the only instance I have from the extensive body of verse that comprises the metrical *Dindsenchas* is *ō sun immach* (LL line 26633 [MD 3: 152.4]).

3.4 Genitive

Although examples have not been easy to come by, the following Old Irish instance has, as in the case of *sin*, a possessive pronoun coreferential with

the demonstrative. Its position at the end of a line, and the *rind ocus airdrind* rhyme confirm that we have to do with a separate stressed word *sé*:

- (43) *Is ed a etarcnae sē*
 COP_{3SG.PRES} it its significance_{NOM} **this**_{GEN}
mac ron-ucais, a Maire,
bid flaith cen tosach—cain n-ell!—
ocus flaith cen nach forcenn.

‘This is what this signifies: the son you have borne, Mary, will be lord without beginning (fair time!) and lord without any end.’ (Blathm. verse 190)

3.5 Stressed *séo*

While enclitic *-seo* is attested from the Milan Glosses,³⁸ the stressed form *séo* is not attested in the Old Irish glosses. Some Middle Irish examples are:

- (44) *Cend Guill sēo at-chí im lāim*
 head_{NOM} Goll_{GEN} **this**_{NOM} PV.^{LEN}see_{2SG.PRES} in-my hand_{DAT}
a Lāech
 oh Lāeg_{VOC}

‘This is the head of Goll which you see in my hand, o Lāeg.’ (LL line 12726; author’s trans.)

- (45) *Rop hé sēo Druim nElgga n-oll*
 AUG=COP_{3SG.PRET} he **this**_{NOM} Druim^{NAS}Elgga^{NAS}great_{NOM.SG.NEUT}

‘This hill was known as great Druim Elga.’ (LL line 27297; trans. MD 4: 337)

- (46) *conid de sēo bias Uisnech*
 so.that-COP_{3SG.PRES} from **this**_{DAT} be_{3SG.FUT.REL} Uisnech_{NOM}

‘and hence shall Uisnech be named’ (LL line 27637; trans. MD 2: 45.44)

The evidence surveyed thus far indicates that the demonstratives *sé* / *só* and *sin* were usually singular in Old Irish, and accordingly that plural forms would be expressed by means of the deictic particle *í*, preceded by the article and followed by the demonstratives *sin* and *síu*, on which see further below.

³⁸ See GOI (§ 475) and Schrijver (1997b: 18).

4 Demonstratives with animate referents

According to Pedersen (1909–1913, 2: 186) the demonstratives *sin* and *sé*, etc., were only used with inanimate reference in Old Irish: “Die substantivischen Gruppen *in-so* (in Ml. auch *in-se . . .*) und *in-sin* haben nur neutrale Bedeutung (“dies”, “jenes”) [The substantive groups *in-so* (in Ml. also *in-se . . .*) and *in-sin* only have a neuter meaning (*dies* ‘this’ (neuter), *jenes* ‘that’ (neuter))].” While examples with inanimate referents are plentiful in Old Irish, there are nevertheless some instances with animate referents, although they are not very common. All those I have noted are as the subject of copula:

- (47) *conid ē epscop in sin*
 so.that-COP_{3SG.PRES} he bishop_{NOM} the_{NOM.SG} **that**_{NOM}
citaru oirtned la Laighniu.
 ‘so that he is the bishop who has been first consecrated in Leinster’ (Thes. 2: 241.15)

- (48) *Is é remibí bóairechaib in sin*
 COP_{3SG.PRES} he PV·BE_{3SG.PRES.HAB} bóaire_{DAT.PL} the_{NOM.SG} **that**_{NOM}
 ‘that is one who takes precedence over other *bóaires*’ (Binchy 1941: 10.248)³⁹

- (49) *Sích in suí Sacharias:*
 ‘*Amrae mac in sō;*
 wonderful_{NOM.SG.MASC} boy_{NOM} the_{NOM.SG} this_{NOM}
ma for-cantae bed amrae
fri sodain da-nō.
 ‘Said the sage Zacharias: “This is a wonderful boy; were he to be taught he would be more wonderful still.”’ (IrGospThomas verse 22)

The referent can also be plural, as in:

- (50) *It ē mnā in sō*
 COP_{3PL.PRES} they woman_{NOM.PL} the_{NOM.PL} **this**_{NOM.PL}
nā dlegut lōg n-eneach

³⁹ From *Críth Gablach*; further examples, all with *in sin*, are at lines 280, 350, 448, 459, 475 and 593.

‘These are women who are not entitled to honour-price.’ (CIH 538.19 [Senchas Már]; author’s trans.)⁴⁰

4.1 Animate uses of *sin*

In Middle Irish, however, examples are much easier to come by, and in what follows I separate the examples of *sin* from those of *sé, só*.

4.1.1 Nominative singular *sin* as subject of copula

(51) ‘Can don mnaí?’ ar cāch.

‘Māthair Branduib in sin’, ar Āedān.

mother_{NOM} Brandub_{GEN} the_{NOM.SG} **that**_{NOM}

‘“Whence is the woman?”, said all. “That is Brandub’s mother”, said Aedán.’

(Meyer 1899: 135, 137 § 9)

(52) mac sin Bressail Bēlaich bind.

son_{NOM} **that**_{NOM} Bresal_{GEN} Belach_{GEN.SG.MASC} melodious_{GEN.SG.MASC}

‘The latter was the son of melodious Bresal Bēlach.’ (O’Brien 1952: 161, 167 verse 12c)

(53) Dúalderg ingen Mairge Móir,

ben sein Smucailli meic Smóil,

wife_{NOM} **that**_{NOM} Smucaille_{GEN} son_{GEN} Smól_{GEN}

‘Dúalderg, daughter of Marg the Great, she was the wife of Smucaille, son of Smól.’ (LL line 28898; trans. Ó Murchadha 2009: 23 verse 70)

4.1.2 *sin* as object of a transitive verb

(54) Fergus Lethderg ro slait sein [: anmain]

Fergus_{NOM} side.red_{NOM.SG.MASC} AUG·spoil_{3SG.PRET} **that**_{ACC}

‘It was Fergus Red-side that spoiled her.’ (LL line 29020; trans. MD 4: 251)

⁴⁰ Similarly, CIH 43.10.

- (55) *gabaid sin ol sē 7 berid a chend dē*
 take_{2PL.IMPV} **that**_{ACC} says he and bring_{2PL.IMPV} his_{LEN} head_{ACC} from_{3SG.MASC}
 “Take hold of that person,” said he, “and remove his head from him.”
 (Atkinson 1887: 643 [RIA MS 23 P 16 (Leabhar Breac) folio 9^a9]; author’s trans.)

4.1.3 Nominative plural *sin* as subject of copula

- (56) *Deich meic sin do Chathair chrūaid*
 ten son_{NOM.PL} **that**_{NOM.PL} of_{LEN} Cathair_{DAT} **stern**_{DAT.SG.MASC}
 ‘Those are the ten sons of stern Cathair’ (LL line 26022; trans. MD 4: 285)

4.1.4 Nominative plural *sin* as subject of a passive verb

- (57) *Na torothair danō techtait dā chorp i n-óenacomol*
deligfitir sin tall isind eséirgi
 separate_{3PL.FUT.PASS} **that**_{NOM.PL} beyond in-the_{DAT.SG.FEM} resurrection_{DAT}
 ‘The monsters also, that have two bodies in one union, they will be separated beyond in the Resurrection.’ (LU line 2562 [hand H] [*Scéla na Eséirgi*]; trans. Stokes 1904: 239).

- (58) *cethri sessir garga a ngluind*
ro marbtha sin la Dreacuinn
 AUG·kill_{3PL.PRET.PASS} **that**_{NOM.PL} by Dreco_{ACC}
 ‘Four times six—fierce their deeds! these were slain by Dreco.’ (LL line 30502; trans. MD 4: 15)

- (59) *ro slechta na sechtaib sain*
 AUG·slaughter_{3PL.PRET.PASS} in-their seven_{DAT.PL} **that**_{NOM.PL}
 ‘They were slain in their sevens.’ (LL line 26094; trans. MD 3: 99)

- (60) *is dīa réir ra sēolta sain.*
 COP_{3SG.PRES} to-their will_{DAT} AUG·send_{3PL.PRET.PASS} **that**_{NOM.PL}
gēill na Eurpa co Crūachain
 ‘in express submission to them have been sent hostages from all Europe to Cruachu’ (LL lines 20690–20691; trans. MD 3: 348)

4.1.5 Nominative plural *sin* as subject of an intransitive verb

- (61) *Is aire condeochatar sin i*
 COP_{3SG.PRES} for_{3SG.NEUT.ACC} that·come_{3PL.PRET} **that**_{NOM.PL} in
comdāil Con Culaind
 meeting_{ACC.SG} Cú_{GEN} Culann_{GEN}
 ‘The reason they came to encounter Cú Chulainn was . . .’ (LL line 8796
 [*Táin Bó Cúailnge*]; author’s trans.)

- (62) *Ro scāchatar sin uile*
 AUG·depart_{3PL.PRET} **that**_{NOM.PL} all_{NOM.PL.MASC}
nocho mair díb oenduine ‘All those have departed; not a single one of
 them remains.’ (Meyer 1912: 218 verse 23)

4.1.6 Plural *sin* after a prepositional pronoun

- (63) *Cid ūadib sain no gairthe.*
 even from_{3PL} **that**_{DAT.PL} PV·call_{3SG.PRET.PASS}
eter slūagaib sāmaigthe
 ‘Even from them it was called among leaguered hosts.’ (LL lines
 25401–25402; trans. MD 3: 23)⁴¹

- (64) *rí díb sin [: Femin]*
 king_{NOM} of_{3PL} **that**_{DAT.PL}
 ‘a king of those’ (LL line 29807 [*Metrical Dindsenchas*])

- (65) *cid atāi dóib sin bēus*
 what PV·be_{2SG.PRES} to_{3PL} **that**_{DAT.PL} still
 ‘Why are you still angry with them?’ (LL line 8367; author’s trans.)

4.2 Animate uses of *sé/só*

I give next plural forms of *sé*, *só*; the singular forms are included in the final section of this paper.

⁴¹ The internal rhyme *ūadib* : *slūagaib* establishes that *sain* is a separate word.

4.2.1 Nominative plural *só* as subject of a transitive verb

- (66) *ar cech n-omgním gníset sō*
 for every_{ACC.SG.MASC} ^{NAS}cruel.deed_{ACC} do_{3PL.PRET} **this**_{NOM.PL}
sníset a comlín chucco

‘For every cruel deed they did, they [the Tuatha Dé] inflicted the like number upon them.’ (LL line 25163; trans. MD 3: 5)

4.2.2 Nominative plural *sé/só* as subject of copula

- (67) *it íat in sō rīg na*
 COP_{3PL.PRES} they the_{NOM.PL} **this**_{NOM.PL} king_{NOM.PL} the_{GEN.PL.NEUT}
cóiced bātar acond feis-sin
 province_{GEN.PL} be_{3PL.PRET.REL} at-the_{DAT.SG.FEM} ^{LEN}feast_{DAT}=DIST

‘These are the provincial kings who were at that feast.’ (LL line 37651 [Bórama]; author’s trans.)

- (68) *Cōic rīg cōicat sáethra[i]ch sē*
 five king_{NOM.PL} fifty_{GEN} laborious_{NOM.PL.MASC} **this**_{NOM.PL}

do lāechraid na Crístaide

‘Five and fifty kings—laborious these!—of the warriorhood of Christendom’ (LL line 25209; trans. MD 3: 9)

4.2.3 Genitive plural *só*

Two examples of the genitive plural in Middle Irish commentary on *Senchas Már* are:

- (69) a. 7 *fō coruib sō uili teacar*
 and under-their contract_{DAT.PL} **this**_{GEN.PL} all_{GEN.PL.MASC} come_{3SG.PRES.PASS}
 ‘And it is the contracts of all of these that are impugned.’ (CIH 1794.15; author’s trans.)
- b. *Tecur fō coruib sō sīs*
 come_{3SG.PRES.PASS} under-their contract_{DAT.PL} **this**_{GEN.PL} below
 ‘The contracts of all of these below are impugned.’ (CIH 1833.30; author’s trans.)

5 Demonstratives with the deictic particle *í*

The deictic particle *í* followed by a demonstrative can qualify a noun or, combined with the article alone, can be used as a substantive. The former type is discussed in GOI (§ 475.2), where the examples are punctuated *in fer hí-siu*, *in fer hí-sin*, etc., and the latter in GOI (§ 476), with the punctuation *int-í-siu*, *ind-í-siu*, *an-í-siu*, etc. An immediate problem with this interpretation of the demonstratives as enclitic is why the form for ‘this’ should be *-siu*, when the enclitic forms otherwise are *-so*, *-sa*, *-se*, *-seo* and *-sea*. In actual fact, there is enough metrical evidence to confirm that both *í* and *sin* are stressed in this combination. If the word for ‘that’ is stressed, so also must the word for ‘this’, and accordingly the spelling *siu* is to be read with a long diphthong, viz. *síu*, the (independent) dative singular form of *sé*, *só*.

A metrical example which confirms that *í* is a stressed word is:

(70) *In chroch hí as-mbeirid-si*

nos rega int í
 PV-3SG.FEM·GO_{3SG.FUT} the_{NOM.SG.MASC} one
doda-roächt do ráith cháich
do thaithchreic cach bí.

‘That cross you speak of, he will suffer it who has come to it for the sake of all to redeem every living creature.’ (IrGospThomas verse 39)

I have one instance from an Old Irish text, and two from Middle Irish texts (the first of these is early Middle Irish), where its position at the end of a line and rhyme confirm that *sin* following *í* is a separate stressed word:

(71) *bed Ísu ainm ind í sin,*
 COP_{3SG.IMPV} Jesus name_{NOM} the_{GEN.SG.MASC} one that_{DAT.SG}
don domun bid sláinicith.

‘Let Jesus be his name, he will be the saviour of the world.’ (Blathm. verse 155)

(72) *abuir fri Maol a n-í sin.*
 speak_{2SG.IMPV} to Máel_{ACC} the_{ACC.SG.NEUT} ^{NAS}one that_{DAT.SG}
a ghríar ó Aodh a fhir

‘Tell that to Máel, [he will have] all he wishes for from Áed, o man.’ (Byrne 1908: 70.14; author’s trans.)⁴²

⁴² With full rhyme (*deibide nguilbnech*) between *sin* and *fhir*.

- (73) *Airigis Tadc a nn-ī sin*
 observe_{3SG.PRET} Tadc the_{ACC.SG.NEUT} ^{NAS}one that_{DAT.SG}
eatha ūad co Mag Femín
cosin scēl-sin a Leath Chuind
co sīl nAililla nŪlaim

‘Tadc observed that; messengers were sent by him to Mag Femín, with that story from Conn’s Half to the descendants of Ailill Ūlam.’ (RIA, MS 23 P 2 [Book of Lecan] folio 222^{ra}10; author’s trans.)⁴³

6 Middle Irish analytic forms of the verb

In a discussion of the rise of the use of independent pronouns to mark both the subject and the object of verbs in Middle Irish, and the origin of the pairs *sé/é*, *sí/í*, etc., in the third person forms, where Old Irish had only one form (*é*, *sí*, etc.), Greene (1958: 111) remarked: “Probably the forms *ol sé*, *ol sí* (which certainly had fully stressed pronouns by this time, whatever the situation may have been in Old Irish) also contributed to the new development; it was certainly they which determined that the *s-* forms of the third person pronouns should be used as subjects immediately following active verbs.”⁴⁴

In *ol sé*, Middle Irish *ar sé* ‘inquit’, the *sé* most likely was historically the demonstrative, as Quin (1960) argued, although following the then current understanding of the form in question as *se*, with a short vowel. Nevertheless, by the Middle Irish period it had been assimilated to the personal pronoun, as can be seen from the following two lines in *Saltair na Rann*, where *sé* is used to mark direct speech by Adam, but when direct speech by Eve is reported, *sí* is used:⁴⁵

- (74) *ar sé, ar Ādam, fria dagmnaī*
 says he_{NOM} says Adam_{NOM} to-his good.wife_{ACC}
 ‘said he, said Adam, to his good wife’ (SR line 1306; author’s trans.)

- (75) *ar sī, ar Eua fri Ādam*
 says she_{NOM} says Eve_{NOM} to Adam_{ACC}
 ‘said she, said Eve, to Adam’ (SR line 1942; author’s trans.)

⁴³ From a poem in the account of the battle of Crinna.

⁴⁴ See also Roma (2000b).

⁴⁵ Cf. Mac Cana (1984).

Some years ago, I showed that the form of the stressed demonstrative pronoun is in fact *sé*, with a long vowel, and suggested that this Old Irish word had some role to play in the development of the homonymous independent pronoun *sé* ‘he’ beside *é* in Middle Irish (Breatnach 2003: 140). Furthermore, the gradual disappearance of *sé* as a demonstrative in Middle Irish may well indicate a shift in function.⁴⁶ Interestingly, there are quite a few instances in Middle Irish texts of ambiguity in the case of *sé*, that is, where it is not entirely certain whether we have to do with the demonstrative pronoun or with the third singular masculine pronoun.

Given that the development of the pairs *sé/é*, *sí/í*, etc. must have taken some time, it is more likely that the first four examples below, taken from texts belonging to the late Old Irish /early Middle Irish period, are of the demonstrative pronoun, although, at the same time, it is not difficult to see them as equivalent to the pronoun *é*.

- (76) *Ruire ēchtach Eassa Rūaidh,*
immo tteccraitís mórsluaigh,

ass-ib *digh* *mbáis* *bāeghlach* *sé,*
PV·drink_{3SG.PRET} drink_{ACC} ^{NAS}death_{GEN} dangerous_{NOM.SG.NEUT} **this**_{NOM}
īar ccrādh úi Iese.

‘The great-deeded chieftain of Eas-Ruaidh, about whom great hosts used to assemble, he took a lethiferous drink dangerous truly, after persecuting the descendant of Jesse (i.e. Christ).’ (O’Donovan 1856: s.a. 899)⁴⁷

- (77) *Ní* *sé* *sním* *fil* *forn*
NEG-COP_{3SG.PRES} **this**_{NOM} trouble_{NOM} be_{3SG.PRES.REL} upon_{1PL}

‘This is not what distresses us.’ (Mulchrone 1939: line 116, Stokes 1887: 11 [Vita Tripartita])

- (78) *7 ro mbaitis Pátraic oc Sangul .i. sain aingel⁴⁸ dodechoid dia acallaim-sium*
a llá sin

7 *ni* *sé* *Uictor*
and NEG-COP_{3SG.PRES} **this**_{NOM} Victor_{NOM}

⁴⁶ Examples of *co sé* are occasionally found in Middle Irish and Classical Modern Irish (*go sé*); see Breatnach (2003: 138).

⁴⁷ With slightly altered punctuation and word-division, and the addition of macrons over long vowels. In the text *uí Iese* (leg. *uí Iēse*) is glossed *.i. Criost*.

⁴⁸ This serves, of course, as an etymology of the place-name *Sangul*.

‘And Patrick baptised him at Sangal; that is a different angel went to converse with him on that day, and it is not Victor.’ (Mulchrone 1939: line 2417, Stokes 1887: 207 [*Vita Tripartita*])

(79) *Trí coícait láech . . .*

ba sé lucht línaib dindgna
 COP_{3SG.PRET} **this**_{NOM} content_{NOM} number_{DAT.PL} fortress_{GEN.PL}
cach imda de suidib

‘Thrice fifty heroes . . . that was the tale, according to the counts of fortresses, in every chamber of the number.’ (LL lines 3565–8; trans. MD 1: 33)⁴⁹

(80) *Secht cubait . . .*

ba sé tomus in tellaig.
 COP_{3SG.PRET} **this**_{NOM} measure_{NOM} the_{GEN.SG.NEUT} hearth_{GEN}

‘Seven cubits . . . that was the measure of the hearth,’ (LL lines 3573–3576; trans. MD 1: 32–33, lines 53–56)

In examples from later Middle Irish texts, *sé* could be taken as the pronoun, used in positions where *é* subsequently came to be used after the distribution of *sé* and *é* was regularised, although some of those below could just as well be read as demonstratives:⁵⁰

(81) *ba sé iath con-atchetar*

COP_{3SG.PRET} **this**_{NOM} land_{NOM} PV·ask_{AUG.3PL.PRET}

‘that was the land they asked for’, (LL line 19708; trans. MD 3: 441)⁵¹

(82) *ar cipé tí bid sē fot a sāeguil*

for whoever come_{3SG.PRES.SUBJ} COP_{3SG.FUT} **this**_{NOM} length_{NOM} his_{LEN} life_{GEN}

‘for whoever so comes, that will be the length of his life’, (O’Rahilly 1967: 180y-z [LL line 9081 (*Táin Bó Cúailnge*)])

⁴⁹ This and the following example are from the poem *Domun duthain a lainde*, which although edited by Gwynn (MD 1: 28–37) as the fourth poem on Tara, is not part of the *Dindsenchas* proper in LL; the language is earlier than that of the *Dindsenchas* as a whole, either late Old Irish or early Middle Irish.

⁵⁰ For the forms with and without *s-*, the latter normally being used where the pronoun is the subject of the copula or a passive verb, and deviations from the norm, see Breatnach (1994a: 274).

⁵¹ Note that Gwynn reads the variant, *ba hed iath conaichetar* (MD 3: 440), with the third person neuter pronoun.

- (83) *Cocholl Manc[h]in, cid mait sē*
 cow¹_{NOM} Manchín_{GEN} however-COP_{3SG.PRES} good **this**_{NOM}
 ‘The cowl of Manchín, however good this is.’ (Meyer 1892: 129.1; author’s trans.)⁵²

- (84) *Ro cūrad ro sedlad sē.*
 AUG·chastise_{3SG.PRET.PASS} AUG·maim_{3SG.PRET.PASS} **he**_{NOM}
ro dedlad rā dóennige
 ‘He was chastised, he was maimed, he was parted from his misery.’
 (LL line 25708; trans. MD 3: 69)

- (85) *Cāelchéis diaro sernud sē*
 Cāelchéis_{NOM} when-AUG·dispose_{3SG.PRET.PASS} **he**_{NOM}
 ‘When Caelcheis was driven abroad’ (LL line 30203; trans. MD 3: 439)⁵³

I finish this collection with two examples of *sé* as subject, where not only is it separated from its verb (and thus *é* might be expected), but also the translation ‘this’ is more appropriate than ‘it’:

- (86) *Dia lod d’iarair mo leigis*
iar mblíadain rūin ro gabus
rom chuir hi seirg seimne sé
 AUG-1SG·^{LEN}put_{3SG.PRET} in wasting_{ACC} ***_{GEN.SG} **this**_{NOM}
hím-meirbe ocus hím-mignē.
 ‘When I went to seek my cure, after a year, I had kept a secret, which had thrown me into a wasting, into feebleness and into an evil state.’ (Meyer 1903: 48, 52 § 6)

- (87) *conos tuc i sūanbās sē*
 so.that-3PL·bring_{3SG.PRET} in sleep.death_{ACC} this_{NOM}
cēol ro chachain Craiphtine
 ‘so that this brought them into a death-sleep, the music which Craiphtine played’
 (Ó Cuív 1966: 174)

⁵² The fifth line of a six-line verse in the version of *Aislinge Meic Con Glinne* in Trinity College Dublin Manuscript H 3. 18; here *sē* makes end-rhyme with *dē* and *nglé*.

⁵³ Note that Gwynn reads the variant *sí* (MD 3: 438.13).

In conclusion, the simplest way to account for these examples as a whole is to take them as exemplifying *sé* ‘this’ in the course of a gradual shift from demonstrative pronoun to a personal pronoun in complementary distribution with *é*.⁵⁴ Only in some cases, such as when *sé* is combined with *sin*, as in *Ba sé sin búar Flidais* (LU line 1632), ‘That was the cattle of Flidais’, is it clear that *sé* is the pronoun and not the demonstrative.⁵⁵ Similarly, we cannot be absolutely certain of the existence of analytic forms of the verb until we find first and second person pronouns used as the subject of a verb, and in the case of the third person, a plural pronoun, a metrically confirmed stressed feminine singular *sí*, or a masculine singular *hé* used as the subject. All attestations of such forms are in the late 12th-century Book of Leinster.⁵⁶

While this paper is by no means intended to be a comprehensive account of the demonstratives in Old and Middle Irish, I hope to have gone some way towards elucidating the phonology and range of use of *sé/só* and *sin*, and the development of the independent pronoun and the analytic forms of the verb in Middle Irish, as well as providing possible dating criteria for texts. I will end by stressing that in all such work it is essential to use all the means available, especially metrics, to determine Old and Middle Irish forms, and not simply to assume that what holds for Modern Irish also held for the earlier period.

⁵⁴ This complementary distribution of two forms of separate origin, which differ only in the presence or absence of initial *s*-, will have formed the basis for the creation of the other pairs of forms of the independent pronoun, first in the third person forms, viz. *sí/í*, *sed/ed*, and *siat/iat*, and eventually in the first and second person plural forms, viz. *sinn/inn*, *sib/ib*, for which see Breatnach (1994a: 274, 429).

⁵⁵ That these are two separate stressed words is shown by the rhyme in *sē sein* with *nimib* in SR lines 195–196.

⁵⁶ See Breatnach (1994a: 272–273 § 10.19) and Breatnach (2015: 72–73).

Carlos García-Castillero

6 Paradigmatic split and merger: The descriptive and diachronic problem of Old Irish Class B infixed pronouns

1 Introduction: Infixes pronouns and clause types in the Old Irish verbal complex

Infixes pronouns are one of the formal strategies used in the Old Irish verbal complex to distinguish clause types, in such a manner that the sole formal opposition between Classes A/B and Class C serves to express the opposition between declarative and relative clause type respectively in lexical compound verbs which take an infixes pronoun.

The use of the Classes A and B of infixes pronouns, which mark declarative clause type, is determined by the phonotactic structure of the (first) lexical preverb of the verbal compound appearing in the pretonic part of the verbal complex. The general rule is that lexical preverbs which end with a consonant (henceforth also (-)VC- lexical preverbs), with the exception of *imm-* and *ar-*, take Class B infixes to express declarative clause type. For instance, the Old Irish verb *as·beir* '(s)he says', with the (-)VC- preverb *as-*, takes the Class B third person singular neuter infixes form $-t^L-$, so that $a(s)- + -t^L- \rightarrow at-$ in *at·beir* '(s)he says it'. By contrast, lexical preverbs which end with a vowel (henceforth also CV- lexical preverbs) make use of the so-called Class A of infixes pronouns, so that e.g. the lexical compound *do·beir* '(s)he brings, gives', with the lexical preverb (*to-* >) *do-*, expresses the same third person singular neuter pronominal reference by substituting the *-o-* vowel of the lexical preverb by the corresponding Class A infixes form $-a^L-$, i.e. $d(o)- + -a^L- \rightarrow da·beir$ '(s)he brings, gives it'. The lexical preverbs *imm-* and *ar-* also make use of Class A of infixes pronouns.

The relative forms which minimally contrast in clause type with the above mentioned declarative forms *at·beir* '(s)he says it' and *da·beir* '(s)he brings, gives it' are *ass-id·beir* 'who says it' and *do-d·beir* 'who gives it' respectively, which include the mentioned Class C forms of the third person neuter infixes pronoun (i.e. $-[i]d^L-$).

The ultimate aim of this study is to explain the diachronic origin of Class B of infixes pronouns, but, as a prerequisite for this, and also as a point which is in itself worth discussing, the exact morphological and syntactic circumstances of these infixes pronouns must also be investigated in the corpus

of the contemporaneous Old Irish glosses. This descriptive question can be briefly referred to as follows. The distinction between Classes A/B (for declarative clause type forms) and Class C (for relative clause type forms) is quite regularly made in Old Irish when a third person pronoun is infix in the lexical compound. However, things are different with a first or second (henceforth also non-third) person pronominal infix. In the language of the contemporaneous Old Irish texts, non-third person infix pronouns are much less regular in making that distinction between declarative and relative form and show a very remarkable behaviour especially when the lexical preverb after which the infix pronoun appears is of type (-)VC-, that is to say, when the declarative clause type infix pronoun must be of Class B. In that situation, Class B is most often used in cases in which relative clause morphology (i.e. a Class C form) is expected. Though less frequently, non-third person Class A infix pronouns also appear in cases in which relative morphology is expected.

The descriptive question is then how to deal with this asymmetry in the use of Classes C and B depending on whether the involved infix is of a third or non-third person, and the position taken in this paper is that this situation of asymmetry observed in the contemporaneous Old Irish texts is directly related to the question on the diachronic origin of the Class B of infix pronouns.

In order to answer this question, sections 2 to 4 provide a detailed description of the situation of Class B pronouns as they are used in the Old Irish glosses. In particular, section 2 provides basic information about the Old Irish verbal complex and the category of clause typing expressed in it; section 3 provides a list of lexical preverbs which take either Classes A or B of infix pronouns, presents the whole paradigm of the three classes of infix pronouns, and illustrates the use of Classes A/B instead of the expected Class C; section 4 lists the forms attested in the three main collections of glosses which show a non-third person infix pronoun after a pretonic lexical preverb of the (-)VC-type, that is to say, of verbs which must take Class B for the declarative clause type forms. On the basis of the previous description, section 5 gives a proper formulation for the diachronic question referred to above, discusses previous diachronic explanations, and provides some basic aspects regarding the etymology of the other classes of infix pronouns, as well as of some lexical preverbs. Section 6 elaborates a diachronic explanation for the Old Irish Class B of infix pronouns which is congruent with the previous description and which also provides a justification for this formal distinction in the infix pronouns used for declarative clause type. Section 7 summarises the main points of the paper.

2 Some background about the Old Irish verbal complex

The initial statement of this paper is, as stated above, that Classes A/B of infixed pronouns are used for declarative and Class C for relative clause types. This section only refers to two issues on these pronominal references. For more aspects of the Old Irish verbal complex, I refer to the treatment in García-Castillero (2012, 2014, 2015, 2020).

The first issue is that the Old Irish infixed pronouns are morphological elements which always appear after a previous pretonic element, which may be a conjunct particle (i.e. a pretonic element of a grammatical nature), or a lexical preverb, which constitutes a semantic unit with the verbal stem (see Table 1 below on page 148). Handbook examples of the combination of a verb with conjunct particles are *ní·beir* and *nad·beir*, from the simple *beirid* ‘brings’. Both being the third person singular of the present indicative active, the former is marked as a negative declarative clause type form (‘[s]he does not bring’), and the latter as a negative relative clause type form (‘who does not bring’ or ‘whom/which s/he does not bring’). The form *nad·beir* must be understood as including relative lenition (i.e. the change of underlying /b^j/ to /v^j/ producing /nað^ver^j/), although this mutation is not graphically marked in Old Irish when it applies to voiced obstruents. Relative lenition involves the phonological fricativisation of the first consonant of the basic form of the verb, in this case a voiced bilabial plosive; the sound /f/ is deleted and /s/ becomes an aspiration; vowels are not affected by lenition. The other relative mutation used in the Old Irish verbal complex is the so-called relative nasalisation, which formally involves the addition of a nasal sound to a voiced plosive (i.e. *nad·mbeir*), or to a vowel, and the voicing of a basic voiceless plosive. The functional side of these two relative mutations does not need to be considered now. The important point for the use of Class B infixes is that it is only combined with the lexical preverbs of the type (-)VC- to be observed in the next section, whereas Class A is combined with lexical preverbs (of the type CV-) and conjunct particles; by themselves, these two classes express declarative clause type. Class C infixes, which express relative clause type, are combined with lexical preverbs of whichever phonotactic type and conjunct particles.

The second issue is related to the two grammatical categories which cross-cut in the infixed pronoun, which must therefore be considered basically as a portmanteau morpheme expressing pronominal reference and clause type at the same time. The Old Irish verbal complex regularly distinguishes six clause types by means of several formal procedures. The two most important clause

types for this paper are the declarative and the relative (where the leniting and the nasalising variants mentioned must be included); in addition, the Old Irish verb distinguishes content (or *wh*-)interrogative, polar (or yes-no) interrogative, responsive and imperative clause type forms. Some of these clause types will be mentioned later, and are characterised by the use of one of the classes of pronominal infixes, as also detailed in the next section. As for the distinction between declarative and relative clause types, it must be stated that the formal opposition between Classes A/B and C is one of the formal strategies which suffice by themselves to distinguish those two clause types in the Old Irish verbal complex, as illustrated in the examples of (1) of the next section. The other formal means are different sets of endings (the so-called absolute endings, where there are both declarative and relative absolute endings), the so-called relative mutations (which contrast with the lack of them), as well as special conjunct particles such as the negative declarative *ní*- and the negative relative *nad*- mentioned above.

3 Formal and functional aspects of the Old Irish Classes A, B and C of pronominal infixes

This section provides the basic descriptive tools to understand properly the problems considered in this paper. Section 3.1 illustrates the basic distinction referred to in the previous section with attested forms including third person singular neuter infixed forms. In section 3.2 the whole set of the infixed pronouns used in the Old Irish verbal complex as well as the main issues of their use are introduced, paying special attention to the formal features of Classes B and C. Section 3.3 centres on the morphological process which must be assumed in the expression of the third person singular masculine / neuter of Class A infixed pronouns. Section 3.4 establishes the proper context in which the asymmetry between third and non-third persons mentioned at the outset must be considered, and focuses on the use of Class A instead of expected Class C as a special case of the general phenomenon which involves the possibility of using either declarative or relative clause type marking in the same syntactic context.

3.1 Basic functional distinction between Classes A/B and C

Classes A and B basically mark the corresponding verbal complex as a declarative clause type verb, as already stated, and are also used in the imperative verb, which makes use of a partly different set of inflectional endings. The pronominal

infixes of Class C are used to mark relative clause type and some other subordinate clauses, and are also used in polar interrogative clause type forms, which are constitutively marked by the conjunct particle *in*^N. In lexical compounds which have no conjunct particle in the pretonic slot, i.e. which have a lexical preverb in the pretonic slot, this difference between Class A/B and Class C expresses by itself the difference between declarative and relative clause type respectively. The forms in (1) and (2) illustrate this clause type difference in minimal or quasi-minimal pairs of forms attested in the Old Irish glosses.

- (1) a. *darigni*
 PV-3SG_{NEUT}(A)·do_{AUG.3SG.PRET}
 ‘(David) has done it.’ (Ml. 51^d2)
- b. *dudrigni*
 PV-3SG_{NEUT}(C)·do_{AUG.3SG.PRET}
 ‘who has done it’ (Ml. 124^b3)
- (2) a. *air atroilli* *dia*
 for PV-3SG_{NEUT}(B)·deserve_{3SG.PRES} God_{NOM}
 ‘for God deserves it . . .’ (Ml. 51^d12)
- b. *donaib hí assidroillet*
 to_{3PL}·the_{DAT.PL} one PV-3SG_{NEUT}(C)·deserve_{3PL.PRES}
 ‘to those that deserve it’ (Ml. 54^d6)

The forms in (1) are both based on the third person singular perfect active *do-rigni* ‘(s)he has done’, of *do·gní* ‘does, makes’, where *do-* is a CV-lexical preverb. In (1a), this form takes the third person singular neuter infix of Class A, whereas in (1b), it takes the corresponding Class C form. The two examples of (2) include forms from the verb *ad·roilli* ‘deserves’. In (2a), *at·roilli* has the Class B third person singular neuter infix and therefore counts as a declarative clause type form. Note, in the relative form *assidroillet* in (2b), the use of the preverb *as-* instead of the original *ad-*: as stated by Thurneysen (GOI § 822), this is due to the loss of formal distinctiveness between those preverbs *as-* and *ad-* when they are combined with Class B infixed pronouns, both with the form *at-*: cf. *at·roilli* ‘[God] deserves it’ in (2a) and the form *at·beir* ‘(s)he says it’ quoted in the introduction, which belongs to the basic form *as·beir* ‘(s)he says’. Though this is not exactly the case of the loss of formal distinctivity assumed later on in this paper as the trigger of the creation of the Old Irish Class B of infixed pronouns, it illustrates how formal distinctions can be lost in the pretonic part of the verbal complex.

3.2 The Old Irish lexical preverbs and the distribution of Classes A/B

As also anticipated at the outset, the phonotactic structure of the (first) lexical preverb of the basic lexical compound decides the shape of the infixed pronoun expressing declarative clause type, i.e. whether Classes A or B are to be used. According to the description in GOI (§§ 411–412), the general rule is that, if the lexical preverb ends in a vowel, i.e. if it has the shape CV-, then Class A is used; if the lexical preverb ends in a consonant, i.e. (-)VC- (except *imm-* and *ar-*, which bear Class A infixes, which originally had the shape CV-, i.e. which come from a form with original vocalic auslaut), then the infixed pronouns of Class B are used. Table 1 includes the most relevant lexical preverbs in the shape that they adopt in Old Irish in declarative clauses without infixed pronoun,¹ and – for the (-)VC-preverbs – it also gives the assumed Primitive Irish form between parentheses and the basic form that they adopt in combination with a Class B infixed pronoun.

Table 1: Old Irish lexical preverbs and their infix class in declarative clause type form.

Lexical preverbs with Class A		Lexical preverbs with Class B	
CV-	(-)VC-	(-)VC-	
<i>ro-</i>	<i>imm-</i> ‘about’	(* <i>kom-</i>)	<i>con-</i> ‘with’ → <i>cot-</i>
(<i>to ></i>) <i>do-</i> ‘to’	<i>ar-</i> ‘for’	(* <i>in[d]-</i>)	<i>in(d)-</i> ‘in’ → <i>at-</i>
<i>di/e-</i> / <i>do-</i> ‘from’		(* <i>ad-</i>)	<i>ad-</i> ‘to’ → <i>at-</i>
<i>fo-</i> ‘under’		(* <i>ath[ij]-</i>)	<i>ad-</i> ‘re-’ → <i>at-</i>
		(* <i>ess-</i>)	<i>as-</i> ‘out of’ → <i>at-</i>
		(* <i>uss-</i>)	<i>as-</i> ‘up, out’ ² → <i>at-</i>
		(* <i>frith-</i>)	<i>fris-</i> ‘towards’ → <i>frit-</i>
		(* <i>for-</i>)	<i>for-</i> ‘over’ → <i>fort-/d-</i>
		(* <i>eter-</i>)	<i>eter-</i> ‘between’ → <i>etart-/d-</i>

The meanings adduced must be understood as orientative and, to a great extent, etymologically based. In not a few compounds, however, these meanings have been blurred, so that they are not anymore distinguished.

¹ Class B is also found in combination with the de-adjectival preverb *mí-* ‘badly, mis-’ in *mit-nimret* ‘that they deceive him’ (Ml. 74^b22 = *mi-t.Nimret*, from the verb *imm-beir* ‘plays, handles’). See García-Castillero (2014) for this type of preverbal element, which is much less frequent than the conjunct particles and lexical preverbs.

² For this semantic interpretation, see Russell (1988: 125).

Despite the fact that there are more preverbs which bear Class B, as it can be observed in Table 1, the ones taking Class A are more frequent, i.e. there are more verbal compounds with those lexical preverbs. In addition to that, whereas Class B is only used with the (-)VC- lexical preverbs, Class A is regularly used also with some very frequent conjunct particles such as the declarative negative particle *ní-* seen in the previous section, or the meaningless particle *no-*, one of whose main functions is precisely to provide a pretonic element after which the equally unstressed infixed pronoun can appear.

The combination resulting from the preverb with the Class A infix constitutes a phonotactically adequate sequence in most forms of the corresponding paradigm. The preverbs *imm-* and *ar-*, to use the pretonic forms used in declarative verbal complexes, are characterised by the addition of a vowel after the final consonant of the form before the Class A pronominal form. These two features are illustrated in Table 2, which includes the CV- lexical preverbs *to-* and *di/e-* with Class A infixed pronouns, and the combination of the former with Class C, on the one hand, and the specific form of the preverb *imm-*, on the other. The (-)VC- lexical preverb included in Table 2 is *con-*, a type of preverb which ends with a nasal and in which the difference between Classes B and C is most conspicuous.

Table 2: Some Old Irish lexical preverbs with infixed pronouns of Classes A, B and C.

	Class A			Class B	Class C	
	<i>to-</i>	<i>di/e-</i>	<i>imm-</i>	<i>con-</i>	<i>con-</i>	<i>to-</i>
1sg.	<i>do-m^L</i>	<i>do-m^L</i>	<i>imm-um^L</i>	<i>cotam^L</i>	<i>condam^L</i>	<i>do-dam^L</i>
2sg.	<i>do-t^L</i>	<i>do-t^L</i>	<i>imm-ut^L</i>	<i>cotat^L</i>	<i>condat^L</i>	<i>do-dat^L</i>
3sg.masc.	<i>d(o)-a^N</i>	<i>d(i/e)-a^N</i>	<i>imm-a^N</i>	<i>cot^N</i>	<i>condid^N</i>	<i>do-d^N</i>
3sg.neut.	<i>d(o)-a^L</i>	<i>d(i/e)-a^L</i>	<i>imm-a^L</i>	<i>cot^L</i>	<i>condid^L</i>	<i>do-d^L</i>
3sg.fem.	<i>do-s-</i>	<i>do-s-</i>	<i>imm-us-</i>	<i>cota-</i>	<i>conda-</i>	<i>do-da-</i>
1pl.	<i>do-n-</i>	<i>do-n-</i>	<i>imm-un-</i>	<i>cotan-</i>	<i>condan-</i>	<i>do-dan-</i>
2pl.	<i>do-b-</i>	<i>do-b-</i>	<i>imm-ub-</i>	<i>cotab-</i>	<i>condab-</i>	<i>do-dab-</i>
3pl.	<i>do-s^N</i>	<i>do-s^N</i>	<i>imm-us^N</i>	<i>cota^N</i>	<i>conda^N</i>	<i>do-da^N</i>

3.3 The morphology of the Class A and B infixed pronouns

This section focuses on some pronominal infixes given in the previous section which involve morphological processes other than the mere addition of a segment. Specifically, I refer to the third person singular masculine / neuter forms of Class A and to the combination of some lexical preverbs with the Class B forms.

As for the third person singular masculine / neuter pronouns in CV- lexical preverbs such as (*to-* >) *do-* and *di/e-/do-* (see Table 2), it seems that the synchronically most adequate description of the morphological process concerned is that they are the outcome of a process of replacement or substitution of the final vowel of the preverb by the vowel *-a-* plus the corresponding mutation. That is to say, the lexical preverbs (*to-* >) *do-*, *di/e-/do-*, *ro-*, *fo-* and the conjunct particle *no-* take out their vowel in order to include the vowel *-a-* which characterises those two persons: e.g. *do-* + *-a*^{N/L} → *d(o)-a*^{N/L}, i.e. *da*^{N/L}. This analysis is perfectly compatible with the diachronic process of vowel elision, to be considered in section 5 below. In this light, the Class A third person singular masculine / neuter infixes seem to be a good example of replacive morphology (cf. Spencer 1998: 140–141), and are relevant in the context of this paper for the following two reasons.

Firstly, the outcome of that process of morphological replacement may be the reason for the formal assimilation of the CV- lexical preverbs (*to-* >) *do-* and *di/e-/do-* ‘from’ in pretonic position, a position in which their respective vowels should have been preserved as distinct. To be sure, the forms *de* and *di* can still be found in some verbs such as e.g. *de-meccim* ‘I despise’ (Sg. 39^b1), but the extremely frequent compound *do-gní* ‘does, makes’ quoted above in (1), which is formed with the preverb *de-* ‘from’, has virtually only *do-* in pretonic position. In other words, I contend that one of the factors which led to the lack of distinction of the pretonic version of *di/e-* and (*to-* >) *do-* (both appearing in Old Irish as *do-* in most cases) was the substitution of the characteristic vowel of the preverb when it was combined with the third person singular m./n. infixed pronoun *-a*^{N/L}, where both the preverbs (*to-* >) *do-* and *di/e-* appeared in Old Irish as *da*^{N/L}. Other factors which have surely played a role in that process are the coincidence in the consonant due to the change (*to-* >) *do-* in pretonic position (see GOI § 178.2) and, eventually, the loss of semantic identity of the element involved due to the lexicalisation of the meaning of the compound. All those conditions meet in the verb *do-gní* just quoted. The same reason has been adduced for the confusion of lexical preverbs with the shape (-)VC-, as in e.g. the form *assidroillet* quoted in example (2b) above.³

³ Not every case of loss of distinctiveness between lexical preverbs is left without response in Old Irish. The phenomenon known as ‘split *for*’ is in the end an attempt to maintain the

Second, this idea of replacement as a morphological process operating in the combination of Old Irish lexical preverbs with infixes can be applied perfectly to Class B of infixes. Adopting the same synchronic perspective as for the Class A third person singular masculine / neuter forms above, the phonotactic combinations which can be considered for the (-)VC- lexical preverbs are the following: (a) the combination of the Class B of infixes with the preverbs which end with a nasal (*con-*, *in-*) implies that the final nasal is substituted by the assumed /d/ of the infix (i.e. *con-* + /d-/ → *cot-* /kod/); the same replacive process seems to apply to the combination of Class B infixes with preverbs in -s, i.e. (*ess-*, *uss-* >) *as-* and *fris-*, which give *at-* and *frit-* respectively; (b) when combined with preverbs ending with /r/ (*for-*, *etar-*), then the Class B pronoun is simply added to the preverb form and its initial dental sound is spelt either as -d- or as -t-; see the attested forms in Table 3 of section 4 below; (c) for the lexical preverbs in a dental fricative (e.g. *ad-* /að/), the process at stake seems to be that the assumed /d/ of the Class B infix again takes the place of the final consonant of the preverb or, alternatively, that the consonant of the preverb and that of the infix have ‘merged’ into a fortis consonant.

The diachronic discussion on Class B infixes is a matter of sections 5 and 6 below, but it is worth noting that the interpretation of the form of Class B as containing a /d/ sound is not the only one possible. In particular, the form *at-* corresponding to *ad-* /að/ may well be the outcome of a merger of the final lenis of the preverb and an initial lenis sound, as if it were *ad-* /að/ + /ð/ > *at-* /ad/. A parallel process may be the case of *nepuid* ‘not-being’ /‘nebuð/ derived from **neβ’βuθ*¹ considered in GOI (§ 137).

3.4 The use of Class A instead of Class C with 1st and 2nd person infixes

In compound verbs with a lexical preverb of whichever phonotactic structure in the pretonic position and which include an infix marking declarative clause type (Classes A/B) and the infix marking relative clause type (Class C) is regularly made with third person pronouns. This opposition has been already illustrated in examples (1) and (2) above.

difference between two semantically opposed lexical preverbs, *for-* ‘over’ and *fo-* ‘under’, in some specific morphological combinations in which they could be confused. See García-Castillero (2017) for this question.

However, compound verbs taking a first or second person infix are less systematic in this regard, so that the pronominal infixes of Classes A and B sometimes appear in forms in which relative clause type morphology is expected. In fact, this alternation between declarative and relative clause type morphology is a widespread phenomenon in Old Irish, not restricted to infixed pronouns; the reader may consult Ó hUiginn's (1986, 1998) studies quoted below in this section.

One may therefore distinguish three groups of clauses according to whether declarative or relative clause type morphology (not only infixed pronouns) is associated with them. Group I consists of main declarative clauses and some specific subordinate clauses which are characterised by the regular use of declarative morphology. Group II consists of subordinate clauses which display both declarative and relative verbal forms. Group III consists of subordinate clauses which regularly show relative morphology. The verbal complex which has the relative conjunct particle $-(s)a^N$, as well as other conjunct particles such as the polar interrogative one already mentioned, are not mentioned in any of these groups because such conjunct particles do not appear in the attested forms included in Table 3 below.

- (I) Verbs with declarative clause type morphology are regularly used in:
 - (a) main declarative clauses,
 - (b) cleft sentences with an anteposed oblique constituent (i.e. a prepositional phrase),
 - (c) adverbial subordinates introduced by co^L 'so that', ma^L 'if', cia^L 'though'.⁴
- (II) Verbs with either (nasalising) relative or declarative clause type morphology are used in:
 - (d) in complement (or noun) subordinate clauses,
 - (e) in adverbial subordinate clauses introduced by $iarsindí$ 'after', $lase$ 'when' (mostly with a relative verb), $amal$ 'as', $(h)óre$ 'because'.
- (III) Verbs with relative clause type morphology are regularly used in:
 - (f) restrictive relative clauses of the leniting or nasalising type,

⁴ In the language of the Glosses, a meaningless Class C third person singular neuter infixed pronoun $-d^L$ appears regularly in the verbal complexes in indicative mood after the subordinating conjunctions ma^L 'if' and cia^L 'though', provided that there is no other semantically full infixed pronoun. In line with the description in García-Castillero (2020, Chapter 5), this use of $-d^L$ is to be interpreted as the introduction of a marker of syntactic dependency. If these conditions are not met, these two subordinating conjunctions ma^L 'if' and cia^L 'though' are regularly followed by a declarative clause type verbal complex. I hope to deal with the subordinating conjunction co^L 'so that' and its relationship with the almost synonymous conjunct particle co^N in a future study.

- (g) cleft-sentences with an anteposed subject or object NP (including cases of *figura etymologica*, which involve nasalising relative marking),
- (h) relative clauses after the light heads *a^N* and *intí aní*,⁵ and in subordinate clauses after the temporal conjunctions *a^N* ‘when’ and *inta(i)n* ‘when’,
- (i) relative clauses of types (f) after the stressed interrogative pronouns *cia cid* ‘who, what’.

The observed variation between declarative and relative morphology in Group II is determined by various factors: see Ó hUiginn (1998: 126–130) for the variation in complement clauses, and García-Castillero (2020, ch. 5) for the variation in the third person singular of the copula after *amal* and (*h*)*óre*. Among these factors, person plays a prominent role. By person, I refer to the cases in which the involved verbal complex includes either only one pronominal reference expressed by means of an inflectional ending or two pronominal references, one of them an infix pronoun. The general tendency is that non-third persons, whether in the inflectional ending or in the infix pronoun (or in both), favour the use of declarative clause type morphology.

As an illustration consider (3), taken from Ó hUiginn (1986: 43, 45). Example (3a) shows relative morphology by means of the pretonic particle *no-* followed by nasalisation (graphically not marked in *nocretim* /no’grⁱed-ⁱim/), while *pridchim* in (3b), which lacks the pretonic particle, represents a declarative clause type form.

- (3) a. *hóre nocretim ísu*
 because PV.^{NAS}believe_{1SG.PRES} Jesus_{ACC}
 ‘because I believe in Jesus’ (Wb. 1^a2)
- b. *hore pridchim soscele do gentib*
 because preach_{1SG.PRES} gospel_{ACC} to Gentile_{DAT.PL}
 ‘because I preach (the) gospel to the Gentiles’ (Wb. 5^c6)

The variation between Classes A/B and C for infix pronouns of whichever person in the syntactic structures mentioned in Group II is to be accounted for as a part of the same variation which is observed in verbal complexes without infix pronoun. The pair of glosses in (4) shows the variation between declarative and relative third person singular neuter forms in a verbal complex after

⁵ According to the general definition provided in the works quoted in García-Castillero (2018: 48–49), a “light head” is a demonstrative pronoun which is (more or less exclusively) used as the head of a relative clause.

(*h*)óre ‘because’. In (4a), we have a Class C infix in the perfect form of *as-beir* ‘says’ (to be analyzed as *as-ind-rubartatar*). In (4b), the perfect form *fritracatar*, from the verb *fris-accai* ‘hopes for’, shows the Class B infixed form which characterises it as a declarative clause type form. This sort of variation is more frequent with non-third persons, a nice example being the gloss in (5), which has a verb with a Class C form (i.e. *no-n-dob-molor-sa*, from the simple *molaithir* ‘praises’) coordinated with another verbal complex including a Class A infixed form (i.e. *no-m-móidim*, from the simple *moidid* ‘boasts’), both depending on the previous *hore*, again the syntactic structure of type (e) above.

- (4) a. [. . .] *huare asinrubartatar* tris pueri
 because PV-^{NAS}3SG_{NEUT}(C)·say_{AUG.3PL.PRET} three children_{NOM.PL}
 ‘[. . .] because *tres pueri* had said it’ (Ml. 131^d12)
- b. [. . .] *huare fritracatar som* a deo
 because PV-3SG_{NEUT}(B)·hope_{AUG.3PL.PRET}=3SG_{NEUT} from God_{ABL}
 ‘[. . .] because they have hoped for it *a deo*’ (Ml. 131^c10)
- (5) *hore nondobmolorsa* et *nom móidim* *indib*
 because PV-^{NAS}2PL(C)·praise_{1SG.PRES}=1SG and PV-1SG(A)·boast_{1SG.PRES} in_{2PL}
 ‘because I praise you and boast myself in you’ (Wb. 14^c18)

The important aspect at this moment is that non-third person infixed pronouns of Class A (i.e. infixed pronouns which value as declarative clause type markers) appear even in subordinate clauses included in Group III of the above classification, that is to say, in syntactic contexts in which relative morphology is consistently used. This can be observed in (6) and (7) with verbal complexes which have the conjunct particle *no-* and a CV- lexical preverb respectively. Both cases of (6) involve the same syntactic structure, i.e. a cleft sentence with anteposed subject, i.e. type (g) of Group III: example (6a) uses the expected Class C form *-don-* in *nodonnertani*, from the simple *nertheid*, but example (6b) shows Class A *-m-* in the simple *beoigidir*. The morphosyntactic structure in (7) corresponds to type (h) of Group III, i.e. the verbal complex introduced by the conjunction *a^N* ‘when’, and (7a) shows the expected Class C form *-dat-* in *afundatferai*, from *fo-fera*, whereas the form in (7b) has the Class A form *-m-* in *andumsemmat*, from *do-seinn*.

- (6) a. *is hé nodonnertani*
 be_{3SG.PRES} he PV-1PL(C)·strengthen_{3SG.PRES}=1PL
 ‘It is He that strengthens us.’ (Wb. 6^d11)

b. *is iress crist nombeoigedar*
*be*_{3SG.PRES} *faith*_{NOM} *Christ*_{GEN} *PV-1SG(A)·quicken*_{3SG.PRES}
 ‘It is Christ’s faith that quickens me.’ (Wb. 19^a20)

- (7) a. *afundatferai*
*when-PV-^{NAS}2SG(C)·present*_{2SG.PRES}
 ‘when you (sg) present yourself (sg)’ (Ml. 38^c26–27)
- b. *andumsennat*
*when-^{NAS}PV-1SG(C)·pursue*_{3PL.PRES}
 ‘when they pursue me’ (Ml. 39^c28)

Even though it is far from being usual, because the ‘expected’ use of Class C is also frequently encountered in those types included in Groups II and III, the use of Class A instead of Class C is a well-established fact.⁶ The extent to which non-third person infixed pronouns of Class B are used instead of the expected Class C form, that is to say, cases that run parallel to those in (6b) and (7b), is discussed in the next section.

4 The descriptive problem: The opposition between Class B and Class C

In contrast to what can be observed for the cases in which the forms of Class A are involved, the use of non-third person infixed pronouns of Class B instead of the Class C counterparts seems to be the rule. The list of forms included in Table 3 below is based on the collection provided by Sommer (1897) and has been revised with the aid of Kavanagh (2001), Griffith and Stifter (2013) and Bauer (2015).

Table 3 is to be interpreted in the following way: the dictionary headword of the Old Irish verb is given in the leftmost column, the attested form and relevant syntactic structure with its English translation appear in the central column, and the following three columns to the right, headed by the signs (I), (II) and (III), correspond to the three main groups of syntactic structures considered in the previous section. In each of the rightmost columns, the following

⁶ Of the 77 verbal complexes with a non-third person infix attested in Wb., Ml. and Sg. with a pretonic CV- lexical preverb or with a conjunct particle appearing in a syntactic context of Group III, i.e. cases such as those in (6) and (7), 33 cases bear Class A, i.e. 43%, e.g. (6b) and (7b), and 44 cases bear Class C forms, i.e. 57%, e.g. (6a) and (7a).

Table 3: Old Irish 1st and 2nd person infixes of Classes B and C in their syntactic context.

Lemma form	Attested form	I	II	III
<i>as-beir</i> 'says'	<i>amal asndonberat</i> 'as they say of us' (Wb. 2 ^a 12)		C (e)	
	(plebs dei) <i>asndanberthe ni</i> 'It is plebs Dei that we used to be called.' (Ml. 114 ^a 7)			C (g)
<i>as-rochoili</i> 'determines'	<i>atamrochoilse</i> 'Determine me!' (impv.) (Ml. 24 ^a 15)	B (a)		
<i>as-scarta</i> 'drives away'	<i>ma atamscartisse</i> 'if they were to drive me' (Ml. 59 ^a 21)	B (c)		
<i>con-airléici</i> 'permits'	<i>condammairleicea</i> 'that He should let me' (Ml. 38 ^a 11)		C (d)	
	<i>iarsindi cotanrairlíc</i> 'after He let us go' (Ml. 125 ^a 9)		B (e)	
<i>con-boing</i> 'smashes'	<i>cochtabosadsi</i> 'so that he should crush you' (Ml. 18 ^a 7)	B (c)		
<i>con-delca</i> 'compares'	<i>frinn fanisin cotondelcfam</i> 'with ourselves we will compare ourselves' (Wb. 17 ^b 10)	B (b)		
<i>con-éicnigethar</i> 'compels'	<i>ithéside cotammeignighthersa</i> 'It is these by which I am compelled.' (Ml. 21 ^b 10)	B (b)?		B (g)?
<i>con-erchloi</i> 'leads'	<i>cotomerchloither</i> 'I am led.' (Sg. 17 ^a 7), gl. <i>agor</i>	B (a)		
<i>con-nerta</i> 'strengthens'	<i>cototnertsu</i> 'Strengthen thyself!' (impv.) (Wb. 30 ^a 9)	B (a)		
<i>con-ocaib</i> 'lifts up, raises'	<i>an condammucbaitisse</i> 'when they used to beatify me' (Ml. 39 ^d 11)			C (h)
	<i>cotabucabarsi</i> 'Be lifted up!' (impv.) (Ml. 46 ^a 7)	B (a)		
<i>con-oscaigi</i> 'moves'	<i>cotammoscaigse</i> 'I should move [in the mountains].' (Ml. 29 ^d 3)	B (a)		
	<i>condatoscaigther</i> 'that you might be moved' (Ml. 23 ^d 21), gl. <i>commouere</i>		C (d)	
	<i>cotatoscaigthersu</i> 'Be moved, O God!' (impv.) (Ml. 58 ^d 14)	B (a)		
<i>con-rig</i> 'binds'	<i>cotobárrig</i> '[he] has constrained you' (Wb. 9 ^b 19)	B (a)		
	<i>cotanrirastarni</i> 'We will be bound.' (Ml. 134 ^a 1)	B (a)		
<i>con-secha</i> 'corrects'	<i>cotob sechfider</i> 'Ye will be corrected.' (Wb. 9 ^a 23)	B (a)		

Table 3 (continued)

Lemma form	Attested form	I	II	III
<i>con-utuinc</i> 'builds'	<i>cotofutaincsi</i> (MS <i>cotofutaircsi</i>) 'He upbuilds you.' (Wb. 8 ^c 16)	B (a)		
<i>in-árban</i> 'impels'	<i>atatairbined su</i> 'Let it impel You.' (impv.) (Ml. 86 ^c 10)	B (a)		
<i>in-greinn</i> 'persecutes'	<i>atamgrennat</i> 'They persecute me.' (Ml. 39 ^d 13)	B (a)		
	<i>donaib hí atamgrennat</i> 'to those who persecute me' (Ml. 127 ^c 8)			B (f)
	<i>honaib hí atamgrennat</i> 'by those who persecute us' (Ml. 45 ^a 16)			B (f)
<i>ind-saig</i> (<i>Jad-saig</i>) 'approaches'	<i>frisna preceptorí atobsegatsi</i> 'like the preachers who go to you' (Wb. 14 ^d 37)			B (f)
<i>in-snaid</i> 'inserts'	<i>coatomsnassar</i> 'that I may be engrafted' (Wb. 5 ^b 30), gl. <i>ut ego inserer</i>	B (c)		
<i>in-sorchaigedar</i> 'illuminates'	<i>coatabsorchaighther</i> (MS <i>coatabsorchaither</i>) 'that you may be illuminated' (Ml. 53 ^b 15)	B (c)		
<i>in-togair</i> 'invokes'	<i>indattogarsa</i> 'that I invoke you' (Ml. 72 ^c 4), gl. <i>inuocandi te</i>		C (d)	
<i>ad-aig</i> 'drives'	<i>massuthol atomaig</i> 'if it is desire what drives me' (Wb. 10 ^d 26)			B (g)
	<i>dílmaine aisndáisen atannaighni</i> 'Licence of narration impels us.' (Ml. 93 ^d 12)	B (a)?		B (g)?
	<i>isfoirbthetu hirisse attotaig</i> 'It is perfection of faith that impels thee.' (Ml. 93 ^d 12)			B (g)
	<i>cid atobaich</i> 'What impels you?' (Wb. 9 ^c 20)			B (i)
	<i>cid atobaig dó</i> 'What impels you to it?' (Wb. 19 ^d 10a)			B (i)
<i>ad-anaig</i> 'brings'	<i>atomanaste</i> 'that I should be brought' (Wb. 14 ^c 20), gl. <i>a uobis deduci</i>		B (d)	
<i>ad-cí</i> 'sees'	<i>atatchigestar</i> 'You will be seen.' (Ml. 59 ^c 12)	B (a)		
	<i>atobcísíde</i> 'He perceives you,' (Wb. 25 ^a 26)	B (a)		
<i>ad-cumaig</i> 'happens'	<i>cindas persine attotchomnicc</i> 'What sort of person art thou' (lit. 'what sort of person is it that has befallen you?') (Wb. 6 ^b 13)			B (f)

Table 3 (continued)

Lemma form	Attested form	I	II	III
<i>ad-ella</i> ‘visits’	<i>atdubelliub</i> ‘I will visit you.’ (Wb. 7 ^a 4)	B (a)		
<i>ad-eirrig</i> ‘emends’	<i>atanneirrig</i> ‘who emends us’ (ML. 114 ^d 10), gl. <i>qui nos</i> [. . .] <i>emendat</i> ¹⁰			B (f)
<i>ad-gaib</i> ‘reprehends’	<i>atabgabed</i> ‘Let it reprehend you (pl).’ (impv.) (ML. 20 ^d 11)	B (a)		
<i>ad-gair</i> ‘sues, forbids, fascinates’	<i>adobragart</i> ⁷ ‘He sued you.’ (Wb. 19 ^b 5 [prima manus]), gl. <i>uos fascinavit</i>	B (a)		
<i>ad-gládathar</i> ‘addresses’	<i>lase atat gládainn se</i> ‘when I used to address you’ (ML. 62 ^c 16.)		B (e)	
<i>ad-gnin</i> ‘recognizes’	<i>atatgentarsu</i> ‘You will be known.’ (ML. 121 ^d 22)	B (a)		
<i>ad-indnaig</i> ‘leads’	<i>atdomindnastar</i> in <i>ispaniam</i> ‘I shall be brought in <i>Hispaniam</i> .’ (Wb. 7 ^a 5)	B (a)?	B (d)?	
<i>ad-opair</i> ‘sacrifices’	<i>atamroipred</i> ‘I was offered.’ (ML. 44 ^c 17)	B (a)		
<i>at-reig</i> ‘rises’	<i>anatammresa</i> ‘when I will rise’ (ML. 31 ^c 14)			B (h)
<i>fris-oirg</i> ‘injures’	<i>fritumchomartsa</i> ‘I have been offended.’ (Wb. 33 ^a 12)	B (a)		
	<i>cia erat fritammior sa</i> ‘How long will it afflict me?’ (ML. 32 ^d 27)			B (f)?
	<i>is ed aerat fritammiurat</i> ‘[It is] so long [that] will they afflict me.’ (ML. 33 ^a 1)			B (f)
	<i>fritammorcat</i> ‘who injure me’ (ML. 39 ^c 27)			B (f)
	<i>cum</i> ⁸ <i>fritammoircise</i> ‘when you injure me’ (ML. 44 ^b 26)			B (e)
	<i>frisanhi fritammorcat sa</i> ‘against those that afflict me’ (ML. 62 ^c 21)			B (f)

7 Sommer (1897: 190) is probably right when he explains this form as due to a mistake of the glossator (“Wohl Versehen des Schreibers für *atob-* [probably a mistake of the scribe for *atob-*]”).

8 Stokes and Strachan (1901–1910 = Thes. 1: 126, n. m) note that this Latin conjunction stands for Old Irish *intan* ‘when’ or *lase* ‘while’.

Table 3 (continued)

Lemma form	Attested form	I	II	III
	<i>ciofut fritatníarrsu</i> 'How long will he offend You?' (Ml. 93 ^a 15)			B (f)
<i>fris-tét</i> 'answers'	<i>fritumthiagar</i> 'I am answered.' (Sg. 183 ^a 3), gl. <i>obeor</i>	B (a)		
<i>for-brissi</i> 'breaks down'	<i>sechnicoimnactar arnamait són fortanbristis ni</i> 'That is, our enemies have not been able to crush us.' (Ml. 135 ^d 4)		B (d)	
<i>for-cain</i> 'teaches'	<i>isdo fordoncain</i> 'It is for this it teaches us.' (Wb. 31 ^c 16)	B?(b)		
	<i>fortanroichanni</i> 'You have instructed us' (Ml. 22 ^c 3)	B (a)		
	<i>it [hē] fortan roichechnatarni</i> 'It is they that taught us.' (Ml. 63 ^b 1)			B (g)
	<i>aforcital forndobcanar</i> 'the teaching by which ye are taught' (Wb. 3 ^b 23)			C (f)
	<i>fortab cech ansa</i> 'I will teach you (pl.)' (Ml. 53 ^c 14)	B (a)?		
	<i>fordubcechna</i> 'who shall teach you' (Wb. 9 ^a 16), gl. <i>qui uos commonefaciat</i>			C?(f)
<i>for-díuclainn</i> 'devours'	<i>fortamdiucuilset sa</i> 'that they may devour me' (Ml. 44 ^c 32), gl. <i>uorare me</i>		B (d)?	
<i>for-comai</i> 'preserves'	<i>fordomchomaitheir</i> 'I am preserved.' (Sg. 139 ^b 2)	B?(a)		
<i>for-moinethar</i> 'envies'	<i>fordobmoinetar</i> 'They envy you.' (Wb. 19 ^d 27)	B?(a)		
<i>for-tét</i> 'helps'	<i>cofardumthésidse</i> 'so that you may help me' (Wb. 7 ^a 12)	B?(c)		
	<i>fortat tet su</i> 'It helps you.' (Ml. 43 ^b 11)	B (a)		
<i>etar-díben</i> 'destroys'	<i>co etardamdibet sa</i> 'in order that they might destroy me' (Ml. 44 ^c 31)	B?(c)		
	<i>co etardamdibitisse</i> 'in order that they might destroy me' (Ml. 54 ^d 14)	B?(c)		
<i>etar-scara</i> 'separates'	<i>lasse etardanroscarni</i> 'when he has separated us' (Ml. 120 ^a 3)		C? (e)	

information is encoded: the capital letters B and C refer to the infix class used, and the small letter between parentheses indicates the syntactic structure in which it is used. Bearing in mind the possibility of having a declarative instead of a relative verbal form observed in the previous section, the information encoded in those three columns must be read as follows. The presence of B in column (I) is the expected procedure in the syntactic structures concerned. The presence of B or C in column (II) can be considered as a part of the general phenomenon of variation between declarative and relative clause type forms in those syntactic environments. Finally, in column (III), i.e. in the syntactic environments in which relative morphology (in this case, Class C of pronominal infix) is expected, the presence of a Class B form should be considered as parallel to the use of Class A in those situations, as illustrated in examples (6b) and (7b) of the previous section. The question mark after the capital letter indicates that the infix Class (either B or C) is not clear, something which is not rare at all. After the letter between parentheses, the question mark indicates that the syntactic structure involved is not clear. Note also that imperative forms (e.g. the form of the verb *as-rochoili*) are marked as (a), i.e. they are counted along with the declarative forms.

The descriptive problem posed by the forms included in Table 3 is the considerably high amount of Class B non-third person infixes in verbal forms in which relative morphology is expected. To be more precise, in the syntactic structures of Group III, this is the case of 19 cases out of a total of 23 forms; this 80% of unexpected Class B contrasts with the 43% of unexpected Class A observed at the end of the previous section.

The evidence provided by the Old Irish glosses, as it is presented in Table 3, permits us to make finer distinctions, in this case, according to the lexical preverb involved. The verbs in Table 3 are ordered according to whether they distinguish between Classes B and C, such that those lexical compounds which distinguish (more or less frequently) between the two classes precede those which apparently do not. Thus, a fairly frequent verb with the lexical preverb *as-* such as *as-beir* 'says' uses on two occasions the Class C forms of the non-third persons infixed pronouns, the reason being probably the nasalising character of the relative forms involved. In the case of the preverb *con-*, one out of three cases of forms in which relative morphology (i.e. Class C) would be expected shows the form used as Class B. The verbal forms with the preverb *in(d)-* show only one case of Class C infixed pronoun out of four possible forms. The preverbs *ad-* and *fris-* only display Class B forms, regardless of the expected clause type morphology. Finally, the preverbs *for-* and *etar-* also show a considerable degree of confusion between the spellings with *t* and *d*. On the one hand, the forms spelt with *-t-* can be identified as Class B infixes, but some of them (e.g. *fortan*

roichechnatarni ‘who have taught us’) appear in forms in which relative morphology would be expected, whereas the forms with *-d-*, which therefore seem to be Class C forms, appear in syntactic environments in which declarative morphology is undoubtedly expected (i.e. *co etardamdibet sa* ‘in order that they might destroy me’). As usually acknowledged, the spellings *-t-* and *-d-* seem to interchange freely in the cases in which these two preverbs in pre-tonic position are combined with a pronominal infix.

The conclusion seems plausible that non-third persons actually are on the verge of making no distinction between Classes B and C, that is to say, that a good deal of the compounds with preverbs of the type (-)VC- only use one set of non-third person infixed pronouns, regardless of the expected declarative or relative morphology of the verb. The descriptive problem may therefore be formulated in terms of paradigm defectiveness: is the opposition between Classes B and C an actually effective opposition in the non-third persons, so that the rare cases in which an apparently Class C form may be identified are actually a sort of incipient attempt to establish that distinction?

In this descriptive problem it is not convenient to take for granted the existence of a specific differentiation in a particular NP type (in this case, first and second person pronouns) by the mere fact that that differentiation is carried through in other types of NPs (in this case, third person pronouns). The former constitutes a natural class which may show specific inflectional features not observable in the latter. Witness the various cases of lack of formal expression observable in the first and second pronominal elements of not a few ancient Indo-European languages for a grammatical opposition which is formally marked in the remaining NPs, as detailed in García-Castillero (2001). In itself, the assumption of such an asymmetric situation between non-third and third person pronominal markers would not be something objectable.

The position defended in this paper is that such an asymmetric paradigm in the infixed pronouns attached to (-)VC- lexical preverbs (with the exception of *imm-* and *ar-*) must be taken seriously, so that the differentiation between relative and declarative forms is more or less systematic for the third persons, but not for the non-third persons. Table 4 below can be viewed as a complement of Table 2 above as a means of representing more realistically the situation of the (-)VC- lexical preverbs other than *con-* and *in-*, i.e. the situation of *for-*, *etar-*, *fris-*, and *ad-*, which do not distinguish systematically between B and C Classes of non-third person infixed pronouns.

Table 4: The lexical preverbs *to-* (representing CV-) and *for-* (representing (-)VC- preverbs) and the use of Classes A, B and C of infixes in the language of the Old Irish glosses.

	Class A	Class B	Class C	Class C
	<i>to-</i>	<i>for-</i>		<i>to-</i>
1sg.	<i>do-m</i> ^L	<i>for-t/dam</i> ^L		<i>do-dam</i> ^L
2sg.	<i>do-t</i> ^L	<i>for-t/dat</i> ^L		<i>do-dat</i> ^L
3sg.masc.	<i>d(o)-a</i> ^N	<i>for-t</i> ^N	<i>for-id</i> ^N	<i>do-d</i> ^N
3sg.neut.	<i>d(o)-a</i> ^L	<i>for-t</i> ^L	<i>for-id</i> ^L	<i>do-d</i> ^L
3sg.fem.	<i>do-s-</i>	<i>for-t/da-</i>		<i>do-da-</i>
1pl.	<i>do-n-</i>	<i>for-t/dan-</i>		<i>do-dan-</i>
2pl.	<i>do-b-</i>	<i>for-t/dab-</i>		<i>do-dab-</i>
3pl.	<i>do-s</i> ^N	<i>for-t/da</i> ^N		<i>do-da</i> ^N

5 The diachronic problem: Paradigmatic split or merger?

Up to this point, the diachronic perspective has been adopted only on a couple of occasions to consider the morphological processes leading to the formation of some Class A infixes, side by side with a more descriptive account of the same phenomenon. This section introduces the systematic consideration of the origin of the Class B forms, for which a purely descriptive stance has hitherto been adopted.

Whereas Class B clearly represents a problem in this regard, there is great consensus (if not complete agreement) among scholars about the etymology of the Old Irish Classes A and C of infixes.

To begin with Class C, the *-d-* of this set of forms is etymologically the same element as the *-d-* of the negative relative conjunct particle *nad-* mentioned in section 2 above. Especially in the case of the first and second persons, the Class C forms can be analysed straightforwardly as the combination of that *-d-* with the corresponding infix of Class A: e.g. Class C first person singular *-dom-* equals **-d(V)-* + Class A first person singular *-m-* and so on. For the ultimate (i.e. Proto-Indo-European) origin of this Old Irish *-d-* marker associated to relative clause type

marking, namely, the connective clitic **de*, I refer to the observations and references in Watkins (1963: 24) and McCone (2006: 273–276).

As for the Class A forms, I focus on the third person singular masculine / neuter infix(es), which will play an important role in the next section. It is usually assumed that they represent the clitic accusative masculine and neuter forms of the Proto-Indo-European anaphoric stem **e/i-*, namely, the forms **em* and **ed* respectively. For these Proto-Irish and Proto-Insular Celtic forms, see Schrijver (1997b: 54–56). Assuming that the dental plosive of the neuter form was lost in absolute final position, the resulting forms **em* and (**ed* >) **e* explain the nasalising and leniting effects of the corresponding infixes in Old Irish, which have been noted as *-a^{N/L}*. As for the combination with the lexical preverbs with the shape CV-, the process assumed in section 3.3 which involves the elimination of the first of two vowels standing in hiatus seems to be the best diachronic explanation for these forms. This diachronic origin is perfectly compatible with the synchronic interpretation in terms of replacive morphology, as also stated in section 3.3.

It is also worth noting in this section that the exceptional character of the Old Irish lexical preverbs *imm-* and *ar-* is due to the fact that they come from forms which originally ended in a vowel (see GOI §411). This original vocalic auslaut agrees with or directly explains (a) the use of Class A infixed pronouns with the lexical preverbs *imm-* and *ar-* as observed in Table 2 above, (b) the relative clause type forms with these lexical preverbs (i.e. *imme/a-* and *are/a-*), as well as (c) the palatal character of the preverb *ar-* in the tonic position of the prototonic forms.⁹

The preverb noted in Table 1 as **ath(i)*, which appears as *aith-* ‘re-’ in its stressed form, and as *ad-* in the pretonic position, takes Class B infixes and must be assumed, at least for the form used in pretonic position, as a preverb without the final vowel, in spite of its most probable etymology, which has a final vowel. In order to explain the different behaviour of this **ath(i)* with respect to *ar-* and *imm-*, Uhlich (2009–2010: 154) adduces that the former lacks a prepositional counterpart as the reason for the maintenance of the form without vowel **ath-* > Old Irish *ad-* ‘re-’. The lexical preverbs *imm-* and *ar-* had a

⁹ Consider e.g. the imperative *airbir* (*biuth*) ‘consume!’ (Wb. 29^a25), from *ar-beir*, where the stress falls on the first vowel of the preverb; similarly, *nádairchissa* ‘that he spare not’ (Wb. 5^b35), a present subjunctive from the verb *ar-cessi*. In forms like these, the original shape of the preverb may have been both **are-* and **ari-*. I leave this question of the original auslaut of Old Irish *ar-* open.

prepositional counterpart, in which Uhlich assumes that the final vowel would have been maintained and therefrom extended to the preverb.¹⁰

A clear consequence of this diachronic observation on *imm-* and *ar-* is that the phonotactic structure of the lexical preverb involved must be taken as the definitive factor deciding the Class (whether A or B) of the infix pronoun used to express declarative clause type. In order to include *ar-* and *imm-*, the shape hitherto considered as CV- should be reformulated as *(V)CV-, as (*-)CV- or, more simply, as (-)CV-. This is an important argument to be considered in the diachronic discussion in this section, but is not the explanation itself, since it still does not state the reason for the use of Class B infixes after lexical preverbs of the shape (-)VC-.

As for Class B infix pronouns, and in line with the descriptive question formulated in the previous section, the diachronic problem can be formulated as follows: is the virtual lack of distinction between Classes B and C in the non-third persons a remnant of an original situation in which there was actually no such distinction, or is it the outcome of a process in which two originally different paradigms (Classes B and C) are not distinguished anymore, or only scarcely, in the non-third persons? In other words, one must decide between a process of paradigmatic split or a process of paradigmatic merger, respectively.

The diachronic explanation to be developed in the next section assumes that there was originally a single paradigm of forms and that a process of morphological split has given rise to two forms for some elements of the paradigm, in this case, in the third persons, and tentatively in the other persons and for some lexical preverbs.

The opposite view has also been defended, most conspicuously perhaps by Thurneysen (GOI § 455), who relies on the form adopted by the (-)VC- preverbs *con-* and *in(d)-* in the expression of Class B pronominal infixes. Certainly, *cot-* and *at-* can be the phonologically regular outcome of previous sequences such as **kon-t-* and **in-t-* respectively. On the basis of these forms, Thurneysen assumes that Class B is initially the form of the Proto-Indo-European demonstrative stem **so-/to-*, which was initially used to express some third persons, and was later generalised for the remaining persons. The main problem this assumption faces is that it completely lacks a motivation for the use of two different infixes, say, the acc.sg. masc. **em* (from the PIE stem **e/i-*) as the forerunner of the third

¹⁰ Though it is certainly difficult to demonstrate, it may well be the case that *aith-* / *ad-* 're-', due to the fact that it is less frequent than *ar-* and *imm-* and also due to its formal similarity with *ad-* (in line with the argument in section 3.3 on the loss of formal distinctivity of previously different preverbs, especially in pretonic position), has been secondarily attracted to the group of the (-)VC- lexical preverbs.

person singular masculine of Class A, and the acc.sg. masc. **tom*, as the forerunner of the corresponding form of Class B. Since the distribution of Classes A and B is clearly associated to the phonotactic shape of the lexical preverb involved in the pretonic position, as just stated, one should ask why a lexical preverb such as, say, **di/e-* took **-em-*, whereas forms such as **eks-* > **ess-* would have taken **-tom-* to express the same person in exactly the same syntactic context.

For other proposals for the origin of Class B which are located in a wider discussion but which defend two originally different paradigms, see McCone (2006: 229–231).

6 The present hypothesis: The paradigmatic split into B and C

This section develops the hypothesis of a paradigmatic split according to which a unique paradigm, the one which is Class C in Old Irish, split into two different paradigms, Classes B and C. This explanation has three points: (i) the trigger of the split, dealt with in section 6.1; (ii) the specific syntactic context(s) in which the use of Class C in declarative clause type forms was enabled, in section 6.2; and (iii) the very process of paradigmatic split, in section 6.3.

6.1 Watkins' (1962) 'Forward Reconstruction' and the trigger of the split

The starting point of this diachronic explanation is the consideration of the paradigmatic structure in which the [\pm third person singular masculine / neuter pronominal infixes] feature cross-cuts the declarative vs relative clause type opposition, in line with the second issue considered in section 2 above.

The resulting schema of formal oppositions is illustrated in Table 5 with the lexical compound *do-beir* 'brings, gives', the lexical preverb of which has the CV- shape. On the one hand, the declarative and relative forms without infixed pronoun contrast by the lack or presence respectively of relative mutation (nasalisation or lenition) in the first consonant of the tonic part; this is an example of one of the formal strategies in which this clause type opposition is marked. On the other, the contrast between the latter form, i.e. the relative clause type form without pronominal infix, and the declarative form with such an infix is expressed by means of the different vowel of the pretonic preverb, which are *-o-* and *-a-* respectively; this difference has been discussed in section 3.3 above.

Table 5: [\pm 3rd person singular masculine / neuter pronominal infixes] and the difference declarative vs relative clause type in a lexical compound with a (-)CV- preverb.

	Declarative clause type	Relative clause type
[- 3sg. masc./neut. pronominal infixes]	<i>do-beir</i> '(s)he brings'	<i>do.^{N/L}beir</i> 'who(m) / that ([s]he) brings'
[+ 3sg. masc./neut. pronominal infixes]	<i>da.^{N/L}beir</i> '(s)he brings him / it'	<i>dod.^{N/L}beir¹¹</i> 'who / that brings him / it'

The four Old Irish verbal complexes turn out to be formally distinct, subtle and minimal as the difference may be.

The basic diachronic assumption of this proposal is that the Class B of infixed pronouns is the response to a situation in which (some of) the (-)VC- lexical preverbs were not able to make an important distinction, the one between the nasalising and leniting relative forms, on the one hand, and the declarative form including the third person singular masculine / neuter infixed pronouns as initially expected according to the same origin assumable for those infixes in combination with CV- lexical preverbs (i.e. **em* and **e[d]*), on the other. The reason for this formal coincidence was that the latter regularly lost its palatal character¹² in pretonic position, the only remnant of their presence being nasalisation and lenition. This initial situation is reflected in Table 6, in which the declarative clause type form combined with the third person singular masculine / neuter infixed pronoun (i.e. the form **ad.^{N/L}cí* '(s)he sees him / it') is 'forward-reconstructed' in the sense of Watkins (1962: 2–3) and Eska (2003), i.e. is reconstructed as an expected Old Irish form which, however, is not attested. In the same table, the assumed relative form including the corresponding infixed pronoun has the expected outcome for a Class C form, i.e. **at.^{N/L}cí* /ad-/ 'who sees him / it' < **að-de-e(m)*-. This form, which is in Old Irish declarative due to its Class B infixed form, is marked with an asterisk because it is given as a relative form.

As for *as*-, the form with Class A third person singular masculine / neuter pronominal infix would have been **es^{N/L}*- (cf. the Old Irish conjugated preposition *es(s)*, mostly *ass* 'out of it'), depalatalised as **es^{N/L}*- and then **as^{N/L}*-, in

¹¹ Note that the mutations marked as ^{N/L} in the forms of this table do not always have the same function: whereas in *do.^{N/L}beir* they are the relative mutations, in *da.^{N/L}beir* and *dod.^{N/L}beir* they are the mutations provoked by the third person singular masculine and neuter infixes respectively.

¹² This palatal character is the effect of the so-called second or third palatalizations, which assume palatalization by the effect of a vanishing front vowel (McCone 1996: 117,119).

Table 6: The expected homonymy of declarative form with 3rd person singular masculine / neuter pronominal infixes and relative without such pronominal infix in a lexical compound with preverb (-)VC-.

	Declarative clause type	Relative clause type
[- 3sg. masc./neut. pronominal infixes]	<i>ad-cí</i> '(s)he sees'	<i>ad^{N/L}cí</i> 'who(m) / that ([s]he) sees'
[+ 3sg. masc./neut. pronominal infixes]	(* <i>að-e(m)- ></i>) * <i>að^{N/L}</i> > Old Irish * <i>ad^{N/L}cí</i> '(s)he sees him / it'	* <i>at^{N/L}cí</i> 'who / that sees him / it'

much the same way as most cases of the conjugated preposition just quoted. Other (-)VC lexical preverbs would have evolved as (**wor-e[m]- >*) **foir^{N/L-}* (by depalatalisation of unstressed forms > Old Irish **for^{N/L-}*), and (**kon-e[m]- >*) **coin^{N/L-}* (by depalatalisation) > Old Irish **con^{N/L-}*.¹³ In other words, I am assuming here that the palatal character caused by the (Class A) third person singular masculine / neuter infixed pronoun in (-)VC- lexical preverbs would have been lost in pretonic position and that this brought about the complete homonymy of this form and the same lexical preverb followed by relative nasalisation and lenition.

As for the depalatalisation itself, current treatments as e.g. McCone (1996: 135) and Stifter (2009: 62) assume an Early Old Irish process affecting consonants in unstressed words such as the copula, prepositions, particles, etc., and this inevitably implies a relatively recent chronology for the rise of Class B infixed pronouns, not very long before the classical Old Irish period.

The Early Old Irish texts in which some such functional words still show palatal character are the Cambray Homily and the *prima manus* of the Wb. Glosses, both dated approximately between the end of the seventh and the beginning of the eighth centuries. There is a potential problem in that the Cambray Homily includes a case of Class B pronominal infix, precisely the third person singular neuter form (*autrubert* 'has said it' [Thes. 2: 246.14], from *as-beir*). In this sense, one could argue that, if depalatalisation was still a process not accomplished in these Early Old Irish texts, then the creation of the Class B infixed form in the manner just assumed could hardly have happened. However, the

¹³ The specific situation of the lexical preverb *ete/ir-* is too complicated to be considered in this paper.

forms which still show palatalisation are of a different nature to the lexical preverbs of verbal complexes, as is perhaps suggested by Thurneysen (GOI §168). They are independent conjunctions which still show their original palatal character: *amail* ‘as’ (Thes. 2: 245.14) for later *amal*, *oire* ‘because’ (Thes. 2: 246.5–6) for later *(h)óire*, *(h)uare* (as in e.g. (3), (4) and (5) above). In the case of these and other conjunctions, the depalatalisation is probably a process to be assumed for later phases, as seems to be the case of *intain* ‘when’ (Thes. 2: 247.3), which in Wb. appears mostly as *intain* and in Ml. and Sg. mostly as *intan*; or even *air* ‘for’ (Thes. 2: 245.33), which later appears both as *air*, as in example (2a), and as *ar*. These differences are probably due to the different chronologies of the grammaticalisation processes leading to their character of conjunction, which is probably accompanied by a difference in stress. With respect to the independent conjunction *air* ‘for’, the corresponding stressed conjugated form *aire* ‘for him, for it’ appears in this manner both in the Cambray Homily (cf. Thes. 2: 244.33) and in classical Old Irish.¹⁴

None of the unstressed forms quoted in the previous paragraph, however, are constitutive components of the Old Irish verbal complex, and the process of depalatalisation assumed at this moment is the one affecting lexical preverbs located in the pretonic position of this morphological structure. In this sense, the Cambray Homily already has the form *ma arfoimam* (MS *maar foim am*) ‘if we receive’ (Thes. 2: 245.12) from *ar·foim*, with the lexical preverb *ar-* in pretonic position also found in later Old Irish texts. As stated in the previous section, the declarative form of the preverb *ar-* with third person singular masculine / neuter (i.e. *are/a^{N/L-}*) and the conjugated preposition *aire* ‘for him, for it, therefore’ are best explained from a previous form **ari-*, whereas the conjunction *air* ‘for’ and the bare pretonic form of the lexical preverb (as *ar-* in the previous form of the Cambray Homily) could be derived from both **ari-* and **are-*. The expected palatalisation in the pretonic form of the lexical preverb in its declarative clause type form without infixed pronoun has suffered the process of depalatalisation of unstressed forms now under consideration.

¹⁴ To conclude the treatment of the so-called Early Old Irish attestation, the combination of the preposition *ar-* with the oblique relative conjunct particle *-(s)a^{N-}* in *aire sechthar* ‘that he follow’ (Thes. 2: 244.31) seems to maintain the palatal character of the pretonic sequence, but the same sequence shows no palatal character a couple of lines before in the same text: *ara tinóla* ‘that he gather’, *are n-airema* ‘that he receive’ (Thes. 2: 244.27–28), from *do-inóla* and *ar-eim* respectively. This conjunct particle, which is the outcome of a relatively recent process of internalisation of a previously autonomous sequence of preposition and demonstrative **sa^{N-}* (see the observations in García-Castillero 2018), may well have preserved some features of its previous situation.

The important aspect for the question of the depalatalisation in preverb forms with Class A third person singular masculine / neuter infixes such as (**að-e[m]*- >) **að^j.^{N/L}*- > Old Irish **ad^{N/L}*- and (**ess-e[m]*- >) **es^{N/L}*- > Old Irish **as^{N/L}*-, as well as **foir^{N/L}*- > Old Irish **for^{N/L}*- is that its most adequate parallel is the form of the bare preverb *ar*- just considered, i.e. the form of the preverb used in the declarative clause type form without infixes, and this form already shows the depalatalised form in Early Old Irish, i.e. **ari/e*- > **ar^f*- > *ar*- (in *ar-foimam*). The depalatalisation of unstressed elements can therefore be considered a relatively long process which takes place before and after the Early Old Irish period, but the crucial change assumed above in Table 6 for the lexical preverbs with the shape (-)VC- seems to have happened well before that time.

It is to be said that the homonymy between forms which can be included in the previous table is not always avoided. The two following cases can be mentioned. First, most lexical compounds with the deuterotonic shapes CV·VC(-) and (-)VC·VC(-) make no systematic distinction between declarative and leniting relative clause type forms without infixes, so that e.g. *do-adbat* may be '(s) he shows' and 'who shows' or 'whom/which (s)he shows'; *ad-aig*, '(s)he drives' and 'who drives'. Second, verbs with the lexical preverb *ar*- have the same form for the relative clause type form and the declarative with third person singular masculine / neuter infixes (e.g. *ara^{L/N}gaib* both 'who seizes' and '(s)he seizes it / him'). Note that the second case is the one which is being considered for the lexical preverbs taking Class B infixes.

In other cases, however, a more marked verbal complex is used to express the meaning of another form in which a phonological process has caused a certain disturbance in the distinction of the categories included in Tables 5 and 6 above. I refer to the case of the verb *fo-fera* 'produces, causes' included in Table 7. In this verb, the regular form of the relative without pronominal infix is *fo-era*,¹⁵ but this combination is also expressed by means of the relative form with pronominal infix: the form *fodera* must be interpreted simply as 'which causes' in Wb. 3^c33, Wb. 3^c34, Wb. 5^d5, Wb. 14^c42, Ml. 93^a13, Ml. 55^d11, and as 'which causes it' in Wb. 33^c12, Ml. 32^a5. The same use of the Class C infix deprived of any referentiality can be assumed for *dodesta* 'what is lacking', the relative form from the CV·VC(-) verb *do-esta* 'is lacking' which is used after the light heads *a^N* and *aní* 'that (what)'.¹⁵

¹⁵ This form can be deduced from the past subjunctive *fuerad* 'that [Joshua] provided' (Wb. 33^b13). The lack of *-f*- in the leniting relative form *fo-era* is the regular outcome of the lenition of /f/.

Table 7: [\pm 3rd person singular neuter pronominal infix] and the difference declarative vs. relative clause type in the lexical compound *fo-fera* ‘causes’.

	Declarative clause type	Leniting relative clause type
[- 3sg. neut. pronominal infix]	<i>fo-fera</i> ‘(s)he causes’	<i>fo-era</i> / <i>fo-d-era</i> ‘who causes’
[+ 3sg. neut. pronominal infix]	<i>fa-era</i> ‘(s)he causes it’	<i>fo-d-era</i> \uparrow ‘who / that causes it’

Turning to the situation of homonymy resulting in Table 6 above, the remedy was to make use of the most marked form, i.e. the form originally used to express these infixes in a relative clause type verb, in order to express also the pronoun of the declarative clause type verb. This first step in the process leading to the creation of Class B infixes is illustrated in Table 8.

Table 8: [\pm 3rd person singular masculine / neuter pronominal infixes] and the difference declarative vs. relative clause type in a lexical compound with preverb (-)VC-.

	Declarative clause type	Relative clause type
[- 3sg. masc./neut. pronominal infixes]	<i>ad-cí</i> ‘(s)he sees’	<i>ad^{N/L}cí</i> ‘who(m) / that ([s]he) sees’
[+ 3sg. masc./neut. pronominal infixes]	* <i>ad</i> ^{N/L} <i>et at^{N/L}cí</i> \leftarrow ‘(s)he sees him / it’	(* <i>at^{N/L}cí</i>) ‘who / that sees him / it’

In spite of the different use of the relative form including the object pronominal reference, both processes assumed in Tables 7 and 8 share two remarkable features. First, in the initial situation, one of the expected forms turns out to be problematic, due either to its inherent structure (the hiatus resulting in the form *fo-era*), or to its lack of differentiation from another form (*ad^{N/L}cí* as ‘who-[m] [(s)he] sees’ and ‘(s)he sees him / it’). Second, the clearer (in the sense of formally more perceptible) form with infixes of Class C comes to the rescue in both cases: on the one hand, the relative form with neuter infixes (i.e. with Class C form) *fo-d-era* ‘who causes it’ lacks the phonotactically uncomfortable hiatus of the attested form *fo-era* and is used to express (also) the bare relative form ‘who causes’; on the other, the assumed relative form with pronominal infix *at^{N/L}cí* ‘who sees him / it’ has the advantage that it shows clearly the form of a pronominal infix and is used to express the declarative form with pronominal infix ‘(s)he sees him / it’.

In both cases, of course, it is necessary to find a syntactic environment in which the use of the surrogate form can be justified.

6.2 The bridging context

In the previous section, the relative form with pronominal infix of both *fo·fera* and compounds with a lexical preverb ending in consonant such as *ad·cí* has been assumed to be the surrogate form for other forms which became problematic from the point of view of their morphological distinctiveness. The portmanteau morpheme which is the Class C infixed pronoun, expressing both person/number and relative clause type character, is used to express only one of these two categories.

The specific syntactic environment and pronominal forms in which those two changes may have taken place constitute the ‘bridging context’ defined by Heine (2002: 86) as “a specific context giving rise to an inference in favor of a new meaning”.¹⁶ As has been made clear in the previous section, the differences and similarities between the cases of the verb *fo·fera* and of the verbs with a lexical preverb ending in a consonant must be clearly stated.

As for the use of *fodera* as the relative form of *fo·fera*, the most plausible scenario is provided by the cataphoric use of the third person singular neuter infix, a context in which the meaning of the pronominal marker can be lost. This cataphoric use of the Old Irish pronominal elements attached to the verbal complex has been studied by Lucht (1994) and, more recently, by Eska (2010). An example among many others is *rafoiligestar* (i.e. *r[o]-a^(L)·foiligestar*) in (8), in which the third person singular neuter infix *-a^(L)-* cataphorically, i.e. proleptically (see GOI § 421), refers to the object NP introduced by the light head *a^N* (*anadfiadar is indsalml so*) appearing later in the same sentence.

- (8) *rafoiligestar* *nathan* *duduid*
 AUG-3SG_{NEUT}·reveal_{3SG.PRET} Nathan_{NOM} to=David_{DAT}
anadfiadar *is* *indsalm* *so*
 that^{-NAS}PV·relate_{3SG.PRES.REL.PASS} in the_{DAT.SG.MASC}=psalm_{DAT} PROX
 ‘Nathan revealed to David what is related in this psalm.’
 (lit. ‘Nathan revealed it to David, that which is related in this psalm’)
 (ML. 109^{b2})

¹⁶ The “bridging context”, which Heine assumed to be a step in the process of grammaticalisation, should therefore also be considered in the process of paradigmatic split as the environment in which a functional change is possible.

As for the use of a Class C pronominal infix, i.e. of an infix which is initially expected in a relative clause type form, to express the same pronominal reference in a situation in which a declarative clause type form could be expected, the most adequate bridging context is in the syntactic structures included above in Group II, since complement clauses and adverbial subordinates introduced by conjunctions such as *amal* ‘as’ or (*h*)*óre* ‘because’ show in the Old Irish glosses a certain degree of variation between declarative and relative clause type. The possibility of variation in those syntactic contexts basically involves the use of Class A instead of Class C, and this would have been interpreted as a neutralisation of that difference.

An example of this situation is given in (9), where the third person plural infixed pronoun of Class A *-s-* observed in Table 2 above (i.e. a declarative clause type marker) is used in the verb after the subordinating conjunction *hóre* ‘because’, which is often accompanied by a nasalising relative verb. For this use, see Ó hUiginn (1986).

- (9) *hóre nosmóidet iprecept*
 because PV-3PL(A)·boast_{3PL.PRES} in=preaching_{DAT}
 ‘because they boast in preaching, [. . .]’ (Wb. 17^c5)

The form *fritracatar* quoted in (4b) above and repeated here in (10), from the verb *fris-accai*, is especially significant at this point. This example was adduced in section 3.4 as a case in which declarative morphology is used in a syntactic context in which relative morphology can also appear. In the light of the present diachronic hypothesis, the Class B third person singular neuter infixed pronoun in *fritracatar* may well be a case in which the original Class C form added to the original form of the preverb involved (i.e. **wriθ-* + /-ð^L-/ > /*frid*^L-/, spelt as *frit-*) was liable to an interpretation as a declarative clause type marker, i.e. in a place in which the declarative clause type Class A infix can also appear.

- (10) [. . .] *huare fritracatar som a deo*
 . . . because PV-3SG_{NEUT}(B)·hope_{AUG.3PL.PRET=3SG_{NEUT}} from God_{ABL}
 ‘[. . .] because they have hoped for it a *deo*’ (ML. 131^c10)

Most cases with first and second person infixed pronoun included in the column (II) of Table 3 perfectly represent the syntactic contexts in which this use of the third person singular masculine / neuter infixed pronouns of Class C

would be reinterpreted as markers of declarative clause type due to their clearer form. In most cases of the column (III) of Table 3, the interpretation of the Class B infixed form as coming from a Class C form makes perfect sense: e.g. the form *anattammresa* ‘when I will rise’ included in Table 3, from the verb *at-reig*, appears in a structure in which relative morphology is otherwise compulsory, and this agrees with its diachronic interpretation as an originally Class C form.

Up to this point, the explanations for the relative form of *fo-fera* and for the rise of Class B infixed pronouns have run in quite a parallel fashion. They are different, however, in that in the first case, only one single pronominal element is involved, the third person singular neuter infixed pronoun, and one single lexical element. In the case of Class B of infixed pronouns, the whole pronominal paradigm comes into play and, in addition to that, there are quite a number of compound verbs involved, i.e. those which had a (-)VC- lexical preverb in the pretonic position.

6.3 The mechanism of paradigmatic split

It is time to delineate the morphological process by means of which two paradigms, i.e. Classes B and C of pronominal infixes, arise from a single one, i.e. Class C. For this purpose, the well-known case of paradigmatic split in which two Latin nouns, *deus, deī* ‘god’ and *dīuus, dīuī* ‘deity, divine’, have developed out of a single original paradigm, the one included in step (i) of Figure 1, may serve to illustrate the basic mechanism of this morphological change. The changes that occurred from step (i) to step (iib) of Figure 1 are of a phonological nature and trigger the later process of paradigmatic split. First, step (iia) shows the effect of the regular changes */wo-/ > /o-/* and */ej-/ > /-ē-/*; forms with and without */w-/* are thus created in the paradigm. In step (iib), the change from closed */-ē-/* to */-ī-/*, which happens in forms such as the genitive and ablative singular but is prevented in other forms such as the nominative and accusative singular due to the shortening of that */-ē-/* in prevocalic position, causes a still clearer differentiation between those two groups of forms. Properly, the split is in step (iii), where each original group of forms analogically creates the missing parts of their paradigms.

The change from (iib) to (iii) in Figure 1 is graphically shown as the substitution of the horizontal by a vertical line, and this is an appropriate representation of the creation of two different paradigms from a situation in which two variants within one and the same paradigm were created.

I therefore assume two main steps for the process of paradigmatic split, first the introduction of some sort of variation within a given paradigm (ii) in

	(i)	→	(iia)	→	(iib)	→	(iii)
nom.sg.	* <i>dejwos</i>		> * <i>dēos</i>		> * <i>deos</i>		> <i>deus</i> <i>dīuus</i>
acc.sg.	* <i>dejwom</i>		> * <i>dēom</i>		> * <i>deom</i>		> <i>deum</i> ↓ <i>dīuum</i>
gen.sg.	* <i>dejwī</i>		> * <i>dēwī</i>		> * <i>dīwī</i>		<i>deī</i> > <i>dīuī</i> ↑
abl.sg.	* <i>dejwōd</i>		> * <i>dēwōd</i>		> * <i>dīwōd</i>		<i>deō</i> > <i>dīuō</i>

Figure 1: The process of paradigmatic split leading to Latin *deus*, *deī* ‘god’ and *dīuus*, *dīuī* ‘deity’.

Figure 1), and second the analogical creation of new forms corresponding to each of the original variants, thus giving rise to two different paradigms ((iii) in Figure 1), the existence of which must be justified on the basis of some functional or semantic difference, the basic requisite for any process of split in diachronic morphology (see García-Castillero 2013).

Accordingly, the paradigmatic split assumed for the creation of Class B as a distinct paradigm from the original Class C is delineated in Figure 2, where the lexical preverbs *aith-* and *uss-* have been omitted. Note that the steps (ii) and (iii) include the forms in their usual Old Irish spelling.

	(i)	→	(ii)	→	(iiia)	→	(iiib)
	Class C		Class C		with non-3rd persons		with 3sg masc./neut.
					Class B Class C		Class B Class C
<i>con-</i>	* <i>con-d-</i>		> <i>con-d-</i>		<i>cot-am-</i> <i>con-d-am-</i>		<i>cot-</i> <i>con-d</i> ¹
<i>in-</i>	* <i>in-d-</i>		> <i>in-d-</i>		<i>at-am-</i> <i>in-d-am-</i>		<i>at-</i> <i>in-d-</i>
<i>ad-</i>	* <i>ad-d-</i>		> <i>at- /ad-/</i>		<i>at-am-</i> ↑		<i>at-</i> <i>ad-id-</i>
<i>ess-</i>	* <i>as-d-</i>		→ <i>at-</i>		<i>at-am-</i>		<i>at-</i> <i>as-id-</i>
<i>fris-</i>	* <i>friθ-d-</i>		> <i>frit-</i>		<i>frit-am-</i>		<i>frit-</i> <i>fris-id-</i>
<i>for-</i>	* <i>for-d-</i>		> <i>fort/d-</i>		<i>fort/d-am-</i>		<i>fort/d-</i> <i>for-id-</i>
<i>etar-</i>	* <i>etar-d-</i>		> <i>etart/d-</i>		<i>etart/d-am-</i>		<i>etart/d-</i> <i>etar-id-</i>

Figure 2: Paradigmatic split from original Class C to Old Irish Classes B and C.

¹This lexical preverb *con-* combined in Old Irish with the Class C 3rd person singular masculine / neuter infix often appears as *conid*^{N/L} (e.g. Ml 106^{b8} *lasse conidreip* ‘when he has entrusted himself’, from *con-erbaí*). Though the consideration of this form *conid-* should also take into account the homonymous forms of the conjunct particle *co*^N ‘so that’ plus the same infix and of this conjunct particle combined with the copula, it seems that the analogical influence of the other forms with renewed Class C 3rd person singular masculine / neuter form *-id-* (i.e. *ad-id-* and so on) suffices to explain the form *conid-* in this combination of lexical preverb plus Class C infix. I owe this observation to Elliott Lash (p.c.).

The forms given in step (i) of Figure 2 are those ones expected for Class C, which is characterised by the addition of the (already) relative marker **-dV-* to

the bare form of the lexical preverb followed by the corresponding affixal pronoun. At this stage, there is no need to differentiate between non-third and third persons, and the regular developments assumed for each form arrived at the situation in step (ii) of Figure 2, with more or less transparent forms such as *con-d-*, *in-d-* and *for-t-*, *etar-t-*, with other forms which are most easily explained as due to the fusion of the final dental with the initial dental of the infix (*ad-*, *aith-*, *fris-* in so far as this is from **friθ-*), but also with forms in which the final consonant of the lexical preverb was apparently substituted by the form of the infixed pronoun beginning with *-t- /d/* (as in the case of *ess-*, *oss-* and – taken at its face value – *fris-*). The substitution of the final consonant which can be assumed for **ess-* > *as-* → *a(s)-t-* (and, *mutatis mutandis*, for *oss-* and *friss-*), would be of course a further case of replacive morphology, similar to that assumed in section 3.3 above for the third person singular masc./neut. forms of Class A with preverbs such as *do-*.

On the basis of that situation, step (iiia) of Figure 2 represents the first move towards the differentiation between Classes B and C for (-)VC- lexical preverbs, and corresponds to McCone's explanation below of the Class B forms *cot-* and *at-*, from *con-* and *in-* respectively.

It has, of course, long been realised that *con-d(-)* and *in-d(-)*, which actually do occur as class C forms in relative clauses, would be the regular outcome of the sequences **kom-de-* and **in(de)-de-* in main clauses too. The simple solution is to posit analogical creation of main-clause *co-t(-)* and *a-t(-)* with loss of the preverb's final consonant as in most other cases such as *a-t(-)* < **ad-de-* or **ey-de-* < **ey(z)-de-* in relation to *ad-* and *as-* (< **ess* < **exs*) respectively or *fri-t-* < **wrid-de-* < **wrid(z)-de-* in relation to *fris* (< **writs*).

(McCone 2006: 229)

Apart from being a step in the development of a new distinction between Class B and Class C, the situation of step (iiia) in Figure 2, which includes the first person singular pronominal infix, reflects quite faithfully the description in section 4 for non-third person infixes of Class B, at least for most of the corresponding lexical preverbs. With respect to that situation, step (iiib) of Figure 2 represents the introduction of new forms for distinguishing Class C forms for the third persons, i.e. the creation of new relative clause type forms for the third person singular pronouns, in which the distinction between declarative and relative is in general more systematic.

The whole process of paradigmatic split may therefore be viewed in the change from step (ii) to step (iiib) in Figure 2: the horizontal line in step (ii) is partially put in the vertical position in step (iiia), as a consequence of the analogical creation of *co-t(-)* and *a-t(-)*, whereas the remaining horizontal part of that line ends up in the vertical position in step (iiib), as a consequence of the analogical creation of the forms for the third persons in the other lexical preverbs.

7 Conclusion

The diachronic explanation for the Old Irish Class B of infixes argued for here is different to previous ones in some important respects. First, it pays special attention to the use of these forms in the contemporaneous Old Irish texts. Second, much in line with the basic tenets in García-Castillero (2015), it considers the interaction between phonological, morphophonological, morphological, and also syntactic aspects of the Old Irish verbal complex: in particular, it takes seriously the assumable phonotactic conditions of some grammatical distinctions such as clause type and pronominal references in lexical compounds with a (-)VC- lexical preverb, and how the morphologically undesirable consequences of some phonological changes can be avoided.

The starting point of the diachronic explanation put forward in this chapter is a situation in which there were only Classes A and C. The forms of the third person singular masculine / neuter infixes of Class C were then used to express those persons when their Class A version had been obliterated by regular phonological changes with most lexical preverbs of the shape (-)VC-. The use of Class C instead of the vanished marking attributable to the original forms of Class A was facilitated in syntactic contexts in which both relative and declarative clause type morphology were possible, and the new ‘declarative’ forms were levelled through the whole paradigm, a process especially easy for the non-third persons, since the verbal complexes including those infixes seem to be less in need to distinguish between relative and declarative clause type forms, at least in view of the frequent use of Class A instead of expected Class C. This step in the development of Class B corresponds to the situation assumed for the non-third persons. In the next step, the third person infixes created a form different to the newly created Class B by adding the ending *-id-* to the bare form of the lexical preverb, thus renewing the form of Class C in those third persons.

This diachronic explanation does not need to go far back in the prehistory of the Irish language in order to explain the origin of Class B infixes and, in fact, it nicely fits in with the following descriptive issues observed in the language of the Old Irish glosses. First, it seems clear that the relevant factor for the use of Class B of infixes is the phonotactic structure of the involved lexical preverb in pretonic position, namely the structure (-)VC-. Second, this diachronic explanation also agrees with the default character of the infixes of Classes A and C, which are used not only with (-)CV-preverbs, but also with the remaining conjunct particles. Third, other situations of homonymy which happen in the case of some specific verbs and under some specific circumstances (e.g. *fo-fera* ‘causes’) are sometimes

corrected by using the more visible form. Fourth, the assumption that the third person singular masculine / neuter infixed pronouns constitute the locus of the whole change accords well with the fact that these forms are the most frequent infixes in the contemporaneous Old Irish texts, as clearly observable in Sommer's (1897) collection of forms. Fifth, it directly explains the remarkable asymmetric situation found in the language of the glosses in the use of the non-third persons of the Class B infixed pronouns, which turns out to be a specific step in the assumed process of paradigmatic split; this asymmetry agrees with general trends in the distribution of declarative and relative clause type marking in Old Irish.

As a general result of this study, the consideration of the process of paradigm split has also revealed the need of a 'bridging context' for this change, which is also essential in grammaticalisation processes. This is surely not a matter of chance. In fact, grammaticalisation is one of the possible sources of a new morpheme in a given paradigm, or of a new paradigm, and the consequence of the creation of these new morphological elements is that there is a morphological split so that the new morpheme or paradigm expresses a specific function or meaning, different to the meanings of the morphemes already existing in that paradigm or to the meanings of the already existing paradigm(s).

Acknowledgement: This is a revised version of a paper presented at the *Teangeolaíocht na Gaeilge / The Linguistics of the Gaelic Languages XVII*, held at the National University of Ireland Maynooth on 10–11 April 2015. I primarily have to express my gratitude to the editors for having accepted it in this volume. I am also indebted to the audience of the 2015 conference, in particular Jürgen Uhlich, for their comments. Elliott Lash, Fangzhe Qiu and an anonymous reviewer have also made important contributions to improve the quality of the paper. Remaining errors are my sole responsibility. Financial support has been received from the research group GIC 10/83, IT 486-10.

Elisa Roma

7 Nasalisation after inflected nominals in the Old Irish glosses: Evidence for variation and change

In this chapter, I discuss the variation in the occurrence of nasalisation on demonstratives, adjectives, nouns and inflected prepositions or adverbs following inflected nouns and adjectives in the Old Irish Würzburg, Turin, St. Gall and Milan Glosses (henceforth Wb., Tur., Sg. and Ml. respectively; see section 1.1 below for the use of edited sources). In these contexts nasalisation is more irregular (at least in spelling) than after proclitics such as articles, possessives and prepositions, and apparently unpredictable.

The data I presented at the Colloquium ‘Variation and Change in the Syntax and Morphology of Medieval Celtic Languages’ (Maynooth, 13 October 2017) were at the time not published yet and had not been presented before. Since in the meantime they have been published in Roma (2018a), this article will not discuss them in detail but only report examples and data for the sake of the argument and dwell on some of their diachronic and synchronic implications.

The paper is organised as follows: section 1 illustrates the contexts taken into account and summarises the data, according to a broad classification of phonetic and syntactic environments (sections 1.2 and 1.3 respectively). Section 2 discusses the data from a distinctive diachronic vs. synchronic perspective. Section 3 sums up the content of the paper and its tentative conclusions.

1 Nasalisation after inflected nominals in the Old Irish glosses: The data

1.1 Sources

The data presented in Roma (2018a), which form the basis of the analysis provided in this chapter, were collected as follows. Instances from Ml. were extracted from Griffith and Stifter (2013), checking all nominal and adjectival entries in the dictionary, while data from Wb., Tur. and Sg. were gathered from

Thes. and cross-checked with Kavanagh (2001), Bauer (2015)¹ and Lash (2018)² for Wb., Sg. and Tur. respectively.

1.2 Phonetic environments

The occurrences of nasalisation or lack thereof were grouped according to a broad classification of possible phonetic and syntactic environments. This was devised expanding on the classification and the results of Thurneysen's (1905) survey, which was the most comprehensive available. Before listing the phonetic environments, a caveat is in order. Nasalisation is expected to be noted in Old Irish spelling only on voiced plosives and on vowels: these were the cases taken into account in the collection of data, in search for a measure of irregularity, which points to variation (see Ó Maolalaigh 2008: 242). Methodologically, therefore, nasalisation as a phonetic or even morphophonological feature cannot be the starting point, as with present-day varieties, but is the demonstrandum.

The phonetic environments are listed below with illustrative examples. The relevant nasalisation marker is highlighted in bold when examples are quoted in the body of the text.

The first kind of environment groups cases where nasalisation would occur after or on a vowel, as in (1). This phonetic environment, where Thurneysen (1905) found that nasalisation was more frequent, includes cases where the triggering word ends in a vowel, as in (1d) and (11), and cases where the nasalised word begins with a vowel, as in (1a)-(1c) and (8), as well as instances where nasalisation occurs between vowels, as in (1e) and (9). Note that instances after a final <n> belong to a separate category, see (2) below. The examples in (1a) to (1e) show nasalisation occurring on different word classes (or nominal case), i.e. on a noun after an agreeing adjective, on an adjective after a head noun, on a demonstrative after a head noun, on a noun in the genitive case, on an inflected preposition (or adverbial), respectively. For all the other phonetic environments listed in this section only a single example will be given, but note that word class of the nasalised word has always been taken into account in the classification and is relevant to the frequency of nasalisation (see section

¹ Bauer (2015) was not available yet when the collection of data from Sg. began. Bauer, Hofman, and Moran's (2018) digital edition of the St. Gall Glosses has also been occasionally consulted.

² I warmly thank Elliott Lash for allowing me to consult the *Minor Glosses Database* before its publication in CorPH.

1.3 below and Roma 2018a: 13). See section 1.3 for details about the classification according to the triggering word.

- (1) a. *arnach* *nindocbáil* *móir*
 for-any_{ACC.SG.FEM} ^{NAS}glory_{ACC} great_{ACC.SG.FEM}
 ‘for some great glory’ (Wb. 23^b12)
- b. *tiagait* *báas* *nanapaig*
 go_{3PL.PRES} death_{ACC} ^{NAS}premature_{ACC.SG.NEUT}
 ‘They go to premature death.’ (Wb. 11^d12)
- c. *ambas* *nisin*
 the_{NOM.SG.NEUT}=death_{NOM} ^{NAS}DEICT_that_{DAT}
 ‘that death’ (Wb. 15^d12)
- d. *ácenele* *nídoine*
 the_{ACC.SG.NEUT}=race_{ACC} ^{NAS}men_{GEN.PL}
 ‘the human race’ (Wb. 5⁵16)
- e. *triguidi* *náiriam*
 through=beseeking_{ACC} ^{NAS}for_{1SG}
 ‘through beseeching for me.’ (Wb. 7^a12)

The second kind groups cases where nasalisation would occur after words ending with *n* before words beginning with a vowel, *d* or *g*, and words ending with <*m*> before words beginning with *b*: in these contexts the lack of a second nasal on the following word may not suggest lack of nasalisation at all (but see section 2 for a discussion of the treatment of this context in Sg.). An example is in (2), where nasalisation occurs on a genitive noun.

- (2) *cotíchtin* *nancríst*
 until=coming_{ACC} ^{NAS}Antichrist_{GEN}
 ‘until the coming of the Antichrist’ (Wb. 25^d1)

The third kind is defined as follows: nasalisation between dental consonants, i.e. all consonants which do not include a labial or velar or two plosives. When both consonants are plosives, the examples have been classified separately (see [5] below; in [3] final <*d*> represents a dental fricative). In (3) nasalisation between dental consonants occurs on a genitive noun.

- (3) *rad* *ndé*
 grace_{NOM} ^{NAS}God_{GEN}
 ‘the grace of God’ (Wb. 7^d3)

The fourth kind groups cases where nasalisation would occur between (non-dental) plosives, as in (4), where nasalisation occurs on an adjective.

- (4) *arná coscrad indeseircc mbráthardi*
 that-NEG·destroy_{3SG.PRES.SUBJ} the_{ACC.SG.FEM}=love_{ACC} ^{NAS}brotherly_{ACC.SG.FEM}
 ‘lest it should destroy the brotherly love’ (Wb. 10^c1)

The fifth environment groups instances where nasalisation is expected between dental plosives, as in (5), where nasalisation occurs on an inflected preposition.

- (5) *suidigfith dia recht ndo*
 establish_{3SG.FUT} God_{NOM} law_{ACC} ^{NAS}to_{3SG.MASC}
 ‘God will establish a law for him.’ (Ml. 46^c20)

Lastly, nasalisation may occur between all other consonants, as in (6), where nasalisation occurs on a genitive noun.

- (6) *isdered mbetho inso*
 is_{3SG.PRES}=end_{NOM} ^{NAS}world_{GEN} the_{NOM.SG}=this_{NOM}
 ‘This is the end of the world.’ (Wb. 10^b3)

The results of the survey are summarised below. I refer to Roma (2018a) for quantitative data.

In Wb. nasalisation is regular only on agreeing nouns, adjectives and on demonstratives, where it mostly occurs after/on a vowel (context as in [1] above). It tends not to be noted after nasals (context as in [2] above) on any following word. On genitive nouns it occurs frequently after/on vowels, and is attested in all interconsonantal environments except between dental plosives. On prepositions nasalisation is rarely found, and never in any interconsonantal position.

In Sg. nasalisation after/on a vowel is always found on agreeing nouns, adjectives and on demonstratives; it is regular on genitive nouns in any phonetic context, while on inflected prepositions it is much more frequent than in Wb., but hardly attested between consonants and never between dental consonants (12 instances without). It tends to be regularly noted after final nasals, and a nasal also appears sometimes when the following word beginning with a vowel is spelt with an initial <h>, as in (7).³

³ Three instances in Sg. (see Roma 2018a: 5). This cannot have any connection with the emergence of a voiceless glottal fricative after or between nasal vowels, in the position once occupied by a nasal consonant, noted by Ó Maolalaigh (2003: 117) for modern Gaelic dialects. See Schrijver (1997a: 219–220).

- (7) *ni fail chumscugud n̄huirdd and*
 NEG·be_{3SG.PRES} change_{ACC} ^{NAS}order_{GEN} in_{3SG.NEUT.DAT}
 ‘There is no change of order there.’ (Sg. 215^{a2})

In Ml. nasalisation occurs in all phonetic environments. Although between dental consonants it tends to be avoided, between non-dental plosives and between other consonants it is attested in a good number of instances on genitive nouns (9/19) and prepositions (10/19). Nevertheless, it is hardly spelt after final nasals on genitives and prepositions.

In Tur. the examples are few, but nasalisation is regular and there is one instance of nasalisation between plosives (*echtar comairbirt m̄biuth pecthæ* ‘outside the practice of the sins’ [Tur. 108]).

The survey presented in Roma (2018a) and summarised here confirms Thurneysen’s observation (1905: 1) that nasalisation is never marked on an initial *d-* after a final *-l* (18 instances in Wb., 10 in Ml., 1 in Sg.), or between a final *-m* (fricative) and an initial *g-* (only occurring in 2 instances in Wb.). While for the latter context the evidence is too meagre to allow any conclusion, for the former it could be argued that a nasal consonant was avoided (and, possibly, not admitted in some varieties).⁴ Nevertheless, counterexamples apparently occur outside my corpus: *iarthimcul n̄dí* ‘after the circuit by it (the sun)’ (Thes. 2: 33.22 [Vienna Bede 23]; on an inflected preposition), where however nasalisation is unexpected after a dative singular noun; *frisinnaraim n̄ grecdi* ‘to the Greek number’ (Thes. 2: 34.28 [Vienna Bede 31]; on an adjective).⁵ Be that as it may, in other interconsonantal contexts the syntactic environment seems crucial: for example, in Wb. there are 3 instances of nasalisation on adjectives out of 8 expected in interconsonantal position (*-s mb-*, *-cc mb-*, *-cc ng-*), but there are none on a preposition in any interconsonantal environment (0/97).

1.3 Syntactic environments

The syntactic environments where nasalisation occurs after inflected nominals have been grouped according to case, gender and number of the triggering word as well as according to word class of the nasalised word (see above section 1.2). They are listed below.

⁴ The nasal in *amal n̄dondfoirde* ‘as signifies it’ (Sg. 26^{b12}), in a different syntactic context, is unusual, and Thes., followed by Bauer (2015), suggests correcting it to *dondfoirde*.

⁵ These are the only two instances of nasalisation in these phonetic environments in the *Minor Glosses Database* (Lash 2018). Lenition after *iar* in Vienna Bede 23 is also irregular.

Nasalisation after a singular nominative neuter noun is exemplified in (1c), (3) and (6). Besides the five word classes listed in section 1.2 and exemplified in (1a) to (1e) respectively, it also occurs once on an agreeing noun in apposition, as in (8).

- (8) *sliab* *nossa*
 mount_{NOM} ^{NAS}Ossa_{NOM}
 ‘Mount Ossa’ (Sg. 63^a16)

Nasalisation after an accusative noun is exemplified in (1a), (1b) (1d), (1e) and (2). It may also occur on an agreeing noun in apposition, as in (9).

- (9) *fridia* *nathir*
 to=God_{ACC} ^{NAS}father_{IACC}
 ‘towards God the father’ (Ml. 127^d8)

Nasalisation after a genitive plural noun is exemplified in (10).

- (10) *itseúit* *macc* *ñgor*
 is_{3PL.PRES}=treasures_{NOM.PL} sons_{GEN.PL} ^{NAS}pious_{GEN.PL.MASC}
 ‘They are the treasures of pious sons.’ (Wb. 23^a9)

Nasalisation after a singular nominative neuter adjective is exemplified in (11).

- (11) *isinse* *ñduit*
 is_{3SG.PRES}=hard_{NOM.SG.NEUT} ^{NAS}to_{2SG}
 ‘It is impossible for you.’ (Wb. 5^b28)

Nasalisation after a noun phrase with an accusative or a nominative neuter noun+adjective is extremely rare: it only occurs once in Ml. 40^a20, reported in (12), out of 11 similar syntactic environments.

- (12) *atá* *debe* *mec* *nand*
 PV-be_{3SG.PRES} difference_{NOM} ^{NAS}little_{NOM.SG.NEUT} in_{3SG.NEUT.DAT}
 ‘There is a little difference there.’ (Ml. 40^a20)

The data clearly show that word class of the nasalised word is relevant for the occurrence of nasalisation. Nasalisation surfaces in the glosses according to the hierarchy in (13) below (see Roma 2018a: 13):

- (13) nasalisation on agreeing words and demonstratives > nasalisation on nominal noun modifiers > nasalisation on any modifier following a noun

I am aware that this picture is somehow biased by the following circumstances:

- prenominal adjectives are mostly proclitic (*cach* ‘every’ and *nach* ‘any’ are the most frequent cases; there are only 3 examples with stressed *cétnae* ‘first’ and 2 with *uile* ‘all’ in Ml.; 3 with *cétna* in Sg.; 1 with *uile* in Tur.); they can therefore be assumed to behave rather similarly to proclitic mutation-triggers such as articles, than as mutation-triggering inflected stressed words.
- demonstratives which can show nasalisation all begin with a vowel (*ucut*) or a vowel initial deictic particle (*ísin*, *ísiu*)

Nevertheless, nasalisation is clearly shown more frequently on adjectives than on genitives and prepositions, as already noted by Thurneysen himself (1905; GOI § 237). In Wb., nasalisation on prepositions is confined to the phonetic context after/on a vowel and does not occur after genitive plural nouns; genitive nouns are mostly not nasalised (overall 67 vs. 40), although nominative singular neuter nouns mostly nasalise a following genitive (18 vs. 12). The same does not hold for Sg. and Ml. In Sg., nasalisation on prepositions is well attested whatever the trigger except neuter adjectives (context (11) above), and even predominant after an accusative noun (12 vs. 9, overall on prepositions 19 vs. 33); nasalisation on genitives is largely predominant whatever the trigger (37 vs. 14). In Ml., nasalisation on prepositions is again frequent after most triggers except neuter adjectives (overall on prepositions 61 vs. 99, after neuter adjectives 7 vs. 16) and predominant on genitive nouns (84 vs. 50).

Perhaps surprisingly, the strongest nasalisation triggers in Wb. are singular neuter nouns, in Ml. genitive plural nouns. In Sg. all nasalisation triggers except neuter adjectives nasalise between 70% and 80% of expected cases.

2 Diachronic and synchronic variation

According to Thurneysen (GOI § 237), the frequent omission of the nasal in interconsonantal position is due to the regular dropping of a nasal in some consonant clusters. Thurneysen presumably drew his conclusions on his survey in Thurneysen (1905), which showed that the nasal was more frequently omitted in interconsonantal position. Nevertheless, this view implies that a nasal was lost between consonants through a highly irregular phonetic process (see below).

Thurneysen's phonological explanation for the lack of nasalisation does not address why the presence of a nasal between vowels apparently increases in time, or at any rate greatly differs from one corpus of glosses to another in certain syntactic environments: in the cases where nasalisation would occur on a vowel and/or follow a vowel (phonetic context as in 1), in Wb. 70 instances show nasalisation, 115 do not, in Sg. 68 show nasalisation and 24 do not, and in Ml. 189 show nasalisation and 38 do not. Even if one excludes nasalisation on prepositions, which appears to be generally rarer, the corresponding figures are Wb. 59 with nasalisation vs. 29 without, Sg. 54 with vs. 8 without, Ml. 145 with vs. 15 without. Of course, consonant dropping in interconsonantal position cannot be the reason for this kind of variation.

It could be assumed that nasalisation was left out in spelling in Wb., although it was realised phonetically, or, vice versa, that Wb. reflects a variety where nasals were dropped more easily between consonants. The former hypothesis would be surprising for initial vowels, and neither hypothesis is supported by independent evidence. Table 1 below reports the spelling of a few sample words with interconsonantal nasals in the three major corpora of glosses.

Table 1: Spelling of *n* between consonants in the interior of words (sample lexemes).

Lexeme	Wb.		Sg. and other Priscian glosses			Ml.	
	with	without	with	without	with	without	
	<i>n̄</i>	<i>n̄</i>	<i>n̄</i>	<i>n̄</i>	<i>n̄</i>	<i>n̄</i>	
<i>freċndairc</i> 'present', <i>freċndarcus</i> 'presence'	11	–	27 ⁶	1	20	–	
<i>aisndís</i> 'declaration' and related words (<i>aisndisse</i> , <i>-í</i>)	4	–	7	2	58	–	
<i>forċgaire</i> "command"	2	–	–	4	2	4	
<i>túailnge</i> "ability" and related words (<i>túailngigidir</i> , <i>túailngigiud</i> , <i>túailngigthe</i>)	1	–	–	–	4	1	

⁶ Once without *d*.

The comparison between Table 1 and the data in section 1 shows that the frequencies with which the nasal is spelt between consonants in the interior of words and in nasalisation contexts do not match. The behaviour of clusters arising from syncope, such as those mentioned by Feuth (1982: 92) and Ó Maolalaigh (1995–1996: 164) and illustrated in Table 1, is different from the behaviour of purported similar phonetic contexts between words. In Wb., while the nasal is regularly spelt in the interior of words, it is most frequently omitted between words, as shown in Roma (2018a) and noted above (sections 1.2 and 1.3). It may be added in this connection that in Ml. there occur a few examples⁷ where nasalisation surfaces on a simple, proclitic preposition, thereby appearing in a phonetic environment which resembles more closely the unstressed position of the interior of words. But, again, the spelling of the nasal in the interior of words in Ml. does not seem to be more regular than in the other corpora of glosses.

Indeed, different spelling conventions apparently hold for nasalisation on consonants and vowels after final nasals in the St. Gall Glosses (see 1.2 above), where nasalisation tends to be spelt regularly (72%), as opposed to both Wb. (5%) and Ml. (42%). This corresponds to Quin's (1979: 256, 258) data on the spelling of nasalisation after the accusative article, since Sg. seems to mark nasalisation after accusative *in* more frequently. This spelling could also reflect a different phonetic realisation. Note that Sg. also features 6 instances of nasalisation after Latin words ending in a nasal (the neuter nouns *nomen*, *pronomen* and *cognomen*, see Roma 2018a: 19).⁸

Bronner (2016) has shown that in the *Additamenta* in the Book of Armagh the kind of nasalisation dealt with in this paper is almost regularly spelt differently from nasalisation after proclitics such as articles, possessives, prepositions, conjunct particles and infix pronouns: only this kind of nasal is written separately between mid-height dots. Note that nasalisation seems to be regularly spelt in the *Additamenta*, but there are only 5 instances on an inflected preposition, of which 3 are of *and* 'in it'; moreover, only 2 instances occur between consonants, 1 of which involves an adjective. This last is reported in (14).⁹

⁷ Ml. 30^b10, 46^a1, 51^a5, 72^c9, 23^a5, 110^d10, 96^a13. See Roma (2018a: 12) for details.

⁸ Among these, *pronomen* is quite clearly a borrowing (so in eDIL), given the dative plural form *pronoibneib* (Sg. 200^b6); cf. nasalisation after the accusative singular but no nasalisation after the dative singular in *pronomen nātárcadach* vs. *o pronomen atárcadach* 'an anaphoric pronoun' (Sg. 209^b10).

⁹ Nasalisation is not marked in *contubart fland feblæ acheill ḋoo* 'Fland Feblae gave his church to him' (Thes. 2: 242.20–21 [Book of Armagh, folio 18^{va}34–5]) (see above section 1.2 on the absence of nasalisation between *-l* and *d-*); possibly also in *cach aleth ódib* 'each of them his way' (Thes. 2: 240.20 [Book of Armagh, folio 18^{ra}28]).

- (14) *arech* *n· donn*
 for=horse_{ACC} ^{NAS}brown_{ACC.SG.MASC}
 ‘for a brown horse’ (Thes. 2: 240.1 [Book of Armagh, folio 17^{va}29])

This kind of spelling seems to be an alternative to the superscript dot or *punctum delens* – which never occurs in the *Additamenta* (Bronner 2016: 42) – and is therefore considered by Bronner (2016: 43, 45) a chronologically and/or geographically restricted variant of such spelling, probably to be ascribed to an experimental phase. The flanking dots of the *Additamenta* occur once in Ml. and once in Sg. (Ml. 69^a23, Sg. 6^b11), according to Bronner (2016: 44), but, perhaps significantly, not in Wb.

Despite spelling variation, though, a purely graphic principle that could account for the observed variation between the corpora of glosses is highly unlikely. We must therefore conclude that either nasalisation was expanding or that Wb. and Ml. reflect the distribution of nasalisation in two different varieties. The situation in Sg. is closer to Ml., but the regular spelling of nasalisation after a final nasal seems to point to a somewhat different variety (see below about another peculiarity of Sg.). We turn therefore to possible diatopic and diachronic scenarios accounting for the observed variation in the occurrence of nasalisation in the contexts examined here. Given the nature of the sources under consideration, the discussion is confined to the diatopic and diachronic dimensions, bearing in mind that register can in principle account for variation within every single corpus of glosses (McCone 1985: 102; Ahlqvist 1988: 27).

Ó Muircheartaigh (2015: 128), in giving a summary of recent scholarship on dialectal variation in Old Irish, suggests that the homogeneity of the Old Irish standard language of the glosses may prevent us from finding dialectal variation in the Early Irish period through comparison of the corpora of glosses. He claims that a detailed examination of the distribution of forms in the glosses may not lead to a full understanding of any dialectal differences and may prove to be fruitless, because the three corpora might all belong to the same larger dialect area, situated in North-Eastern Ireland between the monasteries of Armagh, Bangor and Iona. Ó Muircheartaigh (2015: 124–125) admits that some of the features outlined by *GOI* and reiterated by Ahlqvist (1988) as possible dialectal features in the glosses, such as for instance the superlative suffix (*-em* vs. *-imem*), the form of the demonstrative *són* vs. *ón*, the form of the reflexive *céin* vs. *féin*, may reflect dialectal variation. However, although the variation between the anaphoric pronouns *ón* and *són* may indeed be one of dialect, for example, these anaphoric pronouns have left no trace in the modern

languages, leaving their geographical implications unknown, as concluded by Ahlqvist (1988: 26).

The major corpora of glosses may well belong to the same broad dialectal area. Still, even if later varieties cannot offer direct evidence to locate the outcomes of alternative forms in the Old Irish glosses, an attempt might be made to account for variation, if it surfaces, as I believe is the case for nasalisation. In principle, I expect that variation in this kind of mutation patterns in Old Irish could at first escape both morphological and spelling normalisation. Bronner's (2016) findings concerning the spelling in the Book of Armagh confirm this expectation.

Moreover, despite the fact that the distribution of nasalisation after inflected nominals in Old Irish cannot be plotted with respect to modern Gaelic dialects, as this kind of nasalisation has been lost throughout all varieties, its blocking vs. expansion may lie behind the diverging developments of nasalisation in Scottish Gaelic and Irish.

Drawing on the discussion in Ó Maolalaigh (1995–1996), in Roma (2018a) I suggested that lower frequency of nasalisation across phrasal boundaries and between consonants, as witnessed by Wb., could be linked to the eventual loss of nasalisation in many contexts, and therefore to the Scottish varieties.¹⁰ Nevertheless, Ó Maolalaigh's (1995–1996: 165) explanation for the loss of nasalisation following particles with consonantal codas presupposes the loss of nasalisation in interconsonantal position, a view that I am not inclined to accept, for the reasons shown above. In fact, my previous assumption concerning lower frequency of nasalisation simply pushes back in time the common Gaelic variety that lies behind both Irish and Scottish, but, given the inconsistencies between loss of interconsonantal nasals and absence of nasalisation after inflected words, it does not suggest any pathway for the emergence of the new pattern.

Therefore, it is better to assume that while Wb. must indeed reflect an older or at least a more conservative variety (of course in this respect, i.e. regarding the spread of nasalisation across phrases), Sg. and Ml. might reflect two similar but possibly diatopically distinct later varieties. In either case, the variety in Wb. could indeed reflect a stage which preserved a sequence of nasal + voiced

¹⁰ According to Ó Maolalaigh (2008: 232) some forms in the *Notes in the Book of Deer* may suggest that nasalisation of nominal modifiers following accusative singular nouns was not the norm in certain varieties of Gaelic in 12th-century eastern Scotland. However, other explanations are possible for *acuitid thoisig* 'his toisech's portion (?)' and *incathraig ele* 'the other monastery' or 'into another monastery'.

In *far nóendeilb* ‘according to the same paradigm’ (Sg. 90^{b2}), the same phrase as in (15), *deilb* seems to be a dative form, since the phrase parallels *far cétnu diull* ‘according to the first declension’, where adjective (*cétnu*) and noun (*diull*) are both clearly in the dative case. Therefore, the aphaeretic form of the article in *nóendeilb* would be a dative form. GOI (§ 467) in fact reports aphaeresis of monosyllabic forms of the article after *r* and preceding a numeral. Nevertheless, one example from Sg., (18) below, has eluded both the list in GOI and in Strachan’s (1903b).

- (18) *arbertar* *asnóentairmoirciunn*¹³
 PV·express_{3PL.PRES.PASS} from=the_{DAT.SG.NEUT}=one.ending_{DAT}
 ‘They are expressed by the same ending.’ (Sg. 33^{a19a})

Although Bauer (2015) classifies *as*+nasalisation here as a relative form of the copula followed by relative nasalisation, the phrase clearly parallels Latin *ex eadem forma* and is to be read as the preposition *a(s)*+aphaeretic form of the dative article.¹⁴ It is tempting to see in this form of the article, which looks like a nasal segment floating onto the noun and becoming a nasalisation nasal (both on vowels and on consonants,¹⁵ albeit only after prepositions) in attested phases of the language, an incipient process which led to the generalisation of a nasalising article, as in Scottish Gaelic. This process would be the mirror image of the reanalysis proposed by Ó Maolalaigh (2016: 88–90) for the development of *is ann* as a topicalisation marker of non-nominal elements in Scottish Gaelic, i.e. the reinterpretation of the relative nasal segment following the copula as an independent morpheme (the inflected preposition *ann*).

If the link proposed above between Sg. and Scottish Gaelic is not fallacious, the spread of nasalisation across phrasal boundaries witnessed by Ml. could, on the other hand, reflect an Irish variety without a direct extant offspring. This view relies on the observation that the drift towards the expansion of nasalisation as a phrasal marker, even affecting proclitic prepositions, does not seem to

¹³ The length mark seems rather on <o> than on <e> (despite *oén-* in Thes., Bauer [2015], Bauer, Hofman, and Moran [2018]).

¹⁴ I doubt whether this may suggest an alternative origin for nasalisation after the preposition *ós* ‘above’ or rather adds to the analogical origin proposed by Ó Maolalaigh (2016), i.e. contamination with the relative form of the copula *as*.

¹⁵ The form in *tresñgné* ‘through the type’ (Sg. 73^{b1}) is probably not “the masc. form of the article for the neuter *gné*” (Bauer 2015), despite Strachan’s (1903b: 488) statement that it is “undoubtedly for *tresin ñgné*”, but simply the nasalising accusative singular neuter article, again without the initial vowel, i.e. for *tresañgné* (so eDIL s.v. *tre*).

have been continued any farther than the stage reflected in these glosses. The position of Tur. is very doubtful because the available data are scanty. Nevertheless, this corpus can be paired with Ml., since nasalisation appears regularly.

The two alternative scenarios surmised in Roma (2018a) and suggested here can be sketched as in (a) and (b) in Figure 1 below, respectively, where the horizontal axis roughly reflects diachronic relationship (distance along the time axis) and the vertical axis roughly reflects diatopic variation (distance along the spatial axis). The tentative nature of the hypothesis that links variation within Old Irish with later dialects is represented by question marks.

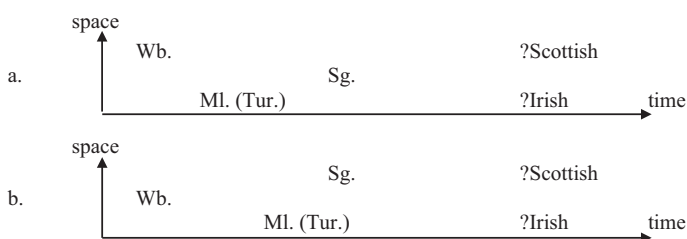


Figure 1: Diachronic and possible diatopic variation in Old Irish.

For the reasons outlined above, I maintain that the scenario in (b) is more likely. Nevertheless, I cannot draw any well-founded conclusion, except that the data on nasalisation after inflected nominals point to an early differentiating feature between Gaelic dialects; this split may be either already reflected in the different pictures offered by Wb., Sg. and Ml., or in the later stage documented by Sg. and Ml., which in some varieties, lost to us for lack of documentation, may have gone even further.

3 Conclusions

In this chapter I have tried to show that the absence of nasalisation after inflected nominals in Old Irish cannot be due in the first place to the loss of a nasal consonant in consonant clusters. Variation across the major corpora of Old Irish glosses is not trivial and must be due to diachronic change, i.e. the absence of nasalisation in some consonantal environments and across phrases is a conservation as opposed to its later expansion, and possibly also to dialectal or, as a consequence, register variation, for which however it must be acknowledged that we do not have

sound independent evidence. Possible connections between the variation in the Old Irish glosses and the divergent developments of nasalisation in Irish and Scottish varieties respectively are in fact hard to determine at the current state of research but it is suggested here that the St. Gall Glosses might reflect a variety that lies behind the developments of Scottish dialects. Since it has been suggested on different grounds that nasalisation may have been one of the earliest differentiating features between Gaelic dialects (Ó Maolalaigh 2008: 247), and that nasalisation transfer is crucial for the Scottish developments (Ó Maolalaigh 1995–1996: 165–167, 2008: 248), further inquiry along these lines might prove fruitful.

Acknowledgement: I am indebted to the editors and to three anonymous reviewers for comments and suggestions which greatly helped to improve an earlier version of this paper.

Jürgen Uhlich

8 On the obligatory use of a nasalising relative clause after an adjectival antecedent in the Old Irish glosses

1 Introduction

According to Thurneysen (GOI § 383), “an adverb formed from the dative of the adjective cannot be used in periphrasis with the copula before its clause . . . [Instead,] the adverbial form is replaced by the nominative sg. neuter of the adjective . . . , and a nasalizing relative clause follows”. The same cleft-sentence construction is referred to in the second part of § 498,¹ and the examples given² are:

- (1) *arndip maith n-airlethar a muntir*
so.that-COP_{3SG.PRES.SUBJ} good_{NOM.SG.NEUT} ^{NAS}care_{3SG.PRES.SUBJ} his household_{ACC}
‘that he care well for his household’ (Wb. 28b32)

(lit. ‘so that it may be good/a good thing how he cares . . .’ [GOI § 383] as opposed to lit. ‘. . . well that . . .,’ with an adverbial antecedent)

- (2) *is lērithir insō no nguidim-se dia*
COP_{3SG.PRES} zealous_{EQ} the=this_{ACC} PV.^{NAS}beseech_{1SG.PRES=1SG} God_{ACC}
‘as zealously as this do I beseech God’ (Wb. 27^d19; author’s trans.)

- (3) *is dīnnimu do-ngní alail*
COP_{3SG.PRES} zealous_{COMP} PV.^{NAS}do_{3SG.PRES} other_{ACC}
‘It is more carelessly that he makes the other.’ (Wb. 4^c33)

Moreover, while “a nasalizing relative clause can be replaced by a formally independent (i.e. principal) clause in almost every instance, . . . this is not

¹ A more detailed description and evaluation of this construction is given by Mac Coisdealbha (1998: 155–157; cf. 257, n. 82). On some of his interpretations of specific cases see individually below.

² Here quoted from Thes., with added editorial macrons and, for these introductory examples, word division and hyphenation. For the length in *in sō* and words of similar structure, see Breatnach (2003).

possible . . . in the constructions described in §§ 499,³ 501,⁴ nor after a neuter adjective in periphrasis with the copula (§ 498)” (GOI § 505).

According to Thurneysen, therefore, the construction under consideration here presents one of the very few grammatical contexts in which a nasalising relative clause is predictable.⁵ If borne out by the detailed re-examination of the material undertaken here, this would allow a confident assessment – within this particular context – of how the formal characteristics of the nasalising relative construction may be affected by adverse linguistic developments. To give two examples: (i) it would be possible to ascertain that in a putative case **is cosmail asbeir* ‘similarly he says’, it is (at least statistically) more likely that the expected relative nasalisation in *asmbeir* has been suppressed in interconsonantal position than that this should be counted among the symptoms for the nasalising relative construction as such already being in the process of loss (see also the discussion in Roma 2018a, and Roma, this volume);⁶ (ii) being able to rely securely on the presence of a nasalising relative construction would also allow one to draw conclusions on the use of the different classes of infixed pronouns in relative sentences overall. For the only other construction that

3 The specific *figura etymologica* construction of a verb connected to its own verbal noun as antecedent in an adverbial relative construction (of the pattern “the deliverance wherewith he delivered”), for which see further Stüber (2010–2012).

4 With an object antecedent, where the nasalising relative alternates synchronically with a leniting relative; see Schrijver (1997b: 91–113).

5 This ruling has been variously questioned by citing formally deviant examples; see Pedersen (1899: 391, 413, 414), Mac Coisdealbha (1998: 155), Ó hUiginn (1986: 58) and Isaac (in Mac Coisdealbha 1998: 257). Their objections and supporting examples will be assessed in sections 5 and 6.

6 The very initial symptoms of this linguistic innovation are described in GOI § 506. For the (partly sporadic) loss of interconsonantal nasals see GOI § 180 (2–3); Thurneysen (1905: 1–2), cf. Quin (1979–1980: 256). To the examples given there need to be added cases that show that even a grammatically functional nasal could be suppressed in this way (as is merely hinted at in GOI § 504 [c]): thus, while there are numerous instances where nasalisation is expressed between consonants, such as *arnach n dermandadar dia* ‘that God should not forget him’ (ML. 32^d5), its loss is seen in *connach[n]gabad huall de* ‘that pride might not seize him’ (ML. 69^a17); cf. further the parallel examples of *indhual rodngabsom* ‘the pride that had seized him’ (ML. 61^a1), vs. *huanduail rod[n]gab* ‘by the pride that seized him’ (ML. 49^a3) [‘/n/’ in both cases inserted by me]. Accordingly, while Schrijver (1997b: 97, 100) differentiates between cases like *asrect maid asmbeir do airiuc tías*, ‘that it is a good law, which he says above he has found’ (Wb. 3^d10), and *taidbsin afrinne asber* ‘it is an exposition of his righteousness which he utters’ (ML. 40^a15), as showing variation between nasalising and leniting relatives after object antecedents, the second example – together with numerous others of similar structure listed in his footnotes (Schrijver 1997b: 97, nn. 1–2, 100, nn. 1, 3–5) – may also be taken as showing suppression of nasalisation.

according to Thurneysen strictly predicts a nasalising relative, the adverbial *figura etymologica*, a full re-examination of the Old Irish material has already been carried out by Stüber (2010–2012), confirming that the nasalising relative is indeed compulsory, and while there are no instances of the pattern **as[m] beir* in her collection, a couple of cases that involve infixed pronouns are worthy of note (and will be addressed below).

2 Differentiations

2.1 Adverbial cleft sentence

Before addressing the adjectival cleft construction itself, it will be useful to differentiate it from some other patterns that are similar in form or in meaning. Beginning with the latter, the observed rule (see also section 3) that a de-adjectival adverb cannot be fronted in a cleft sentence means that the adverbial cleft pattern, consisting of an adverbial expression followed by a non-relative verb, is confined to prepositional phrases (for more information, see Griffith's chapter, this volume) and lexicalised adverbs,⁷ compare:

- (4) *issamlid* *sin* *danō* *bid* *ícc*
 COP_{3SG.PRES}=like_{3SG.NEUT.} that_{ACC} then COP_{3SG.FUT} salvation_{NOM}
dāsi *tuistiu* *c[h]laindde*
 to_{3SG.FEM}=3SG_{FEM} bearing_{NOM} children_{GEN.PL}
 'It is thus then that the bearing of children will be salvation to her.' (Wb. 28^b17)

Thus, while the adverbial element is contained in the fronted item *samlid* 'like it, thus', the same is not permissible with, for instance, (*in/co*) *maith* 'well' (see GOI §§ 379, 381), and instead of †*is (in/co) maith airlithir*, literally 'it is well that he cares', the adverbial element is shifted to the relative connection itself (= the relative pronoun of other languages), resulting in *is maith n-airlethar*, literally 'it is good how he cares'.

⁷ As well as subordinate clauses, see GOI § 814.

2.2 Adjective + subject complement clause

Additionally, a fronted adjective may not only be the antecedent of an adverbial relative, but also of a subject clause (with the meaning ‘[the fact] that . . .’), for which a nasalising relative is merely an option (cf. GOI § 503), yielding a formally similar or indeed identical construction:

- (5) a. with non-relative continuation, copula:

is *follus* *trisodin* *is*
 COP_{3SG.PRES} clear_{NOM.SG.NEUT} through=that_{ACC} COP_{3SG.PRES}
asintsalm *hōdūaid* *d[u]iuc*
 out-the=psalm_{DAT} from=David_{DAT} PV·bring_{AUG.3SG.PRES}
 ‘It is clear thereby that it is out of the psalm from David that he brings . . .’ (ML. 25^d18)

- b. with non-relative continuation, stressed verb:

is *glé* *limm* *nīdigénte*
 COP_{3SG.PRES} clear_{NOM.SG.NEUT} with_{1SG} NEG·commit_{2PL.COND}
 ‘It is clear to me that ye would not commit. . .’ (Wb. 9^d9)

- c. with relative continuation, copula:

isfollus *doib* *asnoipred*
 COP_{3SG.PRES}=clear_{NOM.SG.NEUT} to_{3PL} COP_{3SG.PRES.REL}=^{NAS}working_{NOM}
fir *oirdnithi*
 man_{GEN} appointed_{GEN}
 ‘It is manifest to them that it is the working of a supreme being.’ (Wb. 1^b14)

- d. with relative continuation, stressed verb:

Is *follus* *rundgabsat*
 COP_{3SG.PRES} clear_{NOM.SG.NEUT} AUG-^{NAS}3SG_{NEUT}·take_{3PL.PRET}
terchoiltisiu *indiumsa*
 thy=determinations_{NOM.PL}=2SG in_{1SG}=1SG
 ‘It is clear that Thy determinations are in me.’ (ML. 74^d7)

Here, both semantic considerations and, in three cases, the intervening elements (underlined) make it clear that these are not an adjectival cleft ‘it is clearly/in a clear fashion that . . .’, literally ‘it is clear how . . .’, but non-cleft copula sentences with a complement clause as subject: ‘it is clear that . . .’. The distinction to be observed is thus between ‘(the circumstance) by which’ of the construction under discussion and ‘(the fact) that’ with a subject complement clause. In individual cases – particularly when the main verb is not the copula – doubts could arise as to which of the two constructions is intended; compare:

- (6) *is derb contoroe farao achrid*
 COP_{3SG.PRES} certain_{NOM.SG.NEUT} PV·turn_{AUG.3SG.PRET} Pharaoh_{NOM} his=heart_{ACC}
do miscuis macc n israhel
 to hatred_{DAT} children_{GEN.PL} ^{NAS}Israel_{GEN}
 ‘Certainly Pharaoh had turned his heart to hatred of the Children of Israel.’ (Ml. 123^{b7}) gl. *bene conuertit Pharaoh cōr suum ad odiendum populum, quem . . .*

Here only semantic considerations suggest that the concept of ‘certainly’ (with an English adverb rendering the Old Irish adjective *derb*) is meant to qualify the implied superordinate verb (e.g. in *‘One can state with certainty that . . .’) rather than the verb of the associated sentence (‘had turned in a certain way’). Similarly, *bene conuertit* of the Latin original can hardly stand for ‘he turned well’ = ‘he did well to turn’, but must mean elliptically ‘it may be well stated that . . .’. The following gloss appears to be more ambiguous:

- (7) *ciaso demnithir sō forcomnuoir bieid*
 although-COP_{3SG.PRES} certain_{EQ} this_{ACC} PV·turn_{AUG.3SG.PRET} be_{3SG.FUT}
aimser nad creitfider et dosluinfider
 time_{NOM} NEG_{SUB} believe_{3SG.FUT.PASS} and PV·deny_{3SG.FUT.PASS}
 ‘Though it is so certainly that it has happened, there will be a time when it will be disbelieved and denied.’ (Wb. 28^{c14})

As presented in Thes., this glosses *Spiritus [autem] manifeste dicit, quia . . .*, ‘Now the Spirit expressly says that . . .’, suggesting that here, too, *demnithir sō* ‘as certain as this’ qualifies an implied superordinate verb (e.g. in *‘one can say as certainly as this that . . .’), just as *manifeste* qualifies *dicit*, in which case ‘certainly’ in Thes. would have to be changed to ‘certain’. However, Charles-Edwards (1971) has shown that the full Latin context, quoted only partially in Thes., includes the previous sentence (1 Timothy 3:16–4:1):

- (8) *et manifeste magnum est pietatis sacramentum quod manifestatum est in carne et iustificatum est in spiritu apparuit angelis praedicatum est gentibus creditum est in hoc mundo adsumptum est in gloria[.] Spiritus manifeste dicit quia in novissimis temporibus discendent quidam a fide . . .*⁸

⁸ This follows the Würzburg manuscript itself (see Stern 1910: folio 28^{va}17–23, here given with slight normalisation, mainly concerning abbreviations and word-internal spaces), which deviates from that in the *Vulgate* (Weber 2007) in some minor detail only.

‘Great indeed, we confess, is the mystery of godliness: He was manifested in the flesh, vindicated by the Spirit, seen by angels, proclaimed among the nations, believed on in the world, taken up in glory. Now the Spirit expressly says that in later times some will depart from the faith . . .’ (ESV).

In this, the Old Irish equative *demnithir sō* in fact refers back to the first *manifeste*,⁹ which qualifies the following statement directly (literally ‘and manifestly great is . . .’) and not via a superordinate *verbum dicendi*, meaning that the intended construction is an adjectival cleft after all, including the correct translation ‘certainly’. As a final example, both constructions are found in:

- (9) *is gnāth is trom achothud*
 COP_{3SG.PRES} usual_{NOM.SG.NEUT} COP_{3SG.PRES} heavy_{NOM.SG.MASC} his=sleep_{NOM}
adi 7 is cián m̄ bis ánd
 ANAPH and COP_{3SG.PRES} long_{NOM.SG.NEUT} ^{NAS}be_{3SG.HAB.REL} in_{3SG.NEUT}
 ‘His sleep is wont to be heavy and he is wont to be long therein.’ (Ml. 100^a10)
 (lit. ‘it is usual that . . . and it is long how . . .’ / or: ‘a long time by which . . .’)

2.3 Substantivised adjective as object antecedent

In another superficially similar construction, the connection between the fronted adjective and the relative clause cannot be interpreted adverbially; rather, the predicate adjective serves as a substantive and is interpreted as the object of the headless relative clause:

- (10) *isbecc rofitemmarni irrúnaib dáé*
 COP_{3SG.PRES}=little_{NOM.SG.NEUT} AUG-know_{1PL.PRET/PRES}=1PL in=mysteries_{DAT.PL} God_{GEN}
 ‘It is little we know in God’s mysteries.’ (Wb. 12^c5)

⁹ Charles-Edwards (1971: 189) argues further that “*demnithir* in the Irish gloss refers to the first *manifeste* (the one not given in the *Thesaurus*) and compares it with the second *manifeste*. The glossator’s point is that it is just as certain a scriptural truth that Christ was incarnated etc. as that some will lapse from the faith. The two are equally manifest”. This, however, is not borne out by the text, where *demnithir sō*, in referring to the first *manifeste* as ‘as certain as stated above,’ does so by qualifying *forcommucuir* ‘it has happened,’ without any connection to the second part of the gloss that alone corresponds to what follows the second *manifeste*.

The preposition in *i rrúnaib* ‘in the mysteries’ here necessitates this interpretation, as opposed to a putative adjectival cleft **is becc ro-fitemmar-ni rúna dé*, literally ‘it is small how we know God’s mysteries’. The same construction is editorially assumed in:

- (11) *hūare rombu mór dorat dūaid*
 because AUG-^{NAS}COP_{3SG.PRET} big_{NOM.SG.NEUT} PV-give_{AUG.3SG.PRET} David_{NOM}
[du]lāri frit adrad su
 of=diligence_{DAT} to=your_{SG} worshipping=2SG
 ‘because David has given much diligence to worshipping Thee’ (ML 136^c11)

But the emendation in Thes. is not necessary if instead the manuscript reading is taken as an adjectival cleft, with an English adverbial translation ‘it is greatly that David has given diligence . . .’ (Griffith and Stifter 2013). On the other hand, this is not possible (*pace* Mac Coisdealbha 1998: 155) in the following case:

- (12) *ismó rochéess críst airi .i. báas*
 COP_{3SG.PRES}=more AUG-^{LEN}suffer_{3SG.PRET} Christ_{NOM} for_{3SG.MASC} i.e. death_{NOM}
 ‘It is more that Christ has suffered for him, to wit, death.’ (Wb. 6^c8)

The main sentence as far as *airi* could be understood as an adjectival cleft with adverbial meaning, ‘it is more greatly that Christ has suffered for him’, but only if an innovatory leniting relative is admitted (cf. GOI §506). This, however, would leave the added explanatory object *báas* out of construction. Therefore, Sims-Williams (1984: 193, albeit without quoting *.i. báas* in support), is right in understanding *mó* “substantivally as object antecedent . . . ‘it is a greater (thing) (more) that Christ has suffered for him’” (see also Griffith’s contribution in this volume).

- (13) *bid mó dongēnaesiu oldaas rofoided cucut*
 COP_{3SG.FUT} more PV-^{NAS}do_{2SG.FUT}=2SG ‘than’ AUG-send_{3SG.PRET.PASS} to_{2SG}
 ‘Thou wilt do it more than has been sent to thee.’ (Wb. 32^a25),
 gl. *sciens quoniam et super id quod dico facies* (Philemon 21) ‘knowing that you will do even more than I say’ (ESV).

The translation offered in Thes. presupposes reading *dongēnae* as *don[d]gēnae* with an infix pronoun. However, Pedersen (1899: 391) suggests that “dagegen

gehört *bid mó dongenaesiu* 32 a 25 eher in § 71 [however, *bid mó dongenaesiu* 32^a25 belongs rather to § 71]”, referring to page 392f. where he deals with nasalising relatives connecting to an object antecedent. That this – ‘thou wilt do more . . .’ – is the correct interpretation is proven by the Latin context quoted above.

3 Validity of a rule excluding de-adjectival adverbs from fronting

A few glosses superficially give the impression that in them, a de-adjectival adverb is clefted, and they now need to be addressed individually.

- (14) *ba infortgidiu* 7 *ba hitemul*
 COP_{3SG.PRET} the_{DAT.SG.NEUT}=covert_{DAT.SG.NEUT} and COP_{3SG.PRET} in=darkness_{DAT}
dugníth saul
 PV·do_{3SG.IMPF} Saul_{NOM}
 ‘it was covertly and it was in darkness that Saul . . . used to make . . .’
 (ML. 30^a3) (for *infortgidiu* the manuscript has *imfortgidiu*)

If extracted and viewed in isolation, *ba in [sic leg.] fortgidiu du-gníth* would indeed constitute an adverbial cleft with a non-relative verb – for predicted **ba fortgide du-gníth* ‘it was covert how he used to make’ – but with the actual pairing of two diverse fronted elements, the continuous phrase *hi temul du-gníth* is a normal adverbial cleft beginning with a prepositional phrase, and with such mixed fronting, the construction agrees most naturally with the second phrase *hi temul*, and the first phrase has been secondarily adapted to suit this syntactic context. For a similarly mixed fronting construction, compare ML. 41^d9 in (88) below.

- (15) a. *non difculter*^{gl. 21} *eueniat*^{gl. 22};
i. ní baindodaing
 i.e. NEG COP_{3SG.FUT}·the_{DAT.SG.NEUT}=difficult_{DAT.SG.NEUT}
 ‘i.e. it will not be with difficulty’ (ML. 61^a21)
- b. *dufórban*
 PV·happen_{3SG.PRES}
 ‘it happens’ (ML. 61^a22; author’s trans.)

Here one might expect that the connected Latin phrase *non difculter eveniat*, ‘it may/will not happen in a difficult way’, were explained by an equally unified gloss, and in that case, *in dodaing*¹⁰ would be a fronted de-adjectival adverb, and the whole sentence would stand for predicted **níba dodaing du-forban* (with nasalising relative). However, in the manuscript there is a clear space between *indodaing* and *duforban*, with gloss 61^a21 being almost exactly coextensive with the Latin phrase it explains and 61^a22 only beginning over the second part of the *u* of *eueniat*. Therefore, these are indeed separate glosses, and the adverb *in dodaing* renders *difculter* in isolation, not as the first part of a cleft sentence. The same is even more clearly the case in:

- (16) *multum .i is indil asferr*
 much i.e. COP_{3SG.FUT} the_{DAT.SG.NEUT}=many_{DAT.SG.NEUT} COP_{3SG.PRES.REL}=better
iudeus quam gentilis
 Judaeus than Gentilis
 ‘*multum* i.e. it is greatly that *Judaeus* is better than *Gentilis*.’ (Wb. 2a4)
 gl. *multum per omnem modum* (Rom. 3:2) ‘much in every way’ (ESV)

As Thurneysen remarks on this isolated example, “the construction seems un-Irish” (GOI § 383 n.). While the combination of a fronted adverb with a relative verb could be justified as an incipient innovation (for which see GOI § 506), the adverbial formation *ind il* itself (from *il* ‘many, much’, with *ind* as described in GOI § 379) is entirely unparalleled. Instead, in order to express the concept of ‘greatly’ in this construction, **is mór as ferr* might be expected – compare the material collection in section 4.2 below that does not include cases of the fronted positive *mór*, but, in many instances, the corresponding comparative *mó* ‘more’ instead. Rather than representing a natural Irish expression, then, *ind il* is best explained as a mechanical rendering of the Latin adverb *multum*, and to the extent that this is an artificial process, a correspondingly artificial translation ‘muchly’ may be proposed. This analysis ties in well with the more general observation that the unmarked Old Irish equivalent of what would be de-adjectival adverb formation in other languages is precisely the adjectival cleft under discussion, thus most clearly with comparatives and superlatives, for which direct adverbs like *indluindiu* (Ml. 32^d1, ‘more angrily’, glossing *commotius*) or *innáam* (Wb. 1^c20, ‘most greatly’, glossing *primum*), “are never

¹⁰ Griffith and Stifter (2013) take *indodaing* instead as containing the preposition *in*, but since *dodaing* is an adjective (albeit as such capable of substantivisation) and *difculter* an adverb, direct equivalence of *indodaing* to the latter is more likely.

found in a clause, but occur only as isolated glosses, the language of which is probably somewhat artificial” (GOI § 382). Against this background, Mac Coisdealbha (1998: 156) suggests that:

such a situation obtained in part also for the non-comparative adverbial derived from the adjective, i.e. that it was expressed as a fronted element in the COP. EMPH. construction. . . . This suspicion is strengthened by the general paucity of such *ind*-derivatives in the Würzburg period especially in complete clauses (as isolated translations of Latin adverbs they are more frequent).

In the present case, the only difference is that such a gloss on a Latin word in isolation has been embedded unchanged into the adjectival cleft structure. The exact same mixed construction is found, with an embedded Latin adverb, in:

- (17) *nibbu machdad tra bed figurate-*
 NEG-COP_{3SG.PRET} wonder_{NOM} then COP_{3SG.PST.SUBJ} figuratively
nombed a· uirgo filius asbeir hieronymus
 PV^{NAS}be_{3SG.PST.SUBJ} the uirgo filius PV-say_{3SG.PRES} Jerome
 ‘and it were no wonder then that *uirgo filius* that Jerome speaks of, was figuratively’ (Sg. 62^{b2})¹¹

On the other hand, an ellipsis of the natural Irish construction, i.e. even without the following main verb, may be seen in:

- (18) *imgabaid etbadtreit aris huisse*
 shun_{2PL.IMPV} and-COP_{3SG.IMPV}=quick_{NOM.SG.NEUT} for-COP_{3SG.PRES} proper_{NOM.SG.FEM}
aimgabáil.
 its=shunning_{NOM}
 ‘Shun ye and let it be quickly, for it is proper to shun it.’ (Wb. 9^{d6})
 (gl. *fornicationem fugite*)

for which the complete expression of the second part may be predicted as **bad treit imme-n-imgabaid*, ‘let it be quick how you shun (it)’ (while conceding that formally, *treit* could be either adjective or adverb).

¹¹ The adverb *figurate* was likely obtained from the wider context of this passage (Thes. 2: 116.1 = Hertz [1855–1858] 2009, 1: 145.20). It is used shortly before (at Thes. 2: 115.16 = Hertz 2009, 1: 145.15).

4 Old Irish corpus of adjectival cleft sentences¹²

4.1 Fronted *cían*

To begin with, a separate section is devoted to *cían* merely because in this construction, it is impossible to decide if *cían* is used with its adjectival value ‘long’ or in its equally common substantival function ‘a long time’ (compare also [89] in section 5.4 below).

4.1.1 With overt spelling of nasalisation

- (19) *nī ba* *cián* *ṁ bete* *oca* *cloīnib*
 NEG-COP_{3SG.FUT} long_{NOM.SG.NEUT} ^{NAS}be_{3PL.FUT.REL} at-their wickednesses_{DAT.PL}
 ‘They will not be long at their wickednesses.’ (Ml. 28^a10)

- (20) *nī ba* *cīan* *mbias* *in* *pecthach*
 NEG-COP_{3SG.FUT} long_{NOM.SG.NEUT} ^{NAS}be_{3SG.FUT.REL} the_{NOM.SG.MASC} sinner_{NOM}
 ‘The sinner will not abide long.’ (Ml. 56^c22)

- (21) *nī ba* *cián* *ṁ bete* *and*
 NEG-COP_{3SG.FUT} long_{NOM.SG.NEUT} ^{NAS}be_{3PL.FUT.REL} in_{3SG.NEUT}
 ‘They will not be there long’ (Ml. 66^d14)

- (22) *is* *cián* *ṁ bis* *ánd*
 COP_{3SG.PRES} long_{NOM.SG.NEUT} ^{NAS}be_{3SG.HAB} in_{3SG.NEUT}
 ‘He is wont to be long therein.’ (Ml. 100^a10)

4.1.2 Orthographically¹³ ambiguous regarding nasalisation

- (23) *iscián* *arfolmas* *dún* *insin*
 COP_{3SG.PRES}=long_{NOM.SG.NEUT} PV·undertake_{3SG.PRET.PASS} for_{1PL} the_{NOM.SG}=that_{NOM}

¹² The following collection is intended to be complete for all sources edited in Thes., amounting in the main to the Würzburg, Milan and St. Gall Glosses.

¹³ In what follows, a distinction is made between *orthographically ambiguous* for cases in which a nasalisation, if present, would have been audible – such as in *ar-folmas* in (23), where <f> could represent either unlenited *f* /f/, lenited *f* /f̊/ or nasalised *f* /v/ –, and *phonologically ambiguous* for cases in which a nasalisation could not have affected the stressed anlaut in pronunciation beyond non-lenition – such as the *r* of *do-réracht* (24).

‘It is long since that has been destined (has been imminent) for us.’¹⁴
(Wb. 21^a2)

For relative *ar-* in (23) rather than *ara-*, see further under (90) below.

4.1.3 Phonologically ambiguous regarding nasalisation

- (24) *is* *cīan* *doréracht* *Emain*
 COP_{3SG.PRES} long_{NOM.SG.NEUT} PV·abandon_{3SG.PRET.PASS} Emain_{NOM}
 ‘Long since has Emain been forsaken.’ (Thes. 2: 317.6 and 317.15 [Hymn ii])

4.2 With comparative (and equative or superlative)

According to Thurneysen (GOI § 383), the adjectival cleft sentence “is the normal construction with adverbial forms of comparison” – albeit in a somewhat condensed expression for “replacing” or “corresponding to adverbial forms of comparison in other languages”, since within the Irish construction itself, only the basic adjectival forms (i.e. those not overtly marked as adverbs) may be used. Compare similarly Mac Coisdealbha’s (1998: 156) description: “The comparative and superlative attributive adjectives and corresponding adverbs *must* be formed predicatively with the copula.” On account of this observation, a separate section is here dedicated to fronted degrees of comparison, and most of the extant examples involve a comparative.

Commenting on the basic ‘*is maith* construction,’ Sims-Williams (1984: 193) remarks further: “Note, however, that a nas. rel. clause is not regular in the similar constructions with a comparative (Ml. 22^c14 = [102] below) or a superlative, *in fer as deg do-cheil bile*, ‘the man who best hides a tree’, Thurneysen 1946: 322 (cf. 681 n. 126) – on the Welsh construction which Thurneysen compares see P. Mac Cana, *Celtica* 7 (1966) 91–115.” The two examples adduced, however, are not parallel. While the second, construed superlative case illustrates a process of syntactic raising of the second relative clause to the level of the first – a process that will be addressed, with some real examples, in section 5.2 below – Ml. 22^c14 begins with *air is mōu ru-icim les . . .* ‘for it is more that I need . . .’ and thus with a non-relative copula that does not deliver a context for raising. Instead, this is to

¹⁴ Following the translation in Kavanagh (2001: 103 s.v. *ar-folmathar*), vs. ‘it is long since he destined (?) that to us’ (Thes. 1: 631), but see GOI (§ 708 note) on the analogical spread of the third person singular passive preterite ending -s.

be recognised as one of the few, innovatory exceptions to the nasalising relative rule in the adjectival cleft construction, see section 6.2.

Apart from such exceptions to be discussed further below, the attested cases involving fronted degrees of comparison are:

4.2.1 With overt spelling of nasalisation

- (25) a. *isléir* *dorigni* *indalalestar*
 COP_{3SG.PRES}=careful_{NOM.SG.NEUT} PV·do_{AUG.3SG.PRET} one.of.two=vessel_{ACC}
 ‘It is carefully he has made one of the two vessels.’ (Wb. 4^c32) (= [46a] below) followed by:
- b. *isdínnimu* *dongní* *alail*
 COP_{3SG.PRES}=careless_{COMP} PV·^{NAS}do_{3SG.PRES} other_{ACC}
 ‘It is more carelessly that he makes the other.’ (Wb. 4^c33)
- (26) *condibferr* *donberaidsi* *oldaas*
 so.that-COP_{3SG.PRES.SUBJ}=better PV-^{NAS}3SG_{NEUT}·give_{2PL.PRES.SUBJ}=2PL ‘than’
cách
 anyone_{NOM}
 ‘that you may give it better than anyone (else)’ (Wb. 16^c9) (*donberaid* for *do-nd-beraid*; see Thes. 2: 477)
- (27) *ismóa* *dongnísom* *oldaas*
 COP_{3SG.PRES}=more PV-^{NAS}3SG_{MASC}·do_{3SG.PRES}=3SG_{NEUT} ‘than’
dontlucham
 PV-^{NAS}3SG_{NEUT}·ask_{1PL.PRES}
 ‘He does it more than we ask it.’ (Wb. 21^d9) (*dongní* . . . *dontlucham* = *don[d]-gní* . . . *don[d]-tlucham*)
- (28) *Corrop* *mōoassamōo* *et corrop*
 so.that-COP_{AUG.3SG.PRES.SUBJ} more=and=more and so.that-COP_{AUG.3SG.PRES.SUBJ}
ferrassaferr *donimdigí[d]* *desseirc dé*
 better=and=better PV·^{NAS}multiply_{2PL.PRES.SUBJ} love_{ACC} God_{GEN}
et commessim
and neighbour_{GEN}
 ‘So that more and more, and so that better and better, ye may abound in love of God and of neighbour.’ (Wb. 23^b1)

- (29) *combad mōu dē donadbastae molad*
 that-COP_{3SG.PST.SUBJ} more of_{3SG.NEUT} PV.^{NAS}show_{3SG.PST.SUBJ.PASS} praise_{NOM}
dæ triachaingnīmu
 God_{GEN} through=his=good.deeds_{ACC.PL}
 ‘that the praise of God might be more shewn (*sic*) forth through His good deeds’ (Ml. 37^b23)
- (30) *ní lugu asnindet lāthar innandūle*
 NEG less PV.^{NAS}show_{3SG.PRES} disposition_{NOM} the_{GEN.PL.FEM}=elements_{GEN.PL}
dodia 7 nundfoilsigedar indāas
 to=God_{DAT} and PV.^{NAS}3SG_{MASC}·manifest_{3SG.PRES} ‘than’
 ‘not less does the disposition of the elements set forth concerning God and manifest Him than . . .’ (Ml. 42^b18)
- (31) *combad mōu dē nongabtis imma*
 that-COP_{3SG.PST.SUBJ} more of_{3SG.NEUT} PV.^{NAS}take_{3PL.PST.SUBJ} the_{ACC.PL.NEUT}
formgaire
 commands_{ACC.PL}
 ‘that they might the more receive the commands’ (Ml. 53^c13)
- (32) *is mōu dundrigēnsat indaas*
 COP_{3SG.PRES} more PV-^{NAS}3SG_{NEUT}·do_{AUG.3PL.PRET} ‘than’
oidrairlēcissiu
 PV-3SG_{NEUT}·permit_{AUG.2SG.PRET}=2SG
 ‘They have done it more than Thou hast permitted it.’ (Ml. 87^a8)
- (33) *cesu meinciu aranecar . . . arecar*
 although-COP_{3SG.PRES} often_{COMP} PV_{REL}·^{NAS}find_{3SG.PRES.PASS} PV·find_{3SG.PRES.PASS}
danō cid sō indhūathad
 yet even this_{NOM} the_{DAT.SG.NEUT}=rare_{DAT.SG.NEUT}
 ‘although it is oftener found . . ., yet even this is found rarely . . .’
 (Sg. 137^b2)
- (34) With equative:
islērithir insō nonguidimse
 COP_{3SG.PRES}=zealous_{EQ} the_{ACC.SG}=this_{ACC} PV.^{NAS}beseech_{1SG.PRES}=1SG
dīa nerutsu amal
 God_{ACC} for_{2SG}=2SG as
 ‘I beseech God for thee as urgently as . . .’ (Wb. 27^d19)

(35) With equative:

is *soirbidir sin forndengatsom*
 COP_{3SG.PRES} easy_{EQ} that_{ACC} PV·^{NAS}oppress_{3PL.PRES=3PL}
inni *bis*
 the_{ACC.SG.MASC=DEICT_{ACC}} be_{3SG.HAB.REL}
 ‘even so easily do they oppress him who is . . .’ (Ml. 75^{b7})

4.2.2 Orthographically ambiguous regarding nasalisation

4.2.2.1 Relativity marked otherwise

(36) *ba mmō immefolngitis brón damsa*
 COP_{3SG.PRET} more PV_{REL}·cause_{3PL.IMPF} grief_{ACC} to_{1SG=1SG}
 ‘[They] used more to cause grief to me.’ (Ml. 86^{d6})

(37) *ní lugu immefolngi sonartai do neuch*
 NEG less PV_{REL}·cause_{3SG.PRES} strength_{ACC} to someone_{DAT}
incothud indaas
 the_{NOM.SG.MASC=sleep_{NOM}} ‘than’
 ‘not less does sleep produce strength to a man than . . .’ (Ml. 135^{a13})

4.2.2.2 Relativity otherwise unmarked

(38) *doadbadar híc bríg inna persine*
 PV·show_{3SG.PRES.PASS} here might_{NOM} the_{GEN.SG.FEM} person_{GEN}
dodiccfa asmó dē focíaltar
 PV·3SG_{NEUT}·come_{3SG.FUT} COP_{3SG.PRES.REL}=more of_{3SG.NEUT} PV·expect_{3SG.PRES.PASS}
 ‘Híc is shown the might of the Person that will so come, who is the more expected.’ (Wb. 29^{c4}) (but see also as [86] below)

(39) *istraitiu adcotar fortacht dáe*
 COP_{3SG.PRES}=quick_{COMP} PV·obtain_{3SG.PRES.PASS} help_{NOM} God_{GEN}
 ‘the help of God is more quickly obtained . . .’ (Ml. 92^{b9})

(40) *combad mōu dē nocrete són*
 that-COP_{3SG.PST.SUBJ} more of_{3SG.NEUT} PV·believe_{3SG.PST.SUBJ.PASS} ANAPH_{NOM}
 ‘that it might be the more believed . . .’ (Ml. 111^{d4})

- (41) *is dēniu adciam hūasūlib risiū*
 COP_{3SG.PRES} quick_{COMP} PV·see_{1PL.PRES} from=eyes_{DAT.PL} before
rocloammar infogur hūachlūasaib ut est
 AUG·hear_{1PL.PRES.SUBJ} the_{ACC.SG.MASC}=sound_{ACC} with=ears_{DAT.PL} that is
is toisigiū adciam teilciud in bēla resiū
 COP_{3SG.PRES} first_{COMP} PV·see_{1PL.PRES} throwing_{ACC} the_{GEN.SG.MASC} axe_{GEN} before
rocloammar a guth sidi
 AUG·hear_{1PL.PRES.SUBJ} its sound_{ACC} ANAPH_{GEN}
 ‘we see more quickly with the eyes before we hear the sound with the
 ears, *ut est*, we see the throwing of the axe before we hear the sound of it’
 (Ml. 112^b12)

- (42) *nī mōa adcosnat*
 NEG more PV·strive_{3PL.PRES}
 ‘(they) do not strive more . . .’ (Thes. 2: 6.29 [Carlsruhe Augustine 12^d1])

- (43) With superlative:
fib as deg ropri[d]ched¹⁵
 as COP_{3SG.PRES.REL} best AUG·preach_{3SG.PRET.PASS}
 ‘as it hath been preached best’ (Wb. 23^a3)

4.2.3 Phonologically ambiguous regarding nasalisation

4.2.3.1 Relativity marked otherwise

- (44) *is mōo sluindes pronomen persin quam*
 COP_{3SG.PRES} more signify_{3SG.PRES.REL} pronoun person_{ACC} than
aliae partes

¹⁵ Strictly speaking, it has yet to be demonstrated if <p> in a nasalising context is ambiguous merely in orthography (= /b/) or also phonologically (= /p/). Caution in this regard is suggested by the behaviour, or perhaps merely representation, of *p* under lenition, where it is, or appears, “sometimes lenited, sometimes not . . . Evidently the process, which had developed by analogy with the other stops, particularly with *b*: *β*, had not yet become universal” (GOI § 231.5) for this originally foreign sound/letter. If the background to this is not merely graphical (cf. McManus 1983: 48, n. 63), but phonological, *p* may have shown similar hesitation initially to undergo nasalisation (a possible early instance of the marking of nasalisation may be seen in the doubling in *ippennit*, ‘in penance’ (Thes. 2: 247.8 [Cambrai Homily])).

other parts

‘The pronoun, more than the other parts of speech, signifies a person.’
(Sg. 197^a11)

4.2.3.2 Relativity otherwise unmarked

(45) *amal as trummu forlūadi hisuidi*
as COP_{3SG.PRES.REL} heavy_{COMP} PV·sway_{3SG.PRES} in=ANAPH_{DAT}
‘as it sways more heavily therein’ (Ml. 79^b5)

(46) a. *isléir dorigni indalalestar*
COP_{3SG.PRES}=careful_{NOM.SG.NEUT} PV·do_{AUG.3SG.PRET} one.of.two=vessel_{ACC}
‘It is carefully he has made one of the two vessels.’ (Wb. 4^c32)
(followed by [25b], repeated here as [46b])

b. *isdinnimu dongní alail*
COP_{3SG.PRES}=careless_{COMP} PV·^{NAS}do_{3SG.PRES} other_{ACC}
‘It is more carelessly that he makes the other.’ (Wb. 4^c33)

4.3 With positive adjective

4.3.1 With overt spelling of nasalisation

(47) *nību degming donet[h]adsom¹⁶ achorp*
NEG-COP_{3SG.PRET} difficult_{NOM.SG.NEUT} PV·^{NAS}go_{3SG.PST.SUBJ}=3SG_{MASC} his=body_{ACC}
fadesin issuidiu
own in=ANAPH_{DAT}
‘It was not difficult for him to go to his own body then.’ (Wb. 13^d20)

(48) *nīcumung donindnagar arforcital dúib*
NEG=straight_{NOM.SG.NEUT} PV·^{NAS}bestow_{3SG.PRES.PASS} our=teaching_{NOM} to_{2PL}
‘Not straitly [sic Thes.] is it that our teaching is given to you.’ (Wb. 16^a11)

¹⁶ Assigning this form to *do-etha* ‘goes to, visits, approaches’. On the other hand, Thes. 1: 726 (addendum 588) reports “MS. *donecadsom* (‘the he should see’), Chroust,” i.e. from *do-éccai* ‘looks at, beholds, sees.’ Finally, Kavanagh (2001: 348) takes the lead from Pedersen (1909–1913, 2: 514) in positing a verb *do-éta* ‘clothes’ (albeit intended by Pedersen implicitly as subj. only), i.e. ‘that he should clothe his own body therein’ (cf. *induere* in the Latin context). The principal point in the present context, the relative nasalisation, remains unaffected.

- (49) *nibrónach donintarrái*
 NEG=sad_{NOM.SG.NEUT} PV.^{NAS}return_{AUG.3SG.PRET}
 ‘It is not sadly¹⁷ that he has returned.’ (Wb. 16^b18)
- (50) *niba úaithed dondriga*
 NEG-COP_{3SG.FUT} few_{NOM.SG.NEUT} PV-^{NAS}3SG_{NEUT}·come_{3SG.FUT}
 ‘It will not be with a few¹⁸ that he will come.’ (Wb. 25^a38)
- (51) *iseícrichnichthe donindnigsom*
 COP_{3SG.PRES}=unlimited_{NOM.SG.NEUT} PV.^{NAS}bestow_{3SG.PRES}=3SG_{MASC}
adagmóini
 his=benefits_{ACC.PL}
 ‘It is unlimitedly that He bestows His benefits,’ (Wb. 28^a17)
- (52) *arndip maith nairlethar a muntir*
 so.that-COP_{3SG.PRES.SUBJ} good_{NOM.SG.NEUT} ^{NAS}care_{3SG.PRES.SUBJ} his household_{ACC}
 ‘that he care well for his household’ (Wb. 28^b32)
- (53) *mad ain[m]netach fondamtar inna*
 if-COP_{3SG.PRES.SUBJ} patient_{NOM.SG.NEUT} PV.^{NAS}suffer_{3PL.PRES.SUBJ.PASS} the_{NOM.PL.NEUT}
imneda inbetha frecndairc
 troubles_{NOM.PL} the_{GEN.SG.MASC}=world_{GEN} present_{GEN}
 ‘if the troubles of the present world be borne patiently’ (Ml. 46^d6)
- (54) *airis menic dondecmaing,*
 for-COP_{3SG.PRES} often_{NOM.SG.NEUT} PV-^{NAS}3SG_{NEUT}·happen_{3SG.PRES}
 ‘For it often happens thus.’ (Ml. 54^a7)
- (55) *amal as trait forídiuclannar ade*
 as COP_{3SG.PRES.REL} quick_{NOM.SG.NEUT} PV.^{NAS}devour_{3SG.PRES.PASS} ANAPH
 ‘as it is quickly devoured’ (Ml. 104^b5)

¹⁷ The interpretation as ‘sadly’ (implying the literal connection ‘sad how’), rather than ‘sad’ (literally ‘sad that’; see section 2.2), is confirmed by the Latin context: *abundantius magis gauisi sumus super gaudio Titi, quia refectus est spiritus eius ab omnibus uobis* (2 Cor. 7:13), ‘we rejoiced still more at the joy of Titus, because his spirit has been refreshed by you all’ (ESV).

¹⁸ Greene (1971) has shown that the DIL entry 1 *úathad/óthad/úaithed* is generally and originally an adjective meaning ‘few’.

asrochoilset

PV·determine_{AUG.3PL.PRET}

‘that they should speedily fulfil what they had determined’ (Ml. 95^c2)

- (63) *commixtum interpretatur .i. cummascdae adfét*
 commixtum interpretatur i.e. mixed_{NOM.SG.NEUT} PV·tell_{3SG.PRES}
in salmso di būaid innam
 the_{NOM.SG.MASC}=psalm_{NOM}=PROX of victory_{DAT} the_{GEN.PL.MASC}
babelóndae
 Babylonians_{GEN.PL.MASC}
 ‘i.e. this psalm speaks mixedly of the victory over the Babylonians.’
 (Ml. 115^d9)

4.3.3 Phonologically ambiguous regarding nasalisation

4.3.3.1 Relativity marked otherwise

- (64) *inbec máo .i. isbec as*
 the(?)=little_{NOM.SG.NEUT} more i.e. COP_{3SG.PRES}=little_{NOM.SG.NEUT} COP_{33SG.PRES.REL}
máo oldáusa .i. is bec
 more ‘than’_{1SG=1SG} i.e. COP_{3SG.PRES} little_{NOM.SG.MASC}
inderscugud,
 the_{NOM.SG.MASC}=distinction_{NOM}
 ‘A little greater i.e. she is a little greater than I, i.e. the distinction is small.’ (Sg. 45^a15)
 (gl. *paruo maior*¹⁹ in *paruo maior quam ego*, ‘a little greater than I.’)

In view of the standard teaching on the formation of adverbs,²⁰ Stokes and Strachan (Thes. 2: 99 n. c) wonder if for *inbec* “leg. *inbiuc*, or is *becmáo* a compound?”. Neither the emendation, however, nor the assumption of an unparalleled compound are necessary, if one (i) derives the element *in(d)* in adverbs not from the article, but from the preposition/preverb *ind(-)*,²¹ and (ii) takes account of the

¹⁹ In the manuscript, the gloss begins above the *p* of *paruo* and continues beyond *maior* into the empty space to the right of the column, whereas *quam ego* begins the following line.

²⁰ “To form an adverb, the dat. sg. of the adjective preceded by the article – or at all events by a word identical in form with the article – is generally used” (GOI § 379); cf. Mac Coisdealbha (1998: 155–157).

²¹ As considered in GOI (§ 379 note) and first argued by Morris Jones (1913: 439), cf. further Vendryes (1928), including his comment on the related Old Latin preposition *endo* that “elle

fact that the dative form is not found with all cases of this formation.²² Examples of what instead must be the accusative are listed in GOI (§ 379) itself, such as:

- (65) a. *indoll* ‘amply’ (Sg. 220^a6; author’s trans.), gl. *ultra* ‘beyond’, rather than **ind ull*—for the expected raised vowel cf. the comparative *huilliu*, e.g. Sg. 70^a6;
 b. *inmade*, *inmadæ* ‘in vain’ (Wb. 19^b10, 19^d16) gl. *sine causa* ‘without cause’—contrast the dative *mudu* (Wb. 16^d4);
 c. *ind immdae* ‘abundantly’ (Sg. 26^a5), gl. *examosin*²³—“beside normal” (GOI § 379) *indimdu* (Ml. 35^b5), gl. *passim* ‘in every part’.

For forms like *ind oll* and *ind immdae*, the article is clearly ruled out, since its accusative singular (masculine/feminine) form would be *in n-*. Instead, the possibility of accusative/directional adverbs (including with prepositions taking the accusative), beside dative ones, is further supported by the alternative formation with *co* ‘to’ (GOI § 381) – consider further the English alternative between ‘in a certain way’ and ‘to a certain extent’. Accordingly, the preposition *in(d)* could take either the dative or the accusative,²⁴ and *in bec* is to be taken as a regularly formed adverb – as an alternative to dative *inbiucc* ‘in small measure’ (e.g. Sg. 39^a25) – glossing *paruo* in isolation (followed by *máo* for *maior*), before the entire Latin phrase *paruo maior quam ego* is rendered in more natural Old Irish by the adjectival cleft sentence *is bec as máo oldáu-sa*.

4.3.3.2 Relativity otherwise unmarked

- (66) *nírbu* *faás* *foruigéni*
 NEG-COP_{AUG.3SG.PRET} empty_{NOM.SG.NEUT} PV·serve_{AUG.3SG.PRET}
 ‘Not void has been his service.’ (Wb. 13^b7)

gl. *et gratia eius in me uacua non fuit* (1 Cor. 15:10) ‘And his grace toward me was not in vain.’ (ESV)

devait admettre après elle le datif aussi bien que l’accusatif [it must have taken after it the dative as well as the accusative]” (Vendryes 1928: 78) and Lambert (1995: 174–175), the latter assigning the preposition *ind* the meaning ‘en direction de, contre [towards, against]’.

²² A fact that is not mentioned by Thurneysen in GOI despite being reflected in his collection of examples, or by Lambert, while some cases were at least pointed out in Thurneysen (1909: § 378).

²³ *Recte examosim* (Sg. 26^a8), for *examussim* ‘according to a rule or measure, exactly, regularly, perfectly’ (Glare [1968–1982] 2012: s.v. *examussim*). The gloss, therefore, is “probably guesswork, concluded from the context” (Hofman 1996, 1.2: 115).

²⁴ Like its allomorph *i n-* etc.; see GOI (§ 842).

- (72) *nípadro(mór)*²⁵ *notboctha*
 NEG-COP_{3SG.PST.SUBJ}=too.great_{NOM.SG.NEUT} PV-2SG.move_{2SG.PST.SUBJ}
 ‘Thou shouldst not boast overmuch.’ (Wb. 5^b32)
- (73) *menicc* *atchíth* *hi físb /*
 often_{NOM.SG.NEUT} PV-3SG_{NEUT}^{LEN}see_{3SG.IMP} in visions_{DAT.PL}
dosnicfed *afrithissi*
 PV-3SG_{FEM}^{NAS}come_{3SG.CND} again
 ‘Often he used to see in visions that he should come to it again.’
 (Thes. 2: 312.4 [Hymn ii])

In ascertaining if, in the adjectival cleft construction, a nasalising relative main verb is compulsory (as indicated in GOI §§ 383, 505), the first observation concerning the examples above is that no nasalisation, or indeed relativity, is marked in them, and *dom-rignis* and *not-boctha*²⁶ could serve unchanged to introduce a main clause, such as described, for instance, in GOI (§ 505) as an alternative to most other nasalising relative constructions. Accordingly, Pedersen (1899: 414) remarks on (71) that “das . . . zu erwartende relative *n* fehlt [that . . . for which the expected relative *n* is missing]”, while Mac Coisdealbha (1998: 155) similarly points to the fact that (72) “do[es] not show nasalization” as an argument against the presence of a nasalising relative. Moreover, Isaac in Mac Coisdealbha (1998: 257) compares (72) directly with (100) below:

As for [these] two [examples], the fact is that the opportunity for nasalization is there. 5b32 [= (72)] could have shown nasalization if it had contained the Class C infixed

²⁵ Vs. “*nípadruo*.., worauf zwei buchstaben etwa abgerieben wurden, Chroust [*nípadruo*.., after which about two letters have been erased]” (Thes. 1: 725 [addendum 528]). Previously, Strachan (1899: 42) had suggested reading *nípa[d] dron* . . . The images both in the facsimile (Stern 1910) and online (<http://vb.uni-wuerzburg.de/ub/mpthf12/pages/mpthf12/9.html>) seem compatible with reading *nípadro* (with a gap between *d* and *r*) and at least one more minim (such as the beginning of an *n* or *m*). The anonymous reader cautions that there is hardly enough space for *ro[mór]* in the manuscript and suggests that the intended adverb may be *rom* ‘early, too soon’, and furthermore that rather than taking *boccaid* plus reflexive infixed pronoun with the unparalleled meaning ‘boasts’, one of the more common meanings ‘softens’ or ‘moves’ (DIL) may be intended – thus implying, for instance, a literal ‘that it not be too soon that you move yourself’.

²⁶ Since (73) is not preserved in a contemporary Old Irish manuscript, the infixed pronoun in *at-chíth* can be taken either as expressing Old Irish prolepsis (cf. GOI § 421) or as showing (in this case scribal) Middle Irish petrification of neuter pronouns (cf. McCone 1997: 172–173). In the latter case, the Old Irish original might have had non-proleptic **ad-cíth*, with *c-* = /g/ marking the expected nasalising relative.

pronoun: **nípadromór nondatbocctha*. But it is perfectly regular for Class A, which cannot realize nasalization, to appear in this position, GOI § 413.2. 13a29 [= (100)] could have read **badféal et badfedte dongneid cachréit*. But it does not. In both cases, the opportunity to nasalize was not taken. Thurneysen's rule can, then, at best be taken as a facultative formal strategy.²⁷

Pace Isaac, however, his two examples represent two different types of exception. In *bad fedte do-gneid*, the opportunity to nasalise is indeed there and was not taken, meaning that this and some similar cases constitute real exceptions to the predicted nasalising relative construction and will be addressed as such in section 6. In cases like (72), on the other hand, that opportunity would have had to be created first by switching the infixed pronoun from class A to class C, and it is the opportunity of marking the relative by using class C that was not taken, not that of applying nasalisation, which is formally impossible with class A. In order to assess the latter type of exception, therefore, it will be necessary first of all to review the use of infixed pronouns in relative clauses.

Thurneysen (GOI § 413.2) observes that in relative verbs, class C “regularly replaces the pronouns of class A in the third person only; but it is frequently (though not invariably) used instead of the first and second persons of A and all the forms of B”. From this it follows – as is in fact conceded by Isaac himself above – that in a relative context, a form with a class A²⁸ or B pronoun is to be considered just as relative in status as one with class C, albeit without any overt marker of relativisation, which can only be expressed – by lenition or nasalisation – on the *-d-* of a class C pronoun (for more discussion on some of these points, see García-Castillero's contribution to this volume).²⁹ To test this, one needs of course to rely on constructions that are unambiguous in requiring a relative verbal form, and the clearest, and entirely undisputed, case in this

²⁷ The same direct comparison is implicit in Ó hUiginn's (1986: 58) more complete collection of formally deviant examples from the glosses, comprising, on the one hand, (71) and (72), as well as, on the other, *arachrin* in (96) and *dogneid* in (100) and some others of the same nature (on which see further sections 6.1 and 6.2 below), leading him to summarise that “it would be more accurate to see the use of the nasalizing relative in such cases as more of a normal custom than a fast rule”. Note, however, that while Ó hUiginn (1986) classifies verbs with non-Class C pronouns as non-relative throughout his collections (e.g. p. 44), his reference (p. 67) to “instances of a class A inf. pron. . . . being retained in a rel. clause” implies – in my view correctly – that a non-class C pronoun alone cannot serve to prove that the verb is non-relative.

²⁸ Compare Strachan (1903a: 67, n. 3): “It does not seem to have been noted that, when the short forms of the infixed pronouns of the first and second persons appear in relative use, relative *-n-* is not inserted before them.”

²⁹ While lenition of *d* cannot be marked within the Old Irish orthographical system, its presence can be deduced from the parallel relative nasalisation yielding written *-nd-* (GOI § 504 [b]).

regard is the leniting relative following a subject antecedent. The first two examples below, (74) and (75), illustrate that while a class C pronoun is admissible in order to express both relativity and lenition (on the *d-* of *-dam-*), it is not compulsory for this person (first singular), and class A *-m-* may be used as well, without affecting the underlying syntax:

- (74) *indí* *fodamsegatsa*
 the_{NOM.PL.MASC}=DEICT PV-1SG·afflict_{3PL.PRES}=1SG
 ‘those who afflict me’ (Ml. 33^a19; author’s trans.)

- (75) *Isiress* *crīst* *nombēoigedar*
 COP_{3SG.PRES}=faith_{NOM} Christ_{GEN} PV-1SG·vivify_{3SG.PRES}
 ‘It is Christ’s faith that quickens me.’ (Wb. 19^a20)

Example (75) is a subject-fronting cleft sentence with mandatory leniting relative connection, so it would not be useful to classify the verb *nom-bēoigedar* ‘which enlivens me’ as non-relative merely because neither relativity nor lenition could be expressed or realised on the surface. Rather, this is a leniting relative clause in status, but without the possibility of relative marking or lenition because of the choice of the more unmarked class A for the infixed pronoun. Furthermore, (76) may serve to illustrate Thurneysen’s rule that within class A, only non-third persons are admissible in relatives:

- (76) *nītú* *nodnai(l)* *acht ishé*
 NEG=thou PV-3SG_{MASC}(C)·^{NAS}nourish_{3SG.PRES} but COP=he
not ail
 PV-2SG(A)·nourish_{2SG.PRES}
 ‘It is not thou that nourishes it, but it that nourishes thee.’ (Wb. 5^b28)

Again, both parts of this sentence are subject clefts, entailing mandatory leniting relatives,³⁰ but the relative is marked only by and on the third person singular masculine class C pronoun *-d-n-* in the first verb. This is not formally possible with the second person singular class A pronoun *-t-* in the syntactically

³⁰ “Altogether distinct from this is the use of a non-relative form in the second of two parallel relative clauses, a construction found in many other languages” (GOI § 505 note), referring to *amal as toisegiu grían indáas laithe , is laithe foilsigedar cech rét síc is toissigiú* ‘as the sun is prior to the day, and it is the day that makes clear every thing, so . . . is prior . . .’ (Ml. 85^b11), which shows a pairing of two (underlying) relatives, not of two cleft sentences, each of which contains a relative.

parallel second verb, but *not-ail* must nonetheless be classified functionally as a leniting relative, and not as a non-relative verb.

Similar considerations apply to the use of class B in relatives.³¹ For this, it is instructive to begin by comparing the only other construction in which “a nasalizing relative clause can[not] be replaced by a formally independent (i.e. principal) clause” (GOI § 505), namely the *figura etymologica* connecting a verbal noun with its own verb via an adverbial nasalising relative (literally ‘where-with’, ‘by which’), as described in GOI (§ 499). Against McCone’s (1980: 23–24) objection, that some examples in Wb. and Sg. show a leniting relative instead,³² Stüber (2010–2012: 235, 240) not only follows Ó hUiginn (1983: 123–124) in distinguishing three syntactic types of *figura etymologica*,³³ but also demonstrates that the perceived exceptions are limited to the first two types, i.e. to object or subject-antecedent constructions that allow or indeed require a leniting relative also elsewhere. Among the strictly ‘third-type’, i.e. adverbial relative *figurae etymologicae*,³⁴ Stüber (2010–2012: 252–254) observes merely two deviant cases:

(77) *frissan* *ingrain* *ataroigrainn* *saul*
 to-the_{ACC.SG.NEUT} persecution_{ACC} PV-3PL(B)-persecute_{AUG.3SG.PRET} Saul_{NOM}
 ‘as to the persecution wherewith Saul persecuted them’ (Ml. 30^b2; emended reading by Griffith and Stifter 2014: 58–59)

(78) *dond* *fritobairt* *mail* *fritataibret*
 from=the_{DAT.SG.FEM} opposition_{DAT} slow_{DAT.SG.FEM} PV-3PL(B)-oppose_{3PL.PRES}
nadorche *donsoilsi*
 the_{NOM.PL.NEUT}=darknesses_{NOM.PL} to-the_{DAT.SG.FEM}=light_{DAT}
 ‘from the slow opposition with which the darkness opposes itself to the light’ (Sg. 183^b3)

³¹ Apart from preverbs ending in a dental such as *ad-*, where both **að-ð-* with class C and **að-d-* with class B yield *at-* /*ad-* through homorganic delentation; cf. *massuthol atomaig*, ‘if it is desire that drives me’ (Wb. 10^d26 (*ad-aig*)).

³² See with further examples Ó hUiginn (1983: 123, n.2), and cf. Ó hUiginn (1986: 34).

³³ “The first of these is that in which the antecedent acts as the grammatical subject of a passive verb . . . In the second type the antecedent functions as the object of an active transitive verb”, while in the third type, “the antecedent verbal noun” is taken up by a “following rel. clause which already has a subject or object” (Ó hUiginn 1983: 124). The latter “tripartite type” (Ó hUiginn 1983: 124) alone necessarily figures the adverbial relative connection under consideration here.

³⁴ Termed “intrumentalisch [instrumental]” by Stüber (2010–2012: 231, 245).

Stüber (2010–2012: 253) considers “dass in diesen zwei Beispielen offenbar keine Relativsätze vorliegen [that these two examples apparently do not involve relative clauses]”, but concedes the alternative that these are examples of “der seltene Fall [the rare case]” of a class B pronoun used in a relative clause, for which she adduces the following parallel:

- (79) *ishē danō cotammidethar*
 COP_{3SG.PRES}=he then PV-3PL(B)·power_{3SG.PRES}
 ‘It is He then that hath power over them.’ (Ml. 17^b2)

All three of these examples clearly show a class B pronoun in a relative verb, contrast the same preverbs with class C, first in the same verb *in-greinn* as in (80), then in *fris-gair* (81) and finally in *con-ocaib* (82) – (80) and (82) feature nasalising relatives, (81) a leniting one:

- (80) *anindagreinnsiu*
 when-^{NAS}PV-(^{NAS})3PL(C)·persecute_{2SG.PRES}=2SG
 ‘when Thou dost persecute them’ (Ml. 36^d2) (for *aninda* manuscript has *anunda*)

- (81) *is ī Bē Fīnd friss doghair*
 COP_{3SG.PRES} she Bé ^{LEN}Fīnd_{NOM} PV-3SG_{FEM}(C)·^{LEN}correspond_{3SG.PRES}
 ‘It is [the name] *Bē Fīnd* that corresponds to her.’ (Bergin and Best 1934–1938, 158, § 23; author’s trans.)³⁵

- (82) *ancondammucbaitisse*
 when-^{NAS}PV-1SG(C)·exalt_{3PL.IMPF}=1SG
 ‘when they used to exalt me’ (Ml. 39^d11)

Compare further the following passage from an originally Old Irish text for the regular interchange in relatives between third persons of class C in place of non-relative class A and unadapted class B:

³⁵ On *fris-dog(h)air*, see also Thurneysen (1940: 28), who points out that the combination of preverbs in -s and a directly following *d-* of a class C pronoun in leniting relatives, i.e. *as-d-* and *fris-d-*, is not found in the glosses and is otherwise very rare. For *as-n/n-d-* in nasalising relatives, see Ml. 31^b22, 93^d14, 114^a7, 124^d7. For further discussion of the *fris-* with class C pronouns, see García-Castillero’s contribution to this volume.

- (83) *Is é dorósat na huili 7*
 COP_{3SG.PRES} he PV·create_{AUG.3SG.PRET} the_{ACC.PL.NEUT} all_{ACC.PL.NEUT} and
rodacruthaigestar 7 fodaloing ó nirt
 AUG-3PL(C)·form_{3SG.PRET} and PV-3PL(C)·sustain_{3SG.PRES} by might_{DAT}
a chumachtai. Iss é nodaaíl 7
 his power_{GEN} COP_{3SG.PRES} he PV-3PL(C)·nourish_{3SG.PRES} and
cotaói 7 nodafáiltigedar 7
 PV-3PL(B)·preserve_{3SG.PRES} and PV-3PL(C)·gladden_{3SG.PRES} and
nodasorchaigedar 7 cotamidethar 7
 PV-3PL(C)·illuminate_{3SG.PRES} and PV-3PL(B)·rule_{3SG.PRES} and
dodaraitchiúir 7 atanúigedar na
 PV-3PL(C)·redeem_{AUG.3SG.PRET} and PV-3PL(B)·renew_{SG.PRES} the_{ACC.PL.NEUT}
huili,
 all_{ACC.PL.NEUT}

‘He it is who has created all things, and who has formed them and who sustains them by the might of his power. He it is who nourishes and preserves and gladdens and defective marking of relative lenition has redeemed and renews all things.’ (Strachan 1907: 3, 8)³⁶

Like in (76), this is a (longer) sequence of subject clefts, entailing mandatory leniting relatives, but again the lenition can be realised on the surface only with the class C pronouns, not with the three cases of class B. That, however, does not mean that these three verbs in the chain are not functionally relative and could serve to define an exception to the mandatory nature of a leniting relative after a subject antecedent.

For the adjectival cleft under consideration, then, examples (71) to (73), by not offering the surface option of marking nasalising relativity because of their choice of the more unmarked infixed pronoun classes A and B, respectively, cannot therefore, it is true, serve as proof of the mandatory status of a nasalising relative in this construction. Neither, however, can they be cited as exceptions to it, since in form they are ambiguous as to whether they are main-clause verbs or functionally nasalising relative ones without overt relative marking.

³⁶ Given here in Strachan’s Old Irish restoration based on two manuscripts, but one manuscript copy agrees in all relevant detail concerning the infixed pronouns (Strachan 1907: 2), while in the second copy, only two of the pronouns have been corruptly transmitted (*rocru-thaigestar*, *donail*, see Meyer 1903: 242).

5.2 Syntactic raising due to embedding

- (84) *hiris innaní as deg*
 faith_{ACC} the_{GEN.PL.MASC} =^{NAS} DEICT_{GEN.PL.} COP_{3SG.PRES.REL} best
rochreitset hicrist
 AUG.^{LEN} believe_{3PL.PRET} in=Christ_{DAT}
 ‘faith of those who have best believed in Christ’ (Wb. 31^a6)

On this case, Pedersen (1899: 351) remarks: “*rochreitset* ist selbstverständlich mit *innaní* zu verbinden, *as deg* als adverbiale bestimmung zu *rochreitset* aufzufassen [Needless to say, *rochreitset* is to be construed with *innaní* and *as deg* to be taken as qualifying *rochreitset* adverbially].” And Thurneysen comments:

An amalgamation of relative constructions . . . is . . . found when a superlative is taken out of the relative clause and placed in front of it in periphrasis with a relative form of the copula . . . Here, however, against the rule in § 498, the second relative clause remains a leniting one. (GOI § 508)

and:

Here it is more probable that *innaní* is felt as the antecedent both of *as deg* and *ro-chreitset* (GOI 681, n. 126.)

As for the superlative “taken out of” the relative clause, this positioning of degrees of adjectival comparison is actually the only option to express them in adverbial function, as described above (section 4.2). But what is remarkable here is the use of a leniting relative in *ro chreitset*, since the basic sentence with a non-relative copula would be expected to be **is deg ro creitset*, literally ‘it is best how they have believed’, with a nasalising relative. After embedding this clause into the context of syntactic dependency from *inna n-í* ‘of those,’ however, *ro creitset* governed by *deg* was ‘raised’ to connect directly to the superordinate *inna n-í* to yield ‘of those who have believed’, with the normal leniting relative after a subject antecedent.³⁷ What makes this example particularly valuable is the retained singular *as* after the plural *inna n-í*, thereby showing the “amalgamation” referred to in GOI – rather than a full adaptation to **ata deg ro chreitset*, literally ‘who are best who have believed’ (cf. [89] below). A syntactically less complex case also adduced in GOI (§ 508) is:

³⁷ Thurneysen’s “remains” (GOI § 508) is accordingly to be understood as denoting the transference of the leniting relative from the now syntactically parallel *as* before it.

- (85) *asmaam* *rošechestar* *arsidetaid*
 COP_{3SG.PRES.REL}=most AUG·LENfollow_{3SG.PRET} antiquity_{ACC}
 ‘who has most followed antiquity’ (Sg. 208^b15)

This example again adapts non-relative **is maam ro sechstar*. Furthermore, Elliott Lash has suggested to me that what was listed above as (38), assuming an orthographically ambiguous nasalising relative, could instead be understood as raised, with defective marking of relative lenition, which is why it is here restated with new numbering:

- (86) *doadbadar* *híc* *bríg* *inna* *persine*
 PV·show_{3SG.PRES.PASS} here might_{NOM} the_{GEN.SG.FEM} person_{GEN}
dodiccfa *asmó* *dē*
 PV·3SG_{NEUT}·come_{3SG.FUT} COP_{3SG.PRES.REL}=more of_{3SG.NEUT}
foc[h?]ialtar
 PV·(LEN)expect_{3SG.PRES.PASS}
 ‘*Híc* is shown the might of the Person that will so come, who is the more expected.’ (Wb. 29^c4)

Compare finally the Modern Irish reinterpretation of this pattern, described as “double relative construction” by Ó Nolan (1920: 114–116):

- (87) *Is é Pól an t-aspal is mó a d’fhág scríofa.*
 COP_{PRES} he Paul the apostle COP_{PRES} more C leave_{PST} writings
 ‘Paul is the apostle who has left the most writings.’ (lit. ‘Paul is the apostle who is biggest/most who has left writings’) (Ó hAnluain 1999: 113–114, § 11.36 [1960: 125–126, §225]; author’s trans.)

In summary, these examples do not establish exceptions to the standard construction of adjectival clefts, because the expected nasalising relative would at any rate have been superseded by a leniting one only secondarily as the result of syntactic ‘raising’.

5.3 Mixed antecedents

A construction with mixed antecedents was already seen in (14) above, to be read as *ba in[d] fortgidiu 7 ba hi temul du-gníth saul . . .*, ‘it was covertly and it was in darkness that Saul . . . used to make . . .’ (ML. 30^a3), where the non-relative verb *du-gníth* follows the construction required by the surface-antecedent

closer to it, *hi temul*, i.e. the adverbial cleft without a relative main verb, while the first surface-antecedent was adapted to this construction by being turned into a fronted adverb that is not otherwise admissible. A different case of mixed antecedents may be found in:

- (88) *cid* *dían* 7 *cían* *nothéisinn*
 although-COP_{3SG.PST.SUBJ} swift_{NOM.SG.NEUT} and far_{NOM.SG.NEUT} PV.^{LEN}go_{1SG.PST.SUBJ}
 ‘though I went fast and far’ (Ml. 41^{d9})

According to GOI (§ 506), the leniting relative here is one of the rare innovative deviations from the nasalising relative rule, which will be addressed in section 6.2. However, while the first antecedent is the straightforward equivalent of an adverb in other languages – entailing an adjectival cleft **cid dían no-téisinn*, ‘though it was fast how I went’ – this is not the case with *cían*: while one can go *swiftly/in a swift fashion*, one does not go *in a far fashion*, but rather *a long distance*. This semantic difference necessitates understanding *cían* as a substantivised³⁸ adjective, with an ensuing object relative, and for the latter, a leniting relative is one of the two options, i.e. **cid cían no théisinn*, ‘though it was a long distance that I went’. Here, too, then, the relative construction agrees with the second, closer, antecedent.³⁹

5.4 Failure to mark relativity in a preverb

- (89) *iscían* *arfolmas* *dún*
 COP_{3SG.PRES}=long_{NOM.SG.NEUT} PV·undertake_{3SG.PRET.PASS} for_{1PL}
insin
 The_{NOM.SG.NEUT}=that_{NOM}
 ‘It is long since that has been destined (has been imminent) for us.’⁴⁰
 (Wb. 21^{a2}; author’s trans.)

³⁸ This is not the same as *cían* in its common use as a temporal noun ‘long time’ (cf. section 4.1 above, and also [89] below), inflected as an *ā*-stem rather than with neuter *o*-inflection more typical of substantivised adjectives. And while one does not go *in a far fashion*, one can of course go *in an extended fashion*, i.e. for a long time, which allows the option of an adjectival antecedent in section 4.1 above.

³⁹ When this paper was delivered at the original conference, Ruairí Ó hUiginn raised the valid objection that there are semantic limits to the pairing of mixed antecedents, so that disparate combinations he exemplified by ‘It was quickly and dinner I was eating’ are unlikely to occur. The same semantic discrepancy, however, is not found in the present case, nor is it in (14) above.

⁴⁰ Already listed as (23) above (see there for further details).

According to GOI (§ 493.4), “the pretonic prepositions **im(m)·** and **ar·** have disyllabic forms in relative clauses: **imme·** or **imm·**, **ara·** (arch. **are·**)”. From this, it would have to be inferred that *ar-folmas* in (90) is non-relative. However, while a secondarily shortened (syncopated) form *ar-* is limited to position before proclitic *ro*, relative *ar-* is occasionally encountered elsewhere, including at least once in *Wb.*,⁴¹ compare the following complete collection:

- (90) a. *ara·*: *ara-lēgthar* (9^{b3}), *ara-thá* (10^{b3}), *ara-foím* (13^{c24}),
ara-bágim-se (16^{d9}), *ara-clessid* (22^{d18}), *ara-neut-*
sa (23^{b27}), *ara-mbere* (28^{c11}), *ara-nethem* (31^{c17})
 b. *arru·*: *arru-dérgestar* (4^{c13}), *arro-dibaid* (11^{a19})
 c. *ara·r-* (other than *·ro-*): *ara-rethi* (6^{b22})
 d. *ara·ro-*: *ara-rograd* (3^{c25}), *ara-róit* (4^{b19}), *ara-rogartsom*
 (5^{c23}), *ara-rogartsom* (5^{c23}), *ara-róitmar* (9^{c10}),
ara-róit (9^{c10}), *ara-roéit* (24^{a32})
 e. *ar-ro-*: *amal ar-rograd* (9^{b13})—contrast 3^{c25} in (91d).
 f. *ar·*: *Ished* *ar-thá*
 for COP_{3SG.PRES}=it PV^{·LEN}remain_{3SG.PRES}
insō
 the=this_{NOM}
 ‘It is this that remains.’ (30^{d13})—contrast *ished insō ara-thá* (10^{b3}).

While (90e) is merely more likely to be relative than non-relative,⁴² (90f) is unambiguously relative judging by both its context and the overtly marked relative lenition and shows – unless dismissed as a copying error – the incipient spread of the more common form *ar-* at the expense of relative *ara-*. Accordingly, *ar-folmas* in (89) could be taken as an innovative variant (cf. Ó hUiginn 1986: 65) for overtly relative *ara-folmas*.

⁴¹ See GOI (§ 493.4, note) and, for *Ml.* in particular, Strachan (1903a: 68) concerning cases of *ar-* before *ro*; his collection does not differentiate between stressed and proclitic *ro/ru*, and furthermore Ó hUiginn (1986: 65–66).

⁴² Compare the findings summarised by Ó hUiginn (1986: 56), according to which “where the verb which follows *ama(i)l* is not in the past subjunctive mood, relative marking is normal”, accounting for seventy-seven cases in the glosses, as against fifteen classified as non-relative – six of the latter, however, merely show a non-third person class A infix pronoun (*Wb.* 2^{a11}, 14^{b13}, 16^{a2}, 17^{b10}, 27^{d19}, *Ml.* 53^{b18} [if emended correctly]), a feature that is here rather taken as inconclusive as to relativity (see section 5.1), and only 16^{a2} clearly attests to non-relative status by using the independent negative (*amal ninfessed*).

- (92) *ishé cruth insō æm nosmessammar*
 COP_{3SG.PRES}=he way_{NOM} the=this_{NOM} truly PV-3PL(A):judge_{1PL.PRES}
 ‘This, truly, is the way we shall judge them.’ (Wb. 9^c10) (on *nosmessammar*, Thes. 1: 553, n. c notes, “leg. *nommessamar*?”)

- (93) *anduslēicet inna rind*
 when-^{NAS}PV-3PL(A):sink_{3PL.PRES} the_{NOM.PL.NEUT} planets_{NOM.PL.}
 ‘when the planets sink’ (Thes. 2: 11 [Carlsruhe Bede 18^b10])

The crucial question in order to assess (91) is whether the verbs in (92) and (93) are unambiguously relative. In Ó hUiginn’s (1986: 56–58) assessment, “(In) *C(h)ruth*” with a relative verb is found six times in Wb., three times in Ml. and five times in Sg., as against six non-relative cases in Wb. (including 92 above), but none in Ml. and Sg. However, while the present case features *cruth* in the nominative, after which an adverbial nasalising relative connection (‘by which’) should at least be an option, the other five non-relative cases all show the dative *in chruth-sin* ‘in that way’ (Wb. 3^d27, 18^b16, 24^a17, 24^b13) or *in chrud-so* ‘in this way’ (Wb. 31^c11) fronted in a cleft sentence, i.e. with an adverbial surface antecedent, after which a non-relative verb is the norm anyway in Old Irish. While the difference between *cruth* and *in chruth* is duly pointed out by Ó hUiginn (1986: 56–57, cf. 65, 66), he includes the five examples with adverbial *in chruth* because the nasalising relative occurs once:

- (94) *hóre isinchruthso rumboī*
 because COP_{3SG.PRES}-the_{DAT.SG.MASC}=way_{DAT}=PROX AUG.^{NAS}be_{3SG.PRET}
dossom
 to_{3SG.MASC}=3SG_{MASC}
 ‘because it is thus that he has been’ (Wb. 33^b1)

However, the latter is a special case due to syntactic raising (cf. section 5.2, as well as possibly [108] below): while the pattern of an adverbial cleft regularly demands a non-relative verb, i.e. **is in chruth-so ru boī do-ssom*, the entire construction is here embedded into an *hóre* sentence, and the second part is then raised to depend directly on *hóre*, using the option of a nasalising relative (cf. e.g. *hūare romboī*, ‘because it was,’ [Ml. 48^d8]).

As for *a*’ ‘when,’ Ó hUiginn (1986: 46–47) counts five cases with a relative verb from Wb., one hundred and twelve from Ml. and fourteen from Sg., as against seven non-relative ones from Ml. and one from Sg. Of these, however, five merely

the double-article rule in *ishé cruth* . . . can be taken as additional evidence for a close, relative connection of the following verb.⁵¹ Accordingly, while *corrup léir roscmallathar* in (91) could be counted as a rare exception in deviating from the requirement of a (nasalising) relative in an adjectival cleft, it could also be taken as a similarly rare case of the third-person (-s- in particular)⁵² class A pronoun in a functionally (though formally unmarked) nasalising relative verb. Additional support for the second interpretation may be seen in the fact that (91) (= *ML. 129^{b2}*) is paired with *129^{b1}* – with these two glosses glossing two parallel Latin phrases – and the latter shows an overt nasalising relative: *coru[p]léir dúngné nech inprecept* (*ML. 129^{b1}* = [57]) above. A second isolated example featuring a third-person class A pronoun is:

- (96) *is dēnithir sin arachrin cumachtae*
 COP_{3SG.PRES} swift_{EQ} that_{ACC} PV-3SG_{NEUT}(A).^{LEN}perish_{3SG.PRES} power_{NOM}
innapecthach
 the_{GEN.PL.MASC}=sinners_{GEN.PL}
 ‘Even so swiftly does the power of sinners perish.’ (*ML. 57^{c12}*; also listed as an exception by Ó hUiginn 1986: 58, n. 35)

This is to be contrasted with the explicitly relative class C pronoun e.g. in:

- (97) *amal arindchrim dáé*
 as PV-3SG_{NEUT}(C).^{LEN}perish_{3SG.PRES} smoke_{NOM}
 ‘as smoke perishes’ (*ML. 57^{a10}*; cf. *Wb. 27^{b1}*, *32^{c10}*, *ML. 85^{d10}*).

The apparent counter-examples are the following:

- (98) *fobithin arachiurat*
 because PV-3SG_{NEUT}(A).^{LEN}perish_{3PL.FUT}
 ‘because they will perish’ (*ML. 59^{b9}*)

⁵¹ See Uhlich (2013) (*pace* GOI § 471).

⁵² In this connection it is significant that, however rare -s- is in relative clauses in Old Irish, this very use must have formed the basis for -s- developing into a mere relative marker in Middle Irish, for which see Strachan (1904: 169–170). And if accordingly, -s- is viewed as a low-register colloquialism (of the kind described by McCone 1985), it is paralleled by what appears to be an early case of hypercorrect use of *ro* for *no* (for this more widespread feature in Middle Irish, see Breatnach 1994a: §§ 11.4–11.5; McCone 1997: 189–190, 197).

- (99) *intan aracrínat*
 when PV-3SG_{NEUT(A)}·^{LEN}perish_{3PL.PRES}
 ‘when [they] perish’ (Ml. 73^c2; GOI § 423)

They differ from (96) in that they are readily explicable by the general exception described in GOI (§ 505).⁵³ As for (96), this gloss also contains the same construction with a clear relative construction in *amal as ndian ade 7 as ngair mbís* ‘as it is swift and lasts for a short time’, Ml. 57^c12 (= 107 below), even though in the latter, *mbís* could also have been created by syntactic raising to be construed directly with *amal* (see section 5.2, and example 94). At any rate, a special explanation is required for (96), and a plausible factor to have caused the use of class A *-a-* rather than class C *-(n-)d-* may be seen in the fact that *ara·ch·rin* is one of a handful of Old Irish verbs that feature a petrified infix pronoun, whose meaning and function ceased to be understood synchronically, so that the basic lexeme could be viewed, and generalised, as *ara·ch·rin* rather than **ar·crin*.⁵⁴

6.2 Absence of feasible relative nasalisation

The considerations presented in section 6.1 leave the following three cases:

- (100) *badféal* *et badfedte*
 COP_{3SG.IMPV}=faithful_{NOM.SG.NEUT} and COP_{3SG.IMPV}=honorably_{NOM.SG.NEUT}
dogneid cachréit
 PV·do_{2PL.PRES} every_{ACC.SG.MASC}=thing_{ACC}
 ‘Let it be faithfully and let it be honorably [?] that ye do everything.’
 (Wb. 13^a29)

- (101) *cid* *toisigiú doberthar*
 although-COP_{3SG.PRES.SUBJ} first_{COMP} PV·do_{3SG.PRES.SUBJ.PASS}

⁵³ While relative *hore arinchrinat* ‘because they decay’ (Wb. 27^b1) is of course also possible.

⁵⁴ See GOI (§ 423). An alternative explanation is implied by Thurnysen’s description (GOI § 423) of *intan aracrínat* in (99) as simply ‘without *d*’, i.e. as if *ara-* were here the relative form of the simple preverb without any pronoun attached. However, this is contradicted by the lenition spelled out in (98), which points not to a nasalising relative as an option after such conjunctions, but to the presence of a neuter pronoun. Thus, *leg. ara-c[h]rinat* with standard defective spelling.

indfochaid

the_{NOM.SG.FEM}=tribulation_{NOM}

‘though the tribulation is inflicted first’ (Ml. 19^b11) (see Thes. 1: 716 [addendum 26.24])

- (102) *airismōu* *ruicim* *les* *mairchissechtae* *indaas*
 for-COP_{3SG.PRES}=more PV-reach_{1SG.PRES} need_{ACC} my=pitying_{GEN} ‘than’
 ‘for I need to be pitied rather than . . .’ (Ml. 22^c14) (cf. also section 4.2; for *airismōu* manuscript has *airimmou*)

In all three examples, “the opportunity to nasalize was not taken” (Isaac, in Mac Coisdealbha 1998: 257, commenting on [100]). Thus, these adjectival cleft sentences clearly do not use a nasalising relative, but the question is, what do they use instead? If Thurneysen’s observation were applied, according to which “a nasalising relative clause can be replaced by a formally independent (i.e. principal) clause in almost every instance” (GOI § 505), all three main verbs above would have to be taken as non-relative. As Ó hUiginn (1986: 66) has shown, this type of systematic exception is predominantly found with the copula. Furthermore, since none of these forms can be proven by their orthography to be non-relative,⁵⁵ the alternative is to take them as leniting relatives. Rather than systematic exceptions, then, these may be isolated examples of the incipient “extension of the leniting at the expense of the nasalizing relative” that will be completed by the end of the Middle Irish period (Ó hUiginn 1986: 70; see also 69–75 for more details; also GOI § 506).

7 Conclusions

Having thus reviewed almost the entire evidence for the construction of adjectival cleft sentences – apart from one special environment to be addressed in the appendix – it emerges (section 4) that the vast majority of examples either shows overt marking of relative nasalisation, or the spelling is at least compatible with this construction – be it ambiguous merely orthographically⁵⁶ or also phonologically.⁵⁷ Section 5 addresses a number of formal deviations that are

⁵⁵ Contrast, for instance, the unambiguous difference, with a simple verb, between relative *hōre pridchas* ‘because he preaches’ (Wb. 7^b15) (vs. non-relative *pridchaid*) and non-relative *hōre pridchīm* ‘because I preach’ (Wb. 5^c6) (vs. relative *no pridchīm*).

⁵⁶ Such as *adcotar* in (39), where the <c> can represent non-relative /k/ or nasalised /g/.

⁵⁷ Such as *dorigni* in (46), where the /r/ is not capable of being nasalised.

argued to be due to additional factors (syntactic raising, mixed antecedents), formal innovation (relative *ar-* replacing *ara-*) and, to begin with, the use of non-class C infixed pronouns. These are formally incapable of marking relativity,⁵⁸ but are regularly admissible (except for class A third persons) nonetheless in indisputably relative contexts like the mandatory leniting relative after a subject antecedent. Therefore, their presence alone cannot serve to prove that the verbal form that contains them is functionally non-relative. In other words, while e.g. third person singular *beires* ‘who carries’ is exclusively relative vs. *beirid* ‘carries’ exclusively non-relative, the same distribution is valid for, say, nasalising relative *do-mbeir* ‘by which . . . brings’ vs. non-relative *do-beir* (with /b/) ‘brings’, but not for the pair *dondom-beir* ‘by which . . . brings me’ vs. *dom-beir*, since the latter is found for both ‘brings me’ and ‘by which . . . brings me’. Relative forms like *dom-beir* may still, it is true, be classified as a type of exception, if viewed from the diachronic perspective described by Ó hUiginn (1986: 67):

It has been held that the creation of the class C inf. prons. represents a relatively late development in the prehistory of Irish and grew out of a need to formally distinguish rel. clauses . . . Pronouns of the first and second persons seem to have been much slower in adopting the new rel. forms . . .

In this scenario, the exception consists in not creating the opportunity for the formal marking of underlying relativity by introducing a class C pronoun. Relativity unmarked, however, does not equal non-relativity,⁵⁹ and synchronically, such forms will still not fall under the rule that “a nasalizing relative clause can be replaced by a formally independent (i.e. principal) clause” (GOI § 505), as is the case e.g. for Thurneysen’s example *hóre ní-ro-imdibed* with its unambiguously non-relative negative *ní*. Therefore, the use of non-class C pronouns (apart from third persons class A) is here treated as formally inconclusive as to relativity and thus does not constitute an assured exception to the use of a nasalising relative in adjectival clefts.

This left only five possible counterexamples (section 6), but even these were argued to be explicable not as systematic exceptions, but as individual innovations affecting all types of nasalising relatives. The result of this study, then, is that Thurneysen’s ruling that adjectival clefts trigger a mandatory (underlying) nasalising relative is to be upheld in principle, because there is no unambiguous evidence that a main-clause verb could be used instead.

58 Relative lenition and nasalisation can only be expressed, and only nasalisation be shown in writing, on the *-d-* of a class C pronoun, cf. (26), (27), (32), (50), (54).

59 As it does not, either, with the phonologically ambiguous forms listed above, such as *dorigni* in (46).

Appendix: The main verb *attá* ‘to be’

A specific application of the adjectival cleft pattern remains to be considered, which is described by Thurneysen in a separate paragraph, namely “when the antecedent supplies the concept that constitutes the predicative nominative of the relative clause” (GOI § 500). One of his examples involves an adjective, as in (105) below, resulting in an adjectival cleft as discussed here, with the only difference that the main verb is ‘to be’, specifically the substantive verb – literally ‘it is X how Y is’, etc.⁶⁰ Still, two subtypes are to be distinguished:

- (103) *asmenic* *mbīs* *confitebor duatlugud*
 COP_{3SG.PRES.REL}=often_{NOM.SG.NEUT} ^{NAS}be_{3SG.HAB.REL} *confitebor* for=giving_{DAT}
bude· cīasu *gnāthiu* *dofōisitīn . . . sed*
 thanks_{GEN} although-COP_{3SG.PRES} common_{COMP} for-confession_{DAT} . . . *sed*
non est hīc
non est hīc
 ‘that *confitebor* is often for returning thanks, though it is more common for confession, *sed non est hīc*’ (Ml. 26^c4)

While all cases feature the substantive verb, this first example would do so even outside the cleft pattern. While structurally required by the prepositional predicate *du atlugud*, this is not immediately suggested by the copula continuation *cīasu gnāthiu dofōisitīn*, but confirmed by the similar reference in:

- (104) *biid* *didiu a* *confessio* *hīsin*
 be_{3SG.HAB} then the_{NOM.SG.NEUT} confession_{NOM} DEICT_that
dofōisitīn *pecthae biid* *danō domolad*
 for=acknowledging_{DAT} sins_{GEN.PL} be_{3SG.HAB} also for=praising_{DAT}
biid *danō do atlugud buide* *do fōisitīn* *didiu*
 be_{3SG.HAB} also for giving_{DAT} thanks_{GEN} for confession_{DAT} then
atāsom *sunt*
 be_{3SG.PRES}=3SG_{NEUT} here
 ‘that *confessio*, then, is wont to be to acknowledge sins, it is also to praise, it is also to offer thanks: here, then, it is to confess.’ (Tur. 58^a)

⁶⁰ To the cases listed below might be added those in section 4.1.1 involving *cīan*, if to be taken adjectivally.

This means that *cīasu gnāthiu dofoīsitin* in (103) must itself be an elliptical cleft sentence, for *cīasu gnāthiu mbís . . .* In the next three cases, however, the surface antecedent itself stands for the predicate of the following verb (as opposed to *menic* above with adverbial connection), which therefore must be used in place of the copula:⁶¹

- (105) *cid* *drūailnide* *m̄bes*
 although-COP_{3SG.PRES.SUBJ} corrupt_{NOM.SG.NEUT} ^{NAS}be_{3SG.PRES.SUBJ}
chechtar *indārann*
^{LEN}each_{NOM.SG.FEM} the_{GEN.DU.FEM}=two_{GEN.DU.FEM}=parts_{GEN.DU.}
isinchomsuidigthiu
 in=the_{DAT.SG.NEUT}=compound_{DAT.SG.NEUT}
 ‘though each of the two parts in the compound be corrupt’ (Sg. 202^{b3})

- (106) *isfaittech* *rondboīsom*
 COP_{3SG.PRES}=careful_{NOM.SG.NEUT} AUG^{NAS}_{3SG.NEUT}·be_{3SG.PRET}=3SG_{MASC}
nant *neque manebunt asrubart*
 NEG_{SUB}^{NAS}COP_{3SG.PRES} neque manebunt PV·say_{AUG.3SG.PRET}
 ‘He was careful that he did not say *neque manebunt*.’ (Ml. 21^{d4})

- (107) *amal as* *ndīan* *ade* 7 *as*
 as COP_{3SG.PRES.REL} ^{NAS}swift_{NOM.SG.NEUT} ANAPH and COP_{3SG.PRES.REL}
ngair *mbīs*
^{NAS}short_{NOM.SG.NEUT} ^{NAS}be_{3SG.PRES.REL}
 ‘as it is swift and lasts for a short time’ (Ml. 57^{c12})

In the absence of evidence that the copula itself could also be used as the main verb in this construction, the following case from the *Additamenta* in the Book of Armagh seems puzzling:

- (108) *fer. . . nadip* *rubecc*
 man NEG_{SUB}-COP_{3SG.PRES.SUBJ} too.little_{NOM.SG.NEUT}
nadipromār *bedasommæ*
 NEG_{SUB}-COP_{3SG.PRES.SUBJ}=too.great_{NOM.SG.NEUT} COP_{3SG.PST.SUBJ}-his=wealth_{NOM}

⁶¹ For this and other regular uses of *attá* in place of the copula, see GOI (§ 774) and Mac Coisdealbha (1998: 154–155).

‘a man . . . whose wealth would not be overlittle or overgreat’ (Thes. 2: 241.8–9 and 27–29) (cf. Bieler 1979: 176–177, § 13 [2];⁶² note also that the example is glossed according to the traditional interpretation.)

Here, *nadip* clearly contains the third person singular present subjunctive of the copula, but what is *bed(a)*? The sequence *bedasommæ* (folio 18^{rb}15) is reproduced as such in Thes., but edited as *bed a sommæ* by both Bieler (1979: 177) and Thurneysen (1949: 33), in which *bed* could be the third person singular past subjunctive (or conditional) relative of the copula. On the other hand, since all parallel instances above feature the substantive verb, one could instead assume a third person singular past subjunctive of *attá*, standing for *no beth/bed* with omission of *no*.⁶³ However, this would still leave a problem of concord, according to which the present subjunctive in *nadip* would be expected to be matched by the main verb of the cleft construction.⁶⁴ A solution may be proposed that will begin by considering the word *sommae*. This is given in DIL (s.v. 1 *sommae*, ‘riches, wealth’) as an *iā*-stem connected with *soim* ‘rich, wealthy,’ implying a standard abstract formation. Thurneysen, on the other hand, lists it as ‘Subst. neut.’ (1949: 103). Among the attestations in DIL for both *sommae* and its counterpart 2 *dommae* ‘poverty, scarcity,’ there is no evidence to ascertain the gender of either in Old Irish.⁶⁵ Moreover, besides the adjectives *soim* and *doim*, there are also 2 *sommae* ‘rich, wealthy’ and 1 *dommae* ‘poor, needy,’ and on the other hand there are also *sommatu* ‘wealth, luxury’ and *dommatu* ‘poverty, want.’ This makes a derivational relationship *soim* → abstract noun *sommae* and *doim* → *dommae* far less obvious, and if an original semantic difference between abstract *sommatu* ‘the status of being rich’ and concrete *sommae* ‘riches’, etc., can be assumed, *sommae* and *dommae* are neuter nouns substantivised directly from the homonymous adjectives. In this case, *sommæ* in (108) could be singular or plural, and for the use of the latter – including a concrete meaning – compare *marba sommai* ‘goods will be destroyed’ (Meyer 1894: 40.9.13). This in turn opens up the option of taking the verb in *bedasommæ* as plural, too, and I suggest the following derivation that solves

⁶² This phrase has no equivalent in the corresponding passage in the *Vita Tripartita*, see Stokes (1887: 188.26–27), Mulchrone (1939: lines 2221–2212).

⁶³ Cf. *combed hed nobed and*, ‘so that that should be there’ (Wb. 3^b10). For the omission of *no* see Kelly (1999). Mac Coisdealbha (1998: 154) lists this as one of three cases that show “the substantive verb in place of the copula” but does not explain the form.

⁶⁴ Apart from cases where the introductory copula is reduced to “the unmarked, neutral present tense” (see Mac Coisdealbha 1998: 144–145).

⁶⁵ Dinneen’s (1927) “soime . . . f., riches” could of course continue either gender.

both problems, that of the verb ‘to be’ and that of concord: assuming that *bedasommæ* is among the numerous cases in the *Additamenta* that reflect an Early Old Irish spelling, its original shape may have been **bede e somme*, involving a third person plural present subjunctive relative of *attá*, which in (Classical) Old Irish is spelled *bete*.⁶⁶ This was misunderstood by a subsequent, Old Irish scribe, who, instead of modernising it correctly to **bete a sommæ*, adapted it mechanically to *bed a sommæ*. And returning to the construction under discussion, a separate ‘it should not be too little nor too great that his riches are’ would be expected to be construed with nasalising relative, i.e. **níp rubecc níp romar (m)bete a sommæ*. That the *b-* is not nasalised may be due to one of two reasons: either, this text is simply too early for relative marking to have been analogically transferred to the initial of simple relative forms.⁶⁷ Or, as Elliott Lash has suggested to me, this may be yet another case of syntactic raising (cf. section 5.2 and [94]), in which expected *(m)bete* was raised to the level of the preceding genitival (leniting) relative *nadip . . .*, as if depending directly on the superordinate antecedent *fer*. Be that as it may, the resulting merged syntagm, which subordinates the adjectival cleft construction ‘it should not be too little or too great that his riches are’ to the genitival antecedent ‘a man (whose)’, cannot be represented literally in English. As possible approximations, however, the following may be suggested, namely either (a) ‘a man concerning whom it should not be too little or too great that his riches are (or: . . . too great whose riches are)’, thereby compromising on the genitival relative connection, or (b) ‘a man whose riches should be such that they are not too little or too great’, with a freer rendering of the clefting construction.

66 Cf. Wb. 10^c22. For the earlier spelling convention of mediae for non-initial voiced stops, cf. *scarde*, ‘who separate’ (Thes. 2: 247.18 and 39 [*Cambrai Homily*]), *adobragart* (Wb. 19^b5 [*prima manus*]), ‘has seduced you’ (cf. GOI § 31, note).

67 Cf. GOI (§ 495 [b]) for lenition, (§ 504 [c]) for nasalisation.

9 The “Cowgill particle”, preverbal *ceta* ‘first’, and prepositional cleft sentences in the Old Irish glosses

1 Outline of the problem

1.1 The “Cowgill particle”

The origin of the absolute / conjunct distinction in the Insular Celtic languages has generated more literature than any other phenomenon in the history of the languages. The theory currently enjoying the greatest regard, generally dubbed the “particle theory”, asserts that a second-position clitic particle, the “Cowgill particle”, was inserted in most sentences and is ultimately responsible for the variety of verbal endings, as well as numerous other phenomena, such as mutations or the lack thereof.¹ In the spirit of the particle theory, one can offer the following simplified derivations for the absolute / conjunct distinction (where E stands for the enclitic Cowgill particle):

Insular Celtic	Pre-Irish ^a	Primitive Irish	Archaic Irish	Old Irish
*bereti-E *nī-E bereti	*bereti-E *nī-E beret	*b'er'eθ'i *n'ī b'er'eθ	*b'er'eθ' *n'ī b'er'	beirid 'he carries' ni-beir 'he does not carry'
*to-E bereti *nī-E tobereti	*to-E beret *nī-E toberet	*to b'er'eθ *n'ī tober'eθ	*to b'er' *n'ī tober'	do-beir 'he gives' ni-tabair 'he does not give'

Figure 1: Simplified derivations for the absolute / conjunct distinction in Old Irish.

^aI am using Pre-Irish in a loose sense to indicate some time after the early *i*-apocope (McCone 1996: 100–102) but before other distinctly Irish sound changes.

While the problem is fiendishly complicated, the above derivations capture the essential facts of the particle theory, which has two basic elements: the involvement of a second-position clitic and an early *i*-apocope. Both elements have been

¹ The standard literature arguing for a single particle is Boling (1972), Cowgill (1975), Schrijver (1994), and Schumacher (2004).

accepted by some and rejected by others, but most authorities accept some version of both as having a role in the rise of the distinction.² The generally accepted form of the enclitic is now **eti* (Schrijver 1994; Schumacher 2004: 90–114), the form of which has been determined based on direct British Celtic evidence, indirect Irish evidence, and etymological considerations.

One of the challenges of the Irish evidence is that the particle **eti* is itself totally lost. The only clue to its original existence is the presence of additional morphological material on the verb in the absolute endings (absolute *beirid* < **bereti-[e]ti* vs. conjunct *·beir* < **bereti*) and the lack of otherwise expected lenition (main clause *do-cing* ‘strides’ < **to-[e]ti-kingeti* vs. relative *do-ching* ‘who strides’ < **to- \bar{i} o-kingeti*). Given this lack of direct evidence, any additional evidence for the presence of the Cowgill particle would be quite welcome. As it happens, there is a small group of forms, hitherto relatively unnoticed, which may offer support for the particle.

1.2 The preverb *ceta* ‘first’

A somewhat uncommon preverb *ceta* ‘first’ appears in the glosses. It is connected etymologically to Gaulish **Cintu* in *Cintu-gnatus* ‘first-born’, Middle Welsh *cynt* ‘earlier’, and Old Irish *cét-* ‘first, early’ (first member of compound, usually with nouns) and is reconstructed as **kentu* ‘first’ (Matasović 2009: 201, LEIA C-103; Evans 1967: 182; Zair 2012: 174). The Old Irish use of interest here is as a preverb:

- (1) *is hé céetne fer ceta-ru-chreti di*
 COP_{3SG.PRES} 3SG_{MASC} first_{NOM.SG.MASC} man_{NOM} PV-AUG·believe_{3SG.PRET} of
áis assiaë hi Crist
 folk_{DAT} Asia_{GEN} in Christ_{ACC}
 ‘He is the first man of the folk of Asia that had first believed in Christ.’
 (Wb. 7^b11)

² For instance, Kortlandt (1979) accepts the particle but rejects *i*-apocope, while McCone (1979) accepts the *i*-apocope (see also McCone 1978) but rejects the particle. Other theories on the origin of the absolute / conjunct distinction (e.g. Sims-Williams 1984, McCone 2006, and Isaac 2007) accept both elements, but operate with a series of second-position clitics rather than with a single particle. That the main-clause negation *ní* is itself made up of two particles **ne* + **est* is not relevant to the discussion of the Cowgill particle, since the development **ne est* > **nēst* > **nist* must be quite old, given the raising of **ē* to **ī*.

Table 1: Pretonic *ceta* ‘first’ in the Old Irish glosses.

Wb.		Ml.	
<i>ceta-ru-chreti</i>	Wb. 7 ^b 11	<i>cetid-deirgni</i>	Ml. 124 ^b 3
<i>cíatu-ru-chreitset</i>	Wb. 14 ^a 29	<i>citid-tucat</i>	Ml. 125 ^d 4
<i>ceta-thuidchetar</i>	Wb. 21 ^c 5	<i>cita-roichet</i>	Ml. 44 ^b 4
<i>setu-ru-bridach</i>	Wb. 26 ^c 4	<i>cita-rochet</i>	Ml. 86 ^d 19a
		<i>cita-cómmaised</i>	Ml. 39 ^c 15
		<i>cita-rogaib</i> ⁶	Ml. 38 ^c 3
<i>ad-cita-acæ</i>	Tur. 60	<i>fris-cita-comrici</i>	Thes. 2: 23.38
<i>cita-ru-oirtned</i>	Thes. 2: 241.16		

eighth century became / was on its way to becoming *cita* by Ml. in the late eighth or early ninth century.⁷

That the *e* in the first syllable of Wb. is regular seems reasonable from the point of view of Irish sound change. Insular Celtic **kentu-* would have given **kēdu-* and then **ked-* in pretonic position, and it is this that we find in Wb. *ceta*. We will return to the final vowel later. Somewhat puzzling is that the pretonic *e* in *ceta* did not give *a* as in *a* ‘his’ from Early Old Irish *e*. This may be due to analogy with the accented prefix *cét-* ‘first’ (or the fact that in *ceta*, the *e* was originally nasal, cf. the suggestion made in Lash 2017a).

A similar sound sequence is found in *etar* ‘between’ < **enter*. Comparing **kentu-* and **enter* is instructive, since both are found in pretonic position and both have a similar phonetic structure. Interestingly, however, their outcomes are not totally parallel. Table 2 gives the outcome of the sound sequence **ent* in both Wb. and Ml. in pretonic position.

It should be clear that Wb. shows no difference for *etar* vs. *ceta*, while Ml. shows a considerable one. In Griffith (2016a: Appendix) it is argued that the pretonic sequence *et* in *etar* was on its way to becoming *it*. The tendency can be

⁶ Note that this example could be analysed as *cita-ro-gaib*. The same could probably be argued for *cita-ro-chet* as well. The exact position of *ro* is not critical for the purposes of this paper, but the fact that the particle appears as *ro* is suggestive of its being accented, since pretonic *ro* tends to become *ru*, especially as a second pretonic preverb (see GOI § 101 for the rule and Stifter [2013] for extended discussion of it).

⁷ The exact date of the Milan Glosses is unclear, but it is generally placed in the late eighth to early ninth centuries (see Lash 2017a: 148 and references therein).

Table 2: Outcomes of pretonic **etar* and **kentu* in Wb., Ml., and Sg.⁸

Wb.	<i>etar</i>	<i>itar</i>	Ml.	<i>etar</i>	<i>itar</i>	Sg.	<i>etar</i>	<i>itar</i>
	31	1		60	4		28	10
	<i>ceta</i>	<i>cita</i>		<i>ceta</i>	<i>cita</i>		<i>ceta</i>	<i>cita</i>
	3	0		1	5		0	0

clearly seen in Table 2. The tendency is, however, much stronger for *ceta* than for *etar*. In Griffith (2016b), it was argued that the only plausible explanation for the difference is that *etar* was a (relatively common) preposition in Old Irish, while *ceta* was not. Schumacher (2012) has documented the frequent analogical interactions between accented and unaccented allomorphs of prepositions (see also Griffith 2016a: Appendix), and it seems that the accented conjugated prepositional forms of *etar* exerted enough analogical influence to slow the change of pretonic *etar* to *itar*. Since *ceta* was not a preposition, this analogical influence was not present, allowing for a much faster, presumably regular phonological change to *cita*. Since the presentation of Griffith (2016b), Lash (2017a) has made a strong case for a slightly different interpretation of these facts. He notes that the *i*-variants are concentrated not only in pretonic position, but in a specific subset of this position: in pretonic complexes (which he defines roughly as a pretonic position containing more than one element). His argument is persuasive and I accept most of his findings here. One of the advantages to his approach is that it can explain the differential rate of change from *etar* to *itar* and *ceta* to *cita* in a straightforward manner without recourse to analogical influence. Since the preverb *ceta* ‘first’ was largely, though not exclusively, found in relative verbs (García-Castillero 2014: 87–89), which are pretonic complexes, the preverb would

⁸ The vowel in the final syllable is written variably in the sources. As it is not important for the topic under discussion here, I have simply written it as *a* here (see the appendix in Griffith 2016a on the final vowel in *etar* / *itar*). The data for in Table 2 comes from Table 1 above for *ceta* and Lash (2017a: appendix, Tables A, B, and C) for *etar*. The relevant parts of Lash’s tables are the columns “prep(osition)” and “preverb”, with the further proviso that only verbal forms where the *etar* is pretonic are included here. I further only consider glosses to Priscian in the St. Gall column and thus exclude other minor gloss collections that appear in his Table B. Finally, I exclude Wb. 28^{b3} *etir fessin et dóini* ‘between himself and men’ because *etir* is a conjugated preposition and thus not relevant to pretonic position. I also include two examples of *etir* from gloss Ml. 97^{a7} where Lash notes only one.

be more likely to change its *e* to *i* than *etar*, which had a more balanced distribution in pretonic complexes and simplexes.

We will shortly have occasion to return to Lash (2017a), which focuses on the initial vowel of *ceta*, but now the final vowel of *ceta* < **kentu-*, which has remained outside the discussion, must be accounted for. The oddity is not so much that the final vowel is variously written as *i* (*catid-deirgni* [Ml. 124^{b3}]), *a* (*cita-roichet* [Ml. 44^{b4}]) or *u* (*cetu-ru-bridach* [Wb. 26^{c4}]), but rather that it is present at all. Despite Thurneysen (GOI § 73), final vowels should be lost in all preverbs. The only recognised exception is in relative position, where there is precedent for preverbs having an extra syllable. This is standard for the preverbs *ar* and *imm* (GOI § 493.4), as the contrast between the (a) and (b) examples below shows:

- (3) a. *ar·beir*
 PV·enjoy_{3SG.PRES}
 ‘he enjoys’ (Ml. 43^d14)
- b. *ara·m·ber*
 PV-REL·enjoy_{3SG.PRES}
 ‘that he enjoys’ (Ml. 69^a18)
- (4) a. *im·folngi*
 PV·make_{3SG.PRES}
 ‘it makes’ (Wb. 4d32)
- b. *imma·folngi*
 PV-REL·make_{3SG.PRES}
 ‘which causes’ (Wb. 16b8)

While Breatnach (1994b) has also found examples of the extra syllable with other preverbs in relative constructions, he notes that for these other preverbs, the extra syllable is not the rule but rather the exception:⁹

- (5) *asa·gusi*
 PV-REL·wish_{3SG.PRES}
 ‘who wishes’ (Ml. 61^b17)

⁹ For instance, there are two examples of the preverb *in(d)* with an extra syllable in relative construction. This might be expected, given that the preverb was **inde* / **eni*, but the preverb normally does not have an extra syllable in relative construction, and it takes class B infixed pronouns, not class A (as noted by Breatnach 1994b: 198). For the preverb **as*, the form *asa* even appears in non-relative position, as noted by Thurneysen (GOI § 834B). This fact will enter the discussion again below.

suggestion that leaves the rules intact is that the retained final vowel in non-relative forms of *ceta* is regular before the Cowgill Particle **eti*. The derivation can be posited as here:¹³

(8) **kentu-eti* > **kentu η eti* > **kēdo η eh* > **kēdo η '* > **kēdoi* > **kēde* > *ceta*

The important assumption of this derivation is that there was no elision of the **e* of **eti* after **u*. The standard assumption underlying the particle theory is that the **e* was elided after any vowel (Schumacher 2004: 98–99; see Jasanoff 1997: 152–153 for a possible – though analogically motivated – exception). No one seems have considered cases after **u*, however, probably because the vowel rarely appears in a position where it would be in contact with the Cowgill particle. As such, it might appear that this rule is *ad hoc*. There are a couple further verbs, however, which support the non-elision argued for here: the verbs *ocu-ben* ‘touches’ and *ceta-bí* ‘perceives, feels’.

The first of these verbs, *ocu-ben* ‘touches’, is fairly rare. The six forms attested in the glosses are all from ML. and are given below in Table 3:

Table 3: *Ocu-ben* in ML. (forms in bold are non-relative).

<i>ocu-biat</i>	126^b12	<i>occu-robae</i>	98 ^d 8
<i>ocu-bether</i>	53^b17	<i>ocu-bendar</i>	54 ^a 12
<i>nicon-rocmi</i>	76^a12	<i>nad-ocmanatar</i>	54 ^a 12

Note that the disyllabic pretonic preverb *oc(c)u* is both relative (2x) and non-relative (2x) and that the preverb ends in a historical **u*: **onku-* (GOI § 848; contra Matasović 2009: 299, who derives it from **onko-*).¹⁴ This verb thus gives

¹³ It should be noted that the precise developments here are uncertain. For example, it is not clear whether **u η* would fall together with **o η* even at this late period, as it did earlier (McCone 1996: 55). If it did not, then **kēntu η eh* > **kēdu η '* > **kēdui* > **kēdoi* etc. is the likely development, since there was no difference between **ui* and **oi* at this stage (Cowgill 1967: 135–137; Greene 1976: 39; Uhlich 1995: 15–16; Schrijver 2007: 362 n.12; see also Bisagni 2012: 14).

¹⁴ Pedersen’s explanation of the retained *-u* in *ocu-ben* as due to a third person singular neuter infix pronoun (Pedersen 1909–1913, 2: 298) seems unlikely. The verb is transitive, and we would not expect a meaningless infix pronoun with a transitive verb. Compare also the verbal noun with an objective genitive: *cíd cuit a ocmaide* ‘even as to touching it’ ML. 39^a10. The objective genitive with verbal nouns is normal for transitive verbs. While an infix pronoun with a verb that is inherently transitive is certainly possible, the resulting verb does not

added support to the suggestion that the Cowgill particle does not elide its vowel after **u*. Why the final vowel in *ocu·ben* is consistently *u* but more regularly *a* in compounds with *ceta* is uncertain. It may reflect the nature of the preceding consonant: velar [g] favours *u* while dental [d] is neutral (cf. McCone 2015: 127 for a comparable observation in the context of consonantal *u*-quality).

The second verb relevant for the question is *ceta·bí* ‘perceives’. As noted above, the preverb is **kanta* ‘along’ and is thus different to *ceta* ‘first’ < **kintu*. It is nonetheless relevant here. In relative verbs, **kanta* would have regularly given *ceta*, much like *ar* / relative *ara* and *imm* / relative *imma* and *imne* (as well as relative *ceta* < **kentu-jo-*). For non-relative verbs, while there are no other exact parallels of preverbs of a shape like **kanta*, **cet* would have been the most likely outcome. Evidence from the glosses (see Table 4) shows that the form of the pretonic preverb is always *ceta*, regardless of whether it is relative or not (non-relative forms are in bold):

Table 4: *Ceta·bí* in the glosses (forms in bold are non-relative).

	Wb.		Ml.
<i>ceta·biin</i>	12 ^c 8	<i>cita·m·bí</i>	36 ^b 1
		<i>cita·m·bé</i>	36 ^b 1
	Sg.	<i>cita·m·bénn</i>	44 ^c 15
<i>ceta·biat</i>	3 ^a 1	<i>cita·m·betis</i>	29 ^c 13
		<i>cita·biat</i>	22 ^d 7
		<i>cita·bé</i>	68 ^d 15
		<i>cita·roba</i>	44 ^b 22

Once again, there is not a large number of forms, but the relative and non-relative forms all have the final vowel *-a* (and incidentally, the vowel of the first syllable in Wb. and Ml. corresponds to the pattern for **kentu-* ‘first’). The likeliest explanation for the unexpected appearance of the final vowel of the preverb in non-relative forms of *ceta·bí* ‘perceives, feels’ is that it was taken over from

remain transitive but becomes intransitive (cf. *at·baill* ‘dies’ < ‘throws it’ or *at·reig* ‘rises’ < ‘raises himself’).

the otherwise identical *ceta* ‘first’, which has been argued to have regularly *ceta* in both relative and non-relative clauses.

If all of this is correct, we have some further, limited support for the second-position clitic main-clause particle in the (Pre-)Old Irish verbal complex. While this particle usually disappears without a trace of influence on the preverb it was attached to, it seems to modify the shape of the preverb in exactly one case: when the disyllabic preverb contained a **u* in the final syllable. There are only two such preverbs known to me: *ceta* ‘first’ < **kentu-* and *ocu* < **onku-*. The preverb *ceta*, found in *ceta·bí*, appears to have followed the pattern of *ceta* ‘first’ in non-relative contexts.

1.3 Cleft sentences with fronted prepositional phrases

As was noted above, the main verb in cleft sentences with fronted prepositional phrases should be non-relative according to the rules of Old Irish grammar (Strachan 1929: 123, n.7). A more complete account says that when the subject or object is fronted, the main verb is relative (see also GOI § 494, § 501), but when anything else is fronted, the main verb is non-relative unless the word must be followed by relative *-n-* (i.e. in temporal clauses, manner or degree clauses, *figura etymologica*, source or cause clauses, with adverbially used adjectives; see GOI (§ 383, §§ 497–502) and Uhlich’s contribution to this volume). The focus here is on fronted prepositional phrases (hereafter PPs), where non-relatives are expected:

- (9) *ar is do thabirt dígla berid in*
 for COP_{3SG.PRES} to^{LEN}bring_{DAT} revenge_{GEN} carry_{3SG.PRES} the_{ACC.SG.MASC}
claideb sin
 sword_{ACC} DIST
 ‘For it is for wreaking revenge that he carries that sword.’ (Wb. 6^a13)

As noted in GOI (§ 506; see also McCone 1985: 96; Ó hUiginn 1986: 63), relative verbs are occasionally found:

- (10) *Ni fris nu·chét a propheta*
 NEG-COP_{3SG.PRES} against_{3SG.MASC/NEUT} AUG-^{LEN}sing_{3SG.PRET.PASS} by prophet_{ABL}
 ‘It is not with reference to it that it was sung *a propheta*.’ (Ml. 64^a13)

- (11) *acht is do sochaidi no·pridchib*
 but COP_{3SG.PRES} to multitude_{DAT} PV·preach_{1SG.FUT}
 ‘but it is to a multitude that I will preach’ (Ml. 45^a8)

The question addressed here is how common the use of relatives in such sentences in the major gloss collections is.

2 Methodology

Examples of cleft sentences were collected inclusively (i.e. with a wide net) from the three major Old Irish gloss collections: Wb., Ml. and Sg. This included examples with fronted prepositions and adverbs (as there is significant overlap in function), as well as fronted subjects and objects. As it turned out, there are no exceptions to the rule that fronted subjects and objects are followed by a relative verb, and these sentences are not considered further here. After collecting the cleft sentences, numerous possible examples were excluded:

- as noted above, subject and object clefts as well as adverbial clefts were excluded;¹⁵
- examples without overt copula were discarded, since other interpretations are possible (see below in [12] for an example);
- examples with nasalising relatives according to GOI (§ 383, §§ 497–502) were set aside;
- some other examples were also excluded, such as noun phrases used adverbially: *in chruth so* ‘in this manner’ or *in méit sin* ‘in that size, so much’.

Two examples of excluded sentences are given here in order to show the types of considerations made during the analysis:

- (12) *per prophetas do·n·icfad cucunn*
 through prophets_{ACC.PL} PV·^{NAS}come_{3SG.CND} to_{1PL}
 ‘[It is?] through the prophets that He would come to us.’ (Wb. 21^a3)

¹⁵ The distinction between adverbial and prepositional cleft is made in formal terms. That is, a number of adverbs are formally conjugated prepositions and are considered prepositional phrases for purposes of this chapter.

gl. UT SIMUS IN LAUDEM GLORIÆ EIUS NOS, QUI ANTE SPERAUIMUS IN CHRISTO

‘in order that we might be in praise of His glory, we who previously hoped in Christ [i.e. through the prophets, that He would come to us]’

This example is excluded on principle because of the lack of a copular form at the beginning, but one should note that the relative is probably dependent on being in indirect discourse after Latin *sperauimus* ‘we hoped’, rather than being in a cleft sentence. This analysis fits with the lack of copula and illustrates why copula-less examples are excluded.

The next example is excluded because the fronted element is not a preposition, but it is interesting because the leniting relative seems out of place given the Thesaurus translation, to which a nasalizing relative would be more appropriate.

- (13) *is mó ro-chéess crist airi*
 COP_{3SG.PRES} more AUG.^{LEN}suffer_{3SG.PRET.PASS} Christ_{NOM.SG} for_{3SG.MASC.ACC}
 .i. *báas*
 i.e. death_{NOM}

‘It is more that Christ has suffered for him, i.e. death. [Therefore cast off the foods that you love].’ (Wb. 6^c8; trans. Thes.)

‘what Christ has suffered for him is greater, i.e. death. Therefore. . .’ (author’s translation)

The translation offered here makes clear that *ro-chéess* is a relative without antecedent (GOI §496; Ó Cathasaigh 1990) in a copular sentence (see also Uhlich’s contribution to this volume).

Having excluded various examples as indicated above, it remains to classify the prepositional clefts. Since the orthography of Old Irish is frequently ambiguous, the remaining prepositional clefts are coded as relative (14), non-relative (15), ambiguous (16),¹⁶ non-nasalising relative (17), and non-leniting relative (18):

¹⁶ There are several ways in which a form can be ambiguous. A form like *do-rat-side* (Wb. 23^c17) is ambiguous because *r* does not show mutations (a geminate spelling *rr* is not probative, since it might indicate nasalisation or lack of lenition). The case is similar for *f*, *l*, *m*, *n*, *p*, and *s*. A further type of ambiguity is due to the irregular use of Class C pronouns in relative contexts. While a Class C pronoun after a preverb is a sure indication of a relative verb, a Class A pronoun is in the analysis here only taken as indicative of a non-relative verb if the infixed pronoun is third person. The other persons are treated as ambiguous. Class B pronouns are treated here as non-relative, but since the orthography does not always distinguish [t] and [d], it is impossible to make a principled decision between the two classes in some cases (mostly following *r* with first and second person pronouns), as in *fordon-cain* (Wb. 31^c16).

- (14) *acht is do sochaidi no·pridchib*
 but COP_{3SG.PRES} to multitude_{DAT} PV·preach_{1SG.FUT}
 ‘but it is to a multitude that I will preach’ (Ml. 45^{a8})
- (15) *huare is hifochaidib bithir hisuidib*
 since COP_{3SG.PRES} in=tribulations_{DAT.PL} be_{3SG.HAB.PASS} in=that_{DAT}
 ‘since it is in tribulations that men are for them’ (Ml. 56^{b15})
- (16) *is airi do-roigu dia geinti*
 COP_{3SG.PRES} for_{3SG.NEUT.ACC} PV·AUG.choose_{3SG.PRET} God_{NOM} gentiles_{ACC.PL}
 ‘It is therefore that God has chosen the Gentiles.’ (Wb. 5^{b12})
- (17) *is airi as-berar*
 COP_{3SG.PRES} for_{3SG.NEUT.ACC} PV·say_{3SG.PRES.PASS}
 ‘It is therefore that it is said.’ (Wb. 3^{c21})
- (18) *is airi ro-cload*
 COP_{3SG.PRES} for_{3SG.NEUT.ACC} AUG·overcome_{3SG.PRET.PASS}
 ‘It is therefore that it has been overcome.’ (Wb. 3^{b1})

One category of verb that does not fit well into this system is that of contracted verbs:

- (19) *ní do dígail for firianu tuccad*
 NEG-COP_{3SG.PRES} for punishment_{DAT} on righteous_{ACC.PL} put_{3SG.AUG.PRET.PASS}
recht
 law_{NOM}
 ‘It is not for the punishment of the righteous that the law has been given.’
 (Wb. 28^{a3})
- (20) *Is do tra duic=sem a*
 COP_{3SG.PRES} for_{3SG.NEUT} then^{NAS(?)} put_{AUG.PRET=3SG.MASC} the_{ACC.SG.NEUT}
ndliged so
^{NAS}expression_{ACC} PROX
 ‘It is for this, then, that he has put this expression.’ (Ml. 115^{b15})
- (21) *is do thucad an ·una*
 COP_{3SG.PRES} for_{3SG.NEUT} LEN put_{3SG.AUG.PRET.PASS} the_{NOM.SG.NEUT} ^{NAS}una_{NOM}
 ‘It is for this that the *una* has been put.’ (Sg. 45^{b19})

While contracted verbs correlate somewhat with leniting relative clauses, there are many exceptions to this trend (see Schrijver 1997b: 113–128 for an analysis and McCone 2006: 87–90 for objections; García-Gastillero 2015 is the most recent contribution to this interesting problem). Under the system of classification adopted here, examples like (19) can be seen as ambiguous or, perhaps, non-nasalising relatives. Because they cannot be clearly categorised and may not be relative at all, they are left out of further consideration. There are six such examples: Wb. 7^a2, 24^b26, 28^a3, Ml. 62^a2, 71^c9, Sg. 161^a1.

Examples (20) and (21) are interesting because they appear to be examples of nasalising and leniting relatives of contracted verbs. There are two examples like (20): Ml. 56^c11 and 111^b15; and there are two examples like (21): Sg. 45^b19 and 77^b5. Though the two Milan forms could involve the writing of nasalised *t-* as *d-* in a nasalising relative, they are ultimately ambiguous: *duic* in 115^b15 could be a simple copying error or a late contracted form of *du-uic*. A parallel can be seen in gloss initial *duic* (Ml. 40^c22), where there can be no question of a relative form. The same explanation is available for *ducad* (Ml. 56^c11).

Forms like *thucad* in Sg. seem to follow the post-Wb. Old Irish tendency to lenite morphologically relative forms (GOI § 495). Nonetheless, it is notable that both examples appear in the sequence *is do thucad X* ‘it is for this that X was put’. The form *do* in these glosses is the third singular masculine or neuter of the preposition *do* ‘to, for’ and is etymologically the bare preposition which is taken over as the conjugated form. It would thus be expected to lenite what followed, and it is just possible that these two forms show this lenition. For a slightly inexact parallel one might compare *air thuccai* (Ml. 42^c8) or *ce thuc* (Thes. 2: 225.19 [Carlsruhe Glosses on Priscian]) for the lenition of the initial of a contracted verbal form. The upshot of this discussion is that the contracted verbal forms in the glosses do not appear to offer solid evidence for relative forms after clefted prepositional phrases.

3 The data: A first-pass analysis

Once the various exclusions noted above were carried out, the remaining examples were classified as above and tallied. The results are found in tables 5 through 7 below:

Table 5: Wb. main verbs after prepositional clefts (229 examples in total).

non-relative	120
ambiguous	45
non-nasalising	46
non-leniting	17
relative	1

Table 6: Ml. main verbs after prepositional clefts (265 examples in total).

non-relative	101
ambiguous	43
non-nasalising	93
non-leniting	18
relative	10

Table 7: Sg. main verbs after prepositional clefts (118 examples in total).

non-relative	62
ambiguous	16
non-nasalising	32
non-leniting	6
relative	2

This quick classification comparing clearly relative forms to clearly non-relative forms shows that relatives after clefted prepositional phrases are very rare in Wb. (< 1%), relatively rare in Sg. (approx. 3%), and uncommon in Ml. (approx. 9%).¹⁷ The examples classified as relative deserve a closer look.

4 The data: a closer look at relatives

4.1 The Wb. data (one relative example)

The single example classified as relative in Wb. is the following:

- (22) *is airi ꝥnabrúi(thea) in*
 COP_{3SG.PRES} for_{3SG.NEUT} PV.^(NAS?)break_{AUG.3PL.PRET.PASS} the_{NOM.PL.MASC}
gésce
 branches_{NOM.PL}
 ‘It is therefore that the branches were broken.’ (Wb. 5^b29)

This is not certainly relative, but at least possibly so. The question is how to interpret the spelling: *con-abrúithea* (non-relative) or *con-n-abrúithea* (relative). The following example makes clear that a non-relative form is possible in principle:

- (23) *rodbo dia ad·roni et ꝥnói*
 either God_{NOM} PV·make_{AUG.3SG.PRET} and_{LATIN} PV·preserve_{3SG.PRES}
 ‘It is either God who has made and preserves.’ (Wb. 29^d29)

Since this sentence contains leniting relatives, the spelling *ꝥnói* must stand for *con-ói*, which implies that *ꝥnabrúi(thea)* can stand for *con-abrúithea*, which is ambiguous in the classification here. This analysis removes the only example given above as relative in Würzburg, meaning that there are no examples of relative verbs after clefted PPs in the Würzburg Glosses.

¹⁷ One might compare the clearly relative forms against all others, in which case Wb. has approximately 0.5% relative forms, Sg. approximately 2%, and Ml. around 5%. This count given here is deliberately somewhat conservative, trying not to bias the discussion unnecessarily.

4.2 Sg. data (two relative examples)

The examples from Sg. are somewhat more interesting. This first of these is:

- (24) *is i foilsigud frechdairc*
 COP_{3SG.PRES} in demonstration_{DAT} present_{DAT.SG.MASC}
asa-gnintar i n-ego 7 tu. tri atarcud
 PV·recognise_{3SG.PRES.PASS} in^{NAS} ego and tu through anaphora_{ACC.SG}
immurgu asa-gnintar hi sui
 however PV·recognise_{3SG.PRES.PASS} in sui
 ‘It is in present demonstration that it is recognised in *ego* and *tu*. [It is] through anaphora, however, that it is recognized in *sui*.’ (Sg. 197^{b4})

Under discussion is the first example of *asa-gnintar*,¹⁸ which could be seen as relative (see Breatnach 1994b on prepositions with added vowels in relative compound verbs). On the other hand, note that the preverb is regularly *asa* for this verb in Sg. There are eight total examples in Sg., and six are certainly non-relative;¹⁹ the two examples above in (24) are probably not relative either.

The second example in Sg. of a relative verb in a prepositional cleft is the following:

- (25) *cid ardid hua thuislib ildaib*
 What for^{NAS} COP_{3SG.PRES} from^{LEN} cases_{DAT.PL} plural_{DAT.PL.MASC}
*disruthaigedar*²⁰
 (PV·)derive_{3PL.PRES.PASS}
 ‘Why is it from plural cases that they are derived?’ (Sg. 198^{b3})

eDIL takes the verb as a simplex (s.v. *dísruthaigidir*), in which case this is indeed a relative form. The evidence, however, makes it more likely that this is a compound verb *di-sruthaigedar*, in which case the form in (25) should be classified as non-leniting. Finite forms of the verb are non-probative as to the simplex / compound nature of the verb, since there are only two examples additional to the one above, and both of these are conjunct / prototonic: *hua-n-dirrudiged(d)ar*

¹⁸ The second is excluded according to the principles outlined above because there is no overt copula in the sentence, though in this case a cleft sentence seems clearly to be the correct analysis.

¹⁹ The examples are 29^{a3} (*bis*), 146^{b16}, 180^{b2}, 209^{b13}, and 210^{a10}.

²⁰ Thes. (2: 192) suggests reading *disruthaigeddar*, which is accepted here as the scribe’s intention.

‘from which they are derived’ Sg. 33^a23 and *ó-diruidichther* ‘from which it is derived’ Sg. 50^a1. Non-finite forms are more suggestive. The substantivised participle *dísruthigthe* ‘derivative’ appears twelve times, always with the spellings *dir(r)- / dír(r)-*.²¹ Similarly, the verbal noun *dísruthigud* ‘derivation’ appears seven times with the spellings *dir-* / *dír-* (4x), *dír-* (1x), and *dírš-* (2x).²²

It is clear from the attestations that spellings of the prototonic forms and nominal forms indicate the lenition and / or assimilation of the *s*, while the one form that could be deuterotonic (the form in (25) above) is also the only one with the spelling *disr-*, which indicates that the *s* is not lenited. It is therefore very likely that this verb (which generally translates Latin *derivatur*) is a compound verb. In the context of this chapter, it should be classified as non-leniting. That is, it is not in a leniting relative, but it could conceivably be a nasalising relative clause.²³

From the analysis of the two possible examples of relative verbs after clefted PPs in Sg., we have seen that neither is actually likely to be relative. That leaves us with no certain examples of relatives with clefted PPs in either Wb. or Sg..

4.3 Ml. data (ten relevant examples)

The Milan Glosses have a larger number of possible examples of relative verbs in prepositional clefts, and it will turn out that a number of them are indisputable, i.e. they cannot be explained away. It is necessary, however, to examine them in more detail, and it is useful to classify the relatives into nasalising relatives (three examples), leniting relatives (three examples), and ambiguous relatives (two examples), as well as two verbs for which the distinction is irrelevant because they are absolute relative forms, which do not distinguish leniting and nasalising contexts.

²¹ Loci: 8^b2, 28^a4, 33^a17, 56^b10, 59^b12, 61^a1, 188^a7, 188^a12 (bis), 188^a13, 188^a16, and 188^a19.

²² Loci: 36^b1, 51^a4, 53^a11, 188^a4 (bis), 188^a8, and 193^a1.

²³ The retention of pretonic *dī* might seem surprising, but this seems to be not uncommon in such learnedisms. One might compare: *do-tá* ‘differs’ glossing *distamus* (*di-taam-ni* [Ml. 117^b9]) and *deferre* (*di-tá* [Ml. 120^a6]); *do-samlathar* ‘compares’ glossing *disimulat* (*di-samlathar* [Ml. 21^b2]) and *disimulans* (*di-samlad* [Ml. 114^c3]); *do-meicceathar* ‘despises’ glossing *detero* (*de-mecimm* [Sg. 39^b1]); and *do-nochta* ‘lays bare’ (not *dínochtaid* as in DIL) glossing *denudatur* (*dí-nochtar* [Ml. 54^d23]). The tendency is not universal, however: *do-gaib* ‘diminishes’ glosses *deminuitur* (*do-ñ-gaibter* [Sg. 218^a9]); elsewhere *di-roghad* [Sg. 9^b16]). I have not found examples in Wb. of this sort of learnedism, which is probably not surprising since Ml. and Sg. have more numerous short glosses which calque the Latin, while Wb. has fewer such glosses.

4.3.1 Nasalising relatives

- (26) *is samlid inso as-m-bertar ut. . .*
 COP_{3SG.PRES} like_{3SG.NEUT} the_{ACC.SG=this_{ACC}} PV^{NAS}·say_{3PL.PASS} so.that
 ‘It is thus that they are said, in order. . .’ (Ml. 23^a12)
- (27) *is samlid insin imme-tét*
 COP_{3SG.PRES} like_{3SG.NEUT} the_{ACC.SG=this_{ACC}} PV-REL·travels_{3SG.PRES}
leu=som int ais lósc
 with_{3PL=3PL} the_{NOM.SG.MASC} people_{NOM} lame_{NOM.SG.MASC}
 ‘It is thus that cripples walk with them.’ (Ml. 45^c9)
- (28) *ni arindí bed n-aiPERT*
 NEG-COP_{3SG.PRES} because COP_{3SG.PST.SUBJ} ^{NAS}uttering_{NOM}
asind-robrad-som
 PV^{NAS}·say_{3SG.NEUT}·say_{AUG.3SG.PST.SUBJ}=3SG_{MASC}
 ‘It is not because it was as an uttering that he would have said it.’ (Ml. 50^b8)

It can be noted that the relative form in (27) could have been influenced by the (regular) nasalising relative *imme-tiagat* appearing earlier in the same gloss, but I am not inclined to accept that as strong evidence against *imme-tét* being relative. Example (28), however, upon closer consideration, can probably be set aside: although *arindí* was originally a prepositional phrase, it appears that it has become fully grammaticalised as a conjunction taking a nasalising relative, like the similarly formed *isindí*, *dindí*, and *lassaní*. This leaves example (26) and (27), and it seems quite likely that nasalising relatives spread to this class of fronted prepositional phrase at the same time such relatives spread to fronted adverbials, with which they are essentially synonymous:

- (29) *is ámin tra as cert in*
 COP_{3SG.PRES} thus then COP_{3SG.PRES.REL} correct_{NOM.SG.MASC/FEM} the_{NOM.SG.MASC/FEM}
testimin so
 text_{NOM} PROX
 ‘It is thus, then, that this text is correct.’ (Ml. 62^c7)
- (30) *is amne as coir a lathar*
 COP_{3SG.PRES} thus COP_{3SG.PRES.REL} fitting_{NOM.SG.NEUT} its explaining_{NOM}
 7 *estoasc a chéille*
 and expressing_{NOM} its ^{LEN}meaning_{GEN}
 ‘It is thus that explaining it and expressing its meaning are fitting.’ (Ml. 114^a9)

As noted in GOI (§§ 505–506), the spread of nasalising relatives in such adverbials is itself secondary, probably an extension of the regular nasalisation found with manner clefts (GOI § 498). While Wb. shows neither the extension of nasalising relatives to manner adverbs like *améin* / *amne* nor the extension to fronted manner prepositional phrases, the fact that Ml. has both is interesting. It is quite likely that the spread of nasalising relatives after PPs and adverbs meaning ‘thus’ (i.e. *samlaid* and *amne* / *amin*) is connected, perhaps as a result of influence from the conjunction *amal*, which takes a nasalising relative and is frequently found in the collocation *amal . . . is samlaid . . .* (12 of the 33 examples of *samlaid* in Ml. are found in this sequence).²⁴ This is naturally speculative, but it seems unlikely that there is no connection between the appearance of nasalising relatives after *samlaid* and *amne* / *amin*.

4.3.2 Leniting relatives

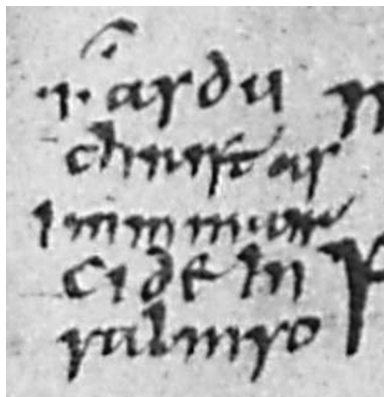
Beside the two or three examples of nasalising relatives, there are three examples of leniting relatives. The first of these is:

- (31) *as* *du Christ* *as* *immaircide*
 COP_{3SG.PRES.REL} to^{LEN} Christ_{DAT} COP_{3SG.PRES.REL} appropriate_{NOM.SG.MASC}
in *salm* *so*
 the_{NOM.SG.MASC} psalm_{NOM} PROX
 ‘that it is to Christ that this psalm is appropriate’ (Ml 16^a7)

The copula is clearly relative,²⁵ and since it does not nasalise the following *immaircide*, it must be a leniting relative. The text as given above is that of Thes. (1: 16.30). A closer look at the manuscript (see Figure 2), however, reveals that the reading is actually *immmaircide*, with three *m*’s. This sort of error is quite easy to explain as due to copying, but I would like to suggest something slightly different, namely, that the exemplar actually had *as n-immaircide* (i.e. a nasalising relative). This was either misread or miscopied as *immmaircide*, a simple error given that the sequences *in*, *ni* and *m* are frequently almost indistinguishable in Insular Minuscule. Later, *immmaircide* was corrected by the

²⁴ The examples of this collocation in Ml. are: 26^b8, 27^d22, 31^b25, 34^b6, 37^a12, 44^a19, 49^a11, 51^d28, 74^d3, 84^c9, 96^c11, and 120^d2.

²⁵ Though *is* and *as* do become interchangeable in Middle Irish, there is no evidence for such confusion as early as Ml., and I reject the possibility that *as* here could be non-relative.

Figure 2: Ml. folio 16r, gloss 16^a7.²⁷

addition of an initial *i* in the margin.²⁶ This plausible sequence of events, while not provable, is attractive in that it can explain the miscopying of *immaircide* as *immaircide*, as well as the fact that the initial *i* is not well-aligned with the margin of the gloss.

The second example of a leniting relative clause is below in (33), along with the Latin text being glossed. In order to understand the context better, the previous gloss is also given, in (32).

- (32) *fris in coais fora-robae som*
 against the_{ACC.SG.FEM} cause_{ACC} on-REL·be_{AUG.3SG.PRET} 3SG_{MASC}
 ‘to the cause that occupied him’ (Ml. 64^a12)

- (33) *ni fris ru-chét a propheta*
 NEG-COP_{3SG.PRES} against_{3SG.NEUT} AUG·^{LEN} sing_{3SG.PRET.PASS} by prophet_{ABL}
 ‘It is not with reference to it that it was sung *a propheta*.’ (Ml. 64^a13)
 gl. *usurpat hoc testimonio etiam beatus apostolus Paulus tamquam simile¹²*
non tamquam proprium¹³, quod non minus Machabeis quam apostolis
conueniret.
 ‘In this passage, as a comparison¹² [and] not as his own¹³, even the
 blessed apostle Paul uses what is not less fitting to the Machabees than to
 the apostles.’

²⁶ An alternative explanation, that *n-immaircide* was correctly copied into the Ml. manuscript, but then later misread by the corrector as *mmaircide* and “corrected”, amounts to roughly the same thing.

²⁷ Photo from Best (1936: plate 16^r) © Royal Irish Academy; reproduced by permission.

While this example may simply represent a legitimate exception to the rule that clefted PPs take a non-relative main clause verb, it is important to examine whether any alternative interpretation exists. One alternative takes the sentence as a non-cleft and *ru-chét* as a headless relative subject ‘what was sung’: ‘that which was sung by the prophet is not in reference to it (i.e. to what Paul is using it for)’. Normally, such sentences would be expected to have a substantive verb²⁸:

- (34) *Ni·fil* *dit* *daidbri-siu* *nachimm-éta-sa*
 NEG·be_{3SG.PRES} from-your poverty_{DAT}=2SG NEG(REL)-1SG-obtain_{2SG.PRES}=1SG
óm *muintir*
 from-my people_{DAT}
 ‘That you do not obtain me from my people is not because of your poverty.’
 (Meid 1974: 130–131, *Táin Bó Fraích*)

Nevertheless, the division of labour of the copula and substantive verb is not as strict as is sometimes implied. The substantive verb is sometimes used where the copula would be expected (GOI § 774; also Stifter 2006: 119). The opposite is rarer, but it does occur in a few constructions and individual examples (GOI § 816; Ahlqvist 2014: 7; see also (36) below for an example: *ní hi súidiu*). The phenomenon is not well researched, so it is unclear whether assuming the copula here in place of the substantive verb is justified or not.²⁹ As a result, it is more likely that we have here a leniting relative in a cleft sentence with fronted prepositional phrase.

The final example of a leniting relative seems secure:

- (35) *mad* *hua* [*a*]*icniud* *bes* *amlabar*
 if-COP_{3SG.PRES.SUBJ} from nature_{DAT} be_{3SG.PRES.SUBJ.REL} dumb_{NOM.SG.MASC}
 ‘[For deafness is usual to one who is dumb] if it is by nature that he is dumb.’ (ML. 59^a12)

One might argue that *bes* is the substantive verb: ‘if it is by nature that the dumb one is’. This interpretation seems forced, however. As a result, the three

²⁸ I would like to thank Elisa Roma for bringing this example to my attention, though I do not assume she agrees with my interpretation here.

²⁹ Less likely is the interpretation: ‘it is not to that which was sung *a propheta* (that the comparison is proper / that the comparison refers)’. This would assume that the antecedent of *ru-chét* is found in the conjugated preposition and that the whole gloss is the fronted material of an implied cleft sentence.

examples with leniting relative verbs can be argued to be rather one example of (originally) a nasalising relative and two examples of a leniting relative.

4.3.3 Ambiguous cases

In the following example, the form of the preverb is ambiguous in that it could be relative or could contain an infix pronoun:

- (36) *cid ho deacht maicc nó ho deacht*
 although-COP_{3SG.PRES.SUBJ} from divinity_{DAT} son_{GEN} or from divinity_{DAT}
athar· ara·foima doinacht maic a
 father_{GEN} PV_{REL}·assume_{3SG.PRES.SUBJ} humanity_{NOM} son_{GEN} the_{ACC.SG.NEUT}
n-í ar·roet ní hi suidiu
 NAS_{one} PV·assume_{AUG.3SG.PRET} NEG-COP_{3SG.PRES} in that_{DAT}
 ‘Whether it should be from the divinity of the Son or from the divinity of the Father that the humanity of the Son would assume that which He has assumed, it is not in the preceding (text).’ (ML. 17^c3)

The verb *ara·foima* may be a relative verb, but it also may contain a pleonastic infix pronoun, coreferential with the neuter object *a n-í* (see Lucht 1994: 92–94 on pleonastic infix pronouns with *a n-í*). In the latter case, this example does not belong here.

The second example in this category is also somewhat uncertain:

- (37) *acht is do sochaidi no·pridchib*
 but COP_{3SG.PRES} to multitude_{DAT} PV·preach_{1SG.FUT}
 ‘but it is to a multitude that I will preach’ (ML. 45^a8)
 gl. IN MEDIO ÆCLESIAE LAUDABO TÉ. ne putaretur singulís⁸ narraturus
 ‘I will praise you in the middle of the church, lest it be thought that I preach to individuals’

Stokes and Strachan (1901 = Thes. 1: 130, n. i), recognising that the *no* seems out of place, suggested reading *not·pridchib* ‘I will preach you’, in which the *no* is necessary to infix the second singular pronoun. The emendation, which occurs on a line break, is possible though not necessarily likely.

4.3.4 Relative endings (i.e. no nasalising / leniting distinction)

Finally, we may turn to cases of simplex verbs with relative endings. As noted above, there is no distinction of leniting or nasalising here, but the forms deserve examination nonetheless.

- (38) *corbu* *du reir* *nach* *aili*
 so.that-COP_{AUG.3SG.PRES.SUBJ} to will_{DAT} some_{GEN.SG.MASC} other_{GEN.SG.MASC}
labraimme
 speak_{1PL.PRES.REL}
 ‘that it should be at the will of some other that we speak’ (Ml. 31^{b16})

- (39) *amal is* *ho imratib* *gnaither* *cech*
 as COP_{3SG.PRES} from thoughts_{DAT.PL} do_{3SG.PRES.PASS.REL} each_{NOM.SG.MASC}
gním
 deed_{NOM}
 ‘as it is from thoughts that each deed may be done’ (Ml. 38^{a5})

Note that vowel distinctions were beginning to become confused already in Milan (Strachan 1903a: 52, 67), so (38) could contain *labraimmi* (i.e. a non-relative verb).³⁰ For *gnaither*, however (the unusual spelling of the first syllable notwithstanding), it seems that this must be accepted as a relative form.³¹

Of the 10 examples in Ml., at least 3 may not be relative after all: 17^{c3}, 31^{b16}, and 50^{b8}. Of the remaining examples, three are nasalising relatives (16^{a7}, 23^{a12}, and 45^{c9}), two are leniting (59^{a12} and 64^{a13}), one is ambiguous (45^{a8}), and one makes no distinction along those lines (38^{a5}).

5 Overview / Conclusions

The conclusions of this study are modest. From the collection of examples, it is clear that neither Wb. nor Sg. has any sure cases of relatives following fronted

³⁰ A reviewer has kindly brought to my attention two interesting examples: *in tan m-bimmi* (Ml. 24^{a18}) and *in tain diagma-ni* (Wb. 3^{a15}), which both show relative nasalisation but a non-relative ending, possibly representing a schwa. This suggests, even as early as Wb., that confusion was beginning to set in in such cases.

³¹ The Ml. scribe’s occasional tendency to write accented [er] as *-er* (cf. *a n-í as-ber titul* ‘that which the title says’ [Ml. 24^{d17}]) is probably not relevant here.

prepositional phrases. This conforms to the standard rules for the grammar of Old Irish and would seem to be an isogloss linking these two gloss collections against ML,³² where, by contrast, there are a number of clear examples of relative verbs in prepositional clefts. There are as many as ten relative examples in ML, with the certain number being maximally seven (see discussion above). Of these examples, just under half are nasalising relatives. It has been suggested that the similarity of some manner adverbials (e.g. *amne* / *amin* ‘thus’) and certain prepositional phrases (e.g. *samlaid* ‘thus’) may have led to the occasional adoption by both of nasalising relatives, perhaps on the model of *amal*, which introduces clauses of manner and regularly takes a nasalising relative. Once nasalising relatives were possible in this small set of clefted PPs, further spread in other categories and encroachment by leniting relatives may also have become possible (GOI § 506). Of the three gloss collections, ML has the strongest representation of the nasalising relative generally (McCone 1980: 15–16; Ó hUiginn 1986: 63). Given that the nasalising relative becomes redundant already by the tenth century, the increase of nasalisation in ML, followed by its rapid decrease and loss, is somewhat puzzling. Nonetheless, I would suggest that the increased number of nasalising relatives is probably connected with the spread of relatives in prepositional clefts.

5.1 *ceta* ‘first’ as evidence for the “Cowgill particle”?

It is now time to return to the case of *ceta* ‘first’. It was argued above that this preverb provides some evidence for a second-position clitic “Cowgill particle”, **eti*. Specifically, it was suggested that **kentu(y)-eti* gives *ceta*. There were, however, only two cases of non-relative *ceta*, and both happened to be in prepositional clefts. The examples are repeated here for convenience:

- (40) *airní* *doib cita-rochet*
 for-NEG-COP_{3SG.PRES} to_{3PL} PV·sing_{AUG.3SG.PRET.PASS}
 ‘For it is not to them that it was first sung.’ (ML. 86^d19a)

³² There is much more to be said here, but this is not the place. Ó Muircheartaigh (2015: 204–217) has argued for Bangor connections for both Milan and St. Gall and affinity to Armagh for Würzburg. How this might play out for specific features, however, is quite an open question.

- (41) *combad* *frisnagruade* 7
 so.that-^{NAS}COP_{3SG.PST.SUB} against-the_{ACC.PL.NEUT}=cheeks_{ACC.PL} and
frisnafortbru *cita·commaired*
 against-the_{ACC.PL.MASC/NEUT}=eyebrows_{ACC.PL} PV-meet_{3SG.PST.SUBJ}
 ‘so that it might be against the cheeks and against the eyebrows that it
 would first meet’ (Ml. 39^c15)

In the light of the examination of prepositional clefts undertaken above, these two examples from the Milan Glosses cannot be considered definitely non-relative. Since one cannot be sure of their evidentiary value, one must ask if there is any solid support left for the idea that the Cowgill particle leaves a trace behind after disyllabic preverbs ending in **u*. The preverb *ocu* < **onku-* in *ocu·ben* ‘touches’ is one such piece of support, as there is no other plausible explanation for the retention of the final syllable.

A second piece of evidence is the preverb *ceta* ‘along’ in *ceta·bí* ‘feels, perceives’. Here, the evidence is indirect. As this *ceta* has a preform **kanta*, it should have developed to relative *ceta·bí* and non-relative **cet·bí*. Since the non-relative form is actually *ceta·bí*, there must be an analogical explanation for it. It seems unlikely that the relative form of the preverb would be taken over directly. The fact that some preverbs in relative contexts had an extra syllable was well at home in Old Irish, being regular for *ar* and *imm* (relative forms *ara* and *imma*), and as Breatnach (1994b) has shown, the pattern even occurred sporadically also for other preverbs. It appears unlikely that an established **cet·bí*, relative *ceta·bí* would have been made into *ceta·bí* for both relative and non-relative without a good model.

The only possible model is *ceta* ‘first’, but interpreting the evidence is difficult. If the two examples (40) and (41) are relative, then we have no positive evidence for what the non-relative form was. There are three realistic suggestions for that form, however: it was *cet*; it was *ceta*; or there was no non-relative form because the preverb was only used in relative contexts. While García Castellero (2014: 87–89) has indeed argued that this preverb originated in relative contexts, it is unlikely that it did not spread from there at all. The textual attestation of the spread may simply be lacking. If the preverb indeed was found in non-relative contexts, it must have taken the form *cet* or *ceta*. If the non-relative form was *cet*, there would have been no model for *ceta* to be taken over in non-relative position in *ceta·bí*. On the other hand, if the non-relative form was actually *ceta* (and we happen not to have attestations of it because both (40) and (41) are actually relative forms), then this would support the argument being made here, and it would provide a model for non-relative

ceta in *ceta-bí*. Finally, it may indeed be the case that one or both of the examples (40) and (41) is non-relative. There would then be a model for non-relative *ceta* in *ceta-bí*, and *ceta* ‘first’ would provide direct positive evidence for the rule that disyllabic preverbs ending in **u* retain the final syllable before the Cowgill particle. Though the evidence is not entirely straightforward, we are left with a problem if the non-relative form of *ceta* ‘first’ was anything but *ceta*.³³

5.2 The origin of the absolute / conjunct verbal endings

We can now briefly return to the debate about the origin of absolute and conjunct verbal endings in Insular Celtic. The evidence cited here will certainly not change anyone’s mind about the validity of the particle theory as explanation for the absolute / conjunct distinction. It does, however, present evidence that disyllabic preverbs ending in **u* retained their second syllable in both relative and non-relative clauses. This does not happen with other vowels and must receive some sort of explanation, regardless of one’s views on the origins of the absolute verbal endings. The particle theory provides a relatively straightforward, though difficult to prove, framework for that explanation.

Acknowledgement: I would like to thank my Utrecht colleagues Peter Schrijver and Mícheál Ó Flaithearta, the conference participants at the “Variation and Change in the Syntax and Morphology of Medieval Celtic Languages” conference, and two anonymous reviewers for many helpful discussions and suggestions on the topics of this paper.

³³ A possible third option is that the forms in (40) and (41) were seen by speakers of Old Irish as ambiguous. If they could be seen as either relative or non-relative, they could be examples of the bridging context by which relatives in prepositional clefts became possible. While this idea has a certain appeal, it seems to be ruled out by the fact that the ambiguity of the forms exists only as written. Spoken aloud, the distinction between relative and non-relative would have been clear.

Appendix: Examples

Below are given all the examples of non-excluded prepositional clefts in Wb., Ml. and Sg., i.e. the examples that make up the data represented in Tables (5), (6), and (7).

Wb.

Non-relative: 1^d4, 9, 2^a3 is, ^b6, 15, ^c6, 13, 3^a10, ^c6, 22, 4^a4, 13, 17, 24, 27 is, ^b27, ^c23, ^d15, 33, 5^b16, 27, 36, ^c16 (bis), 6^a12, 13, 19, 30 *im-tiagam*, ^b4, 14, 8^c6, 16, 9^a1, 18, 23 (bis), ^b5, 7 *bid*, ^c9, 10, ^d27, 10^c2, 3, ^d11, 23, 27, 11^d2, 5, 6, 12^a21, 13^a3, 5, 16, 22, 32, ^b13, 18, 29, ^c11, 12, 14^c8, 24, 40, ^d26, 15^a13, ^b11, 18, 28, ^c23, ^d18, 16^d7 *ar-focarar*, 17^b20, ^c19, 18^c5, 19^a19, 20, 20^b16, ^c21, 21^b2, 7, ^c19, ^d1, 22^c10 *coiscitir*, 17, 23^a2, ^b41, ^c11 *berir*, 28, ^d21, 29, 24^a17, 29, ^d1, 21, 25^a8, ^c16, 26^b11, ^d8, 25 (bis), 27^a11 (bis), 29, ^c18, 22, 28^a19, ^b17, ^c12, 19, 29^a16, ^b12, ^d6 (bis), 23, 30^b25, 31^b11, 32^a6, ^c13, 33^d7, 34^a6.

Relative: 5^b29.

Non-leniting relative: 2^a3 *do-téit*, 3^b1, 4^a27 *for-téit*, ^b14, 5^a5, 6^a29, 30 *ad-ciam*, ^b20, 10^a29, 30, ^c1, 10, 18^d6, 19^b6, ^c6, 21^a12, 25^b28.

Non-nasalising relative: 2^c17, 3^c21, ^d21, 4^d17, 5^a1, 6^a14, ^b7, ^d5, 8^a9, ^c12, 9^b6, 7 *as-berar*, ^c14, ^d25, 10^a4, ^c11, 12, ^d16, 11^a2, 12^c29, 13^b26, 14^c33, 15^a16, 16^c4, ^d14, 17^a2, ^b29, ^c23, 18^d1, 19^b14, 20^d12, 22^a6, ^c10 *do-airbertar*, 23^b12, 17, ^d25, 26, 24^c14, 22, 25^a12, 27^b3, ^c8, ^d20, 29^a21, 31^a10, ^d2.

Ambiguous: 1^c3, 2^b24, 26, ^d25, 4^c7, 27, 32, 35, 37, 5^a4, 12, ^c17, 6^c3, ^d14, 7^a3, 14, ^d15, 8^b2, 10, ^d22, 10^a22, ^d8, 12^a29, 13^a21, ^d26, 17^d16, 18^c13, 20^d9, 10, 21^b4, ^d2, 23^a7, ^c17, ^d4, 30 *immum-ruidbed*, 25^a3, 26^c11, 27^c35, 29^a28, 30, ^d29, 31^c16, ^d6, 32^d10, 14.

Ml.

Non-relative: 3^a4, 14^d10, 15^c10, 17^b8, 20^b13 *ata* (bis), 24^d30, 26^b8, 27^c10 *teit* (bis), 28^c8, 30^d24, 31^b1, 23, 32^d6 *ata*, 10, 34^b6, ^d6 *at-taat*, 35^d26, 37^a8 *berthair*, 8 *berthir*, 10 *téit*, 10 *berthair*, 10 is, 38^a5 *gnitir*, ^c3, 42^b7 *berid*, 7 *beirthi*, 7 *ra-gab*, 43^a2, ^c13, 44^a11, 14, ^b2, 47^a17, 48^a6, 49^a11, 27, ^b7, 50^a5, 8, ^d18, 51^a14, ^b12 *eirbthi*, ^d2 *da-gneth*, 2 *da-rigni*, 10, 53^a19, ^b8, 11 *da-airilbset*, 54^a1, 56^b3, 15 *bithir*, 33, ^c11, 60^b11, 62^c2,

64^a10, 67^d8, *ata*, 8 *trachtid*, 24, 68^b2, 3, 69^b1 *molfait*, ^d3 *at·ror*, 72^d1, 12, 74^b1, ^c21, ^d13, 83^b14, 88^b15, 89^b6, 90^d11, 92^d12, 94^a13, 15, ^c3, 10 *teit*, 96^a10, 97^d17, 100^d4, 101^c4, 6–7 *saidi* (bis), 103^d26, 27 *teit*, 27 *is*, 106^c11, 108^c12 *trachtaid*, 109^a2 (*ter*), 111^b15 *dos-melmais*, 112^b20, 114^a2–3, 118^b6, 121^d8, 123^b13, 124^b3 (bis).

Relative: 16^a7, 17^c3, 23^a12, 31^b16, 38^a5 *gnaither*, 45^a8 *no·prithchib*, ^c9, 50^b8 *bed*, 59^a12, 64^a13.

Non-leniting relative: 2^b6, 25^b6, 30^a9, 32^c17, 34^d6 *no·tesad*, 39^c15, 44^a19, 50^d7 *ro·cuala*, 54^a21, ^c18 *no·teged*, 57^d13, 95^a1, 101^c6-7 *du·tiagar* (bis), 106^c3, 111^c9, 126^b2, 131^c14.

Non-nasalising relative: 2^c3, 14^a4 *ro·gabad*, 4 *robu*, 9, ^c19 *ar·osailther*, 16^a10, 17^b18, 18^a8, 19^b11, 24^c15, ^d10, 26, 29, 26^a8, 30^a3, 31^b17, 32^d6 *du·gnither*, 35^a8 *ro·gabad* (bis), 9, 10, ^b10, 16, 18, ^c21, 36^b3, ^c21, 37^a12, 14, ^b16, ^c20, 40^c20, 42^a15, 44^b1, 45^d7, 8, 46^a21, ^d3, 10, 47^c11, 48^d27, 28, 51^b12 *do·aisilbi*, ^c2, ^d8, 25, 52x0, 53^b11 *do·airilbset* [MS *do airilbset*], ^c13, 54^a22, ^d4, 55^c1, 57^d8, 64^c19, 66^d4, 69^a11, 71^b14, 74^a1, 81^c4-6, 83^d9, 84^c9, 86^d13, 89^a2, 90^b15, 91^b7, 94^c10 *do·adbat*, 96^b18, 98^c10, 100^b12, 108^b4, 109^a1 (bis), 110^d16, 111^c3, 113^c7, 115^a14, 120^d2, 121^c16, 123^c8, 10, 126^c10, 127^d2, 14, 132^a1 *ro·uctha*, 1 *as·berat*, 133^b2, 139^a6 (bis), 8, 9, 10, 11, 142^d1.

Ambiguous: 14^b12, 13, ^c19 *ro·segar*, 17^b2, ^c7 *ar·roét* (bis), 18^c10, 21^a11, 26^b10, 31^a25, ^d12, 33^d12, 37^a16, 45^a8 *as·rubart*, 9, 46^c24, 47^a8, 20, 50^d7 *ru·radus*, 51^a19, ^d28, 53^b11 *do·recachtar*, 11 *do·recatar*, ^d17, 61^d2, 66^c1, ^d15, 69^b1 *ro·fessatar*, ^d3 *ro·pridach*, 14, 72^d9, 85^d10, 86^d19^a, 88^a17, 96^c11, 102^c7, 105^a4, 108^c12 *fu·fálgi*, 113^c2, 119^d3, 125^a11, 130^b8, 145^c4.

Sg.

Non-relative: 7^b14, 9^a8, 19^b2, 26^b7, 28^a3, 32^b2, 36^b1, 38^a1, 41^b3, 42^a9, 52^b1, 54^b3, 6, 56^b8, 57^b1, 66^b9, 10, 71^b8, 76^b7, 90^b2, 95^b1, 104^b5, 113^b3, 138^a4, 139^a1, 144^b3, 152^a1, 159^a3, 168^b1, 169^a1, 173^b2, 179^a2, 181^a2, 5, 183^a2, 188^a3, ^b1, 191^a5, 196^a1, 197^a2 *ata*, 11 *ar·ícht* (*ter*), 199^b3, 200^a2, ^b7, 201^b1, 203^a7, ^b3, 8, 204^b5, 8, 205^a2, 207^a2, 208^a10, 209^b10, 29, 213^a1, 215^a8, 217^a1, 218^a8, 222^a9.

Relative: 117^a5, 148^b7

Non-leniting relative: 149^b6 (bis), 158^a3, 197^b4, 198^b3, 208^a9.

Non-nasalising: 9^b10, 10^a8, 29^b15, 30^a7, 32^a1, 35^b13, 39^a25, 45^b9, 50^a4, 59^a8, 106^b16, 140^a4, 143^a1, 157^b1, 158^a4, 161^a2, 183^b3, 187^b5, 189^b2, 192^b4, 197^a2 *as·bertar* (bis), 203^b5, 206^a3 *do·gni*, 207^b2, 208^a1, 209^a1, 210^a4, 211^a6, 213^a7, K15^a3, K66^a1.

Ambiguous: 18^a1, 6, 28^a9, 40^a17, 69^a5, 74^b8, 103^a1, 136^a1, 153^b6, 154^a1, 157^b3, 188^a14, 195^b, 202^a3, 203^a22, 206^a3 *con·osna*.

Britta Irslinger

10 The functions and semantics of Middle Welsh *X hun(an)*: A quantitative study

1 Introduction

The different types of reflexive markers are a much-discussed areal feature of the languages of Western Europe. While the markers of most European languages, like e.g. German *sich*, French *se* or Italian *si* are based on the PIE reflexive pronoun **s(ǫ)e-*,¹ English and the neighbouring Insular Celtic languages Welsh and Irish² employ different markers originating from intensifiers. As a result, reflexive markers and intensifiers are different in the first group of languages, but not in the second. German expresses reflexivity with the pronouns *mich*, *dich*, *sich*, etc., as in (1a) and (1b), while uninflected *selbst* ‘self’ is used as an intensifier, adnominally in (1c) and adverbally in (1d), English uses *my-*, *your-*, *himself*, etc. in both cases.

- (1) a. German: *Ich sehe mich im Spiegel.*
I see_{1SG.PRES} me_{ACC} in.the_{DAT.SG.MASC} mirror
English: *I see myself in the mirror.*
- b. German: *Er spricht ständig mit sich.*
he speaks_{3SG.PRES} continuously with REFL
English: *‘He keeps talking to himself.’*
- c. German: *Der Präsident selbst wird der Feier beiwohnen.*
the_{NOM.SG.MASC} president self will_{3SG.PRES} the_{DAT.SG.FEM}
ceremony attend_{INF}
English: *The President himself will attend the ceremony.*
- d. German: *Der Präsident schrieb seine Rede selbst.*
the_{NOM.SG.MASC} president write_{3SG.PRET} his_{ACC.FEM.SG.} speech
self
English: *The President wrote his speech himself.*

¹ Haspelmath (2001: 1501), König and Siemund (2000: 44–51).

² See Irslinger (2014b: 161–164) on Modern Irish and (2014b: 179–182) on Old Irish.

A formal similarity of intensifiers / reflexives is to be observed especially between English and Welsh. Both languages have complex markers consisting of a pronoun inflected according to person, number and gender coupled with a second element, which is *self* in English, *hun* in North Welsh and *hunan* in South Welsh (Table 1). In addition, both markers originate from intensifiers and are in use with such function.

Table 1: Paradigms of intensifiers/reflexives in Modern Welsh and English.

		Modern Welsh		Modern English
		North	South	
Sg.	1	<i>fy hun</i>	<i>fy hunan</i>	<i>myself</i>
	2	<i>dy hun</i>	<i>dy hunan</i>	<i>yourself</i>
	3 _{MASC.}	<i>ei hun</i>	<i>ei hunan</i>	<i>himself</i>
	3 _{FEM.}	<i>ei hun</i>	<i>ei hunan</i>	<i>herself</i>
Pl.	1	<i>ein hun</i>	<i>ein hunain</i>	<i>ourselves</i>
	2	<i>eich hun</i>	<i>eich hunain</i>	<i>yourselves</i>
	3	<i>eu hun</i>	<i>eu hunain</i>	<i>themselves</i>

Because English differs from the other Germanic languages, which have reflexives based on PIE **se-*,³ and displays a marker structurally similar to the Welsh one, the hypothesis of a celticism in English or at least of convergent developments has widely been discussed as a possible explanation.⁴

For both languages the double function of intensifier / reflexive is not yet to be found in the earliest documents, i.e. in Old English or Old Welsh.⁵ Both

³ Cf. Gothic dat. *sis*, acc. *sik*, Old Norse *ser*, *sik*, Old Saxon *sik*, Old High German *sih* < Proto-Germanic **siz*, **sike* ‘himself, herself’ (Kroonen 2013: 437).

⁴ See the different treatments e.g. in Preusler (1938: 187), Tristram (1999: 24), Vezzosi (2005: 228–240), Filppula, Klemola, and Paulasto (2008: 95–97), Miller (2012: 37) and Vennemann (2013: 122). According to Poppe (2009: 253–258) the hypothesis remains unproven, albeit attractive. Lange (2007: 186) is sceptical and suggests conducting further research first. Contrary to this, van Gelderen (2019: 225) rejects any influence from Irish or Welsh on the Old English Lindisfarne Glosses.

⁵ Old Welsh is fragmentarily attested in onomastics, glosses and a few short texts, some of which are difficult to understand. This material contains two clear examples of intensifying *X hun(an)*. In addition, there is one reflexive construction, containing a verb with the prefix *im-*

English *X-self* and Welsh *X hun(an)* were first employed as intensifiers and only much later as reflexive anaphors. This means that whatever conclusion can be reached with regard to contact influence in the case of the intensifiers cannot be relevant for the reflexives, as separate processes brought about their emergence in the two languages.

In Old English, co-reference was expressed by the ordinary personal pronouns, which were ambiguous in the third persons, cf. (2) and (3). Disambiguation could be obtained by adding an intensifier like in (4) (König and Siemund 2000: 44–46).

- (2) *hine he bewerað mid wæpnum*
 he_{ACC} he_{NOM} defend_{3SG.PRES} with weapons_{DAT.PL}
 ‘He defended himself with weapons.’ (Zupitza 1966 [Ælfric, *Grammar* 96. 11–12]; late 10th–early 11th century;)
- (3) *ða behydde Adam hine & his wif eac swa dyde*
 and hide_{3SG.PRET} Adam_{NOM} he_{ACC} and his wife_{NOM} also so do_{3SG.PRET}
 ‘and Adam hid himself and his wife did the same’ (Crawford 1922 [Ælfric, *Genesis* 3.8]; late 10th–early 11th century)
- (4) *Hannibal . . . hine selfne mid atre acwealde.*
 Hannibal_{NOM} hide_{3SG.PRET} self_{ACC.SG.MASC} with poison_{DAT} kill_{3SG.PRET}
 ‘Hannibal killed himself with poison.’ (Sweet 1883; [Orosius IV.11]; late 9th century)

For English the expansion of the functional scope of the intensifier and its use as a reflexive marker can be dated precisely. While the earliest examples can be found around 1150, the replacement of the simple pronoun strategy by *X-self* was complete as late as the seventeenth century. Nevertheless, *X-self* was used as a reflexive in the majority of cases already by the end of the fifteenth century. Examples (5) and (6) illustrate the old and new strategies respectively in different editions of the Bible (Peitsara 1997: 288; König and Siemund 2000: 49; Keenan 2002: 333–350; Lange 2007: 173–177).

(Old Welsh for Middle Welsh *ym-*) and possibly an infix pronoun. Although the analysis of the latter is controversial (see the discussion in Irslinger 2014b: 183, 191–193), its analysis as a plain pronoun expressing co-reference is probable in view of Middle Welsh (see below, section 4.2). Overall, there is not enough evidence to draw any firm conclusions regarding the expression of reflexivity in Old Welsh.

- (5) *Adam and his wijf **hidden hem** fro the face of the Lord God*
 ‘Adam and his wife hid themselves from the face of the Lord God.’
 (Peitsara 1997: 321 [Wycliffe, *The Old Testament, Genesis 3.8, 1380*])
- (6) *And Adam **hyd hymselfe** and his wyfe also from the face of the LORDe God.*
 (Peitsara 1997: 322 [Wycliffe, *The Old Testament, Genesis 3.8, 1380*])

Unfortunately, such detailed information is not available for Middle Welsh *X hun(an)*, making it challenging to compare the development of the two languages.

The present chapter makes a first attempt to carry out such a comparison with the help of a quantitative study. The paper is organised as follows: section 2 summarises the relevant typological and diachronic research on intensifiers developing into reflexive markers. Section 3 examines the number of occurrences of *X hun(an)* in the *Rhyddiaith Gymraeg* corpus, their function and their distribution according to text types. Section 4 will then analyse the function of *X hun(an)* as a part of constructions coding reflexive events as well as the semantics and syntax of the verbs with which it occurs considering also the material contained in the *Rhyddiaith y 13eg Ganrif* corpus. Finally, the instances of *X hun(an)* as a reflexive marker will be discussed in detail with regard to date, distribution and possible triggers of the change.

2 Typological and diachronic aspects

An intensifier can be adjoined to each constituent of a clause, referring to the entity expressed by this very constituent. Examples (7a) and (7b) from König (2001: 748) illustrate this use for two different constituents. In (7a) the intensifier is adjoined to the NP coding the agent and refers to it (adnominal use). In (7b), the intensifier is adjoined to the VP (adverbial use). Because a verbal action presupposes the presence of an agent, the intensifier refers not only to the action itself, but also, and even predominantly, to the agent who performs it intentionally. Gast and Siemund (2006: 366) thus propose the term “actor-oriented” instead of “adverbial” for the type in (7b), which will be adopted here.⁶

⁶ Gast and Siemund (2006: 366, 371) describe the function of an ‘actor-oriented’ intensifier as ‘role disambiguation’. The intensifier blocks middle readings of polyfunctional verbal or pronominal middle markers, stating who is the intentional agent.

- (7) a. **The President himself** will attend the ceremony. (adnominal use)
 b. The President **wrote** his speech **himself**. (adverbial/actor-oriented use)
 (König 2001: 748; Gast and Siemund 2006: 349)

The basic function of intensifiers is to evoke alternatives to the referent of their focus. In doing so, they structure the set of referents belonging to a certain situation into a centre expressed by the intensified constituent and a periphery (König 2001: 749).

Intensifiers thus express co-reference with their head like reflexives, but their function is pragmatic instead of syntactic. In combination with reflexives they have a disambiguating function, i.e. adnominal intensifiers are used for referential disambiguation and actor-oriented intensifiers are used for role disambiguation (Gast and Siemund 2006: 363, 370).

Intensifiers are thus often adjoined to “full reflexives”, i.e. transitive events in which the agent performs an action on him- or herself (8b), which he or she normally performs on a patient (8a). The self-direction of the action is unexpected and thus semantically marked, especially in the case of negative actions. The optional intensifier in (8c) from German is actor-oriented and emphasises that the actor intentionally performed this act. English does not allow an equivalent differentiation, because the reflexive and the intensifier are identical and the sequence **himself himself* is ungrammatical (Kemmer 1993: 52; König 2001: 758; Gast and Siemund 2006: 366).

- (8) a. English: *He kills his neighbour.*
 b. English: *He kills himself.*
 c. German: *Er tötet sich selbst.*
 he kill_{3SG.PRES} REFL self
 ‘He kills himself.’

Because of this functional and semantic overlap, intensifiers have the potential to develop into reflexives, and, undergoing grammaticalisation, intensifiers share a first functional expansion as markers of “full reflexives”.

2.1 Grammaticalisation pathway of intensifiers/reflexives

Figure 1 illustrates the grammaticalisation path for Proto-Indo-European (PIE) **s(y)e-*, which probably was originally an intensifier. In the Romance, Germanic, Baltic and Slavic languages, it is the root of reflexive pronouns and of reflexive verbal endings, which prototypically express co-reference of the agent and the

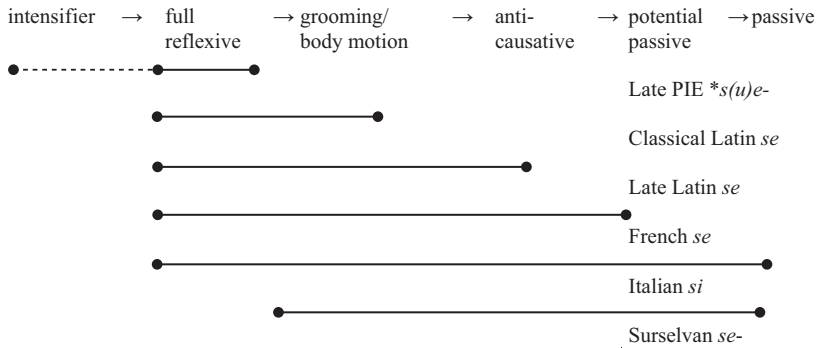


Figure 1: The functional development of of PIE **s(u)e* from PIE to Romance.

patient of a transitive verb. From there, their scope spread to further domains of detransitive voice, such as middle voice, reciprocal, anticausative, impersonal or passive. Although the individual languages have reached different stages of grammaticalisation, the development of their functional extensions follow the same unidirectional grammaticalisation path (Haspelmath 2003: 235 with Figure 8.18).⁷

On the other hand, English *X-self* and Welsh *X hun(an)* cover mainly the first stages of the grammaticalization path. Both markers originate from intensifiers and are still used for this function (Figure 2).

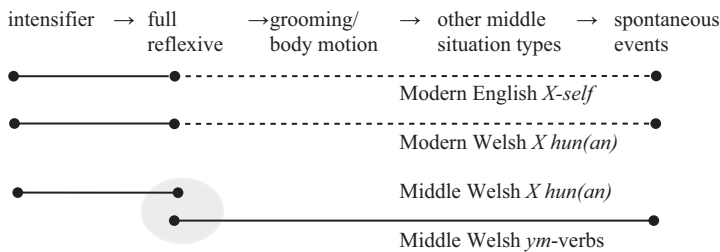


Figure 2: The functional scope of Engl. *X-self* and Welsh *X hun(an)*.

The Modern English and Modern Welsh markers are used as full reflexives. In addition, they can also be found with verbs belonging to various middle situation types, but on the whole marking is much rarer than with corresponding verbs in

⁷ See Irslinger (2014b: 166–168) with an overview of recent studies on PIE **s(u)e* and its developments in different languages. See also Harbert (2007: 327) on Germanic, Stéfanini (1962: 114) on Romance.

languages like German or French. In Figure 2, this expansion is therefore represented by broken lines. The fact that the same expression is employed for intensifiers and reflexive anaphors prevents their use as markers of derived intransitivity to a certain extent (König and Siemund 2000: 65), e.g. in verb pairs belonging to the inchoative / causative alternation.⁸ In German or French, an unmarked transitive can be clearly distinguished from its intransitive counterpart, which is marked by *sich* or *se* respectively, in (9a) and (9b). On the other hand, English and Welsh possess a considerable number of “labile” verbs that can be constructed transitively or intransitively, where no additional marker is needed in the second case, as in (9c) and (9d) (Pope 2009: 262–264).

9. a. German: *Sie öffnet die Tür.*
 she open_{3SG.PRES} the_{ACC.SG.FEM} door
 ‘She opens the door.’
 versus *Die Tür öffnet sich.*
 the_{NOM.SG.FEM} door opens_{3SG.PRES} REFL
 ‘The door opens.’
- b. French: *Elle ouvre la porte.*
 she opens_{3SG.PRES} the_{SG.FEM} door
 ‘She opens the door.’
 versus *La porte s’ouvre.*
 the_{SG.FEM} door REFL opens_{3SG.PRES}
 ‘The door opens.’
- c. English: *She opens the door.*
 versus *The door opens.*
- d. Welsh: *Mae hi ‘n agor y drws.*
 be_{3SG.PRES} she PRED open_{VN} DEF door
 ‘She opens the door.’
 versus *Mae ‘r drws yn agor.*
 be_{3SG.PRES} DEF door PRED open_{VN}
 ‘The door opens.’

However, this parallelism between English and Welsh can be found only in the modern period. In Middle Welsh, the verbal prefix *ym-* is productively employed to transform transitive verbs into intransitive ones, expressing a broad range of middle functions. The marker, which originates from the Proto-Celtic

⁸ Nevertheless, such verbs are not absent from English and the number of lexicalised reflexive verbs, motion middles and anticausatives has been increasing since the Middle English period, cf. Siemund (2010, 2014).

preposition **ambi*- ‘about, at all sides’ is predominantly used as a marker of reciprocity. However, it occurs also with verbs denoting the other middle situation types established by Kemmer (1993: 16–20), such as body care, body motion, change of body position, benefactive middle, cognition middle and spontaneous events (anticausatives) (Irslinger 2017c: 116–123). Occasionally, *ym*-verbs can also act as full reflexives. Given that a full reflexive function is the first step an intensifier goes through when expanding its scope by grammaticalization, an analysis of such a function is crucial to understanding when and how the intensifier *X hun(an)* developed into a reflexive marker.

2.2 Middle Welsh *X hun(an)* in previous research

Evans (1964: 89, §98) introduces *X hun(an)* as the Middle Welsh “reflexive pronoun” in the standard handbook *A Grammar of Middle Welsh*, adducing a great number of examples that illustrate its use. However, he does not make a distinction between the functions of reflexive marker and intensifier, and most of the examples actually contain intensifying *X hun*, like in (10) through (13). When translating Middle Welsh into Modern English, Evans renders Middle Welsh *X hun(an)* in most cases as English *X-self*. The difference becomes apparent only in languages in which reflexives and intensifiers are not identical, like e.g. German.⁹ For the sake of clarity, German translations have been added to Evans’ English ones.

- (10) *e ’r amherauder e hun*
 to DEF emperor 3SG_{MASC.INTS}
 ‘to the emperor himself / zum Kaiser selbst’ (Jones 1939: 336.33 [*Gwyrtheu Mair*])
- (11) *neu ’r diffetheist du hun*
 PTC PERF destroy_{2SG.PRET} 2SG_{INTS}
 ‘thou thyself hast destroyed / du selbst hast zerstört’ (Williams 1951: 20.29 [*Pedeir Keinc y Mabinogi*])
- (12) *yr a gewssynt e hun*
 DEF PTC get_{3PL.PRET} 3PL_{INTS}
 ‘what they themselves had got / was sie selbst bekommen hatten’ (Williams 1951: 46.27 [*Pedeir Keinc y Mabinogi*])

⁹ König (2001: 751–752), Haspelmath (2001: 1501).

- (13) *dy anwybot dy hun*
 2SG_{POSS} ignorance 2SG_{INTS}
 ‘thy own ignorance / dein eigenes Unwissen’ (Williams 1951: 2.12–13
 [*Pedeir Keinc y Mabinogi*])

Only in three examples – (14), (15) and perhaps in (57) below – does *X hun(an)* function as a reflexive marker. In German, this is rendered as the reflexive pronoun *dich*¹⁰ followed by the intensifier *selbst*.

- (14) *na chapla dy hun*
 NEG reprove_{2SG.IMPV} 2SG_{REFL}
 do not reprove thyself / tadle dich nicht selbst’ (Lewis 1925: 23.28 [*Cynghorau Catwn*])

- (15) *ony ledy dy hun*
 unless kill_{2SG.PRES} 2SG_{REFL}
 ‘unless thou dost kill thyself / außer wenn du dich (selbst) tötest’ (Jones 1941: 24.25 [*Cynghorau Catwn*])

This construction co-occurs with the one in (16), in which the verb *hoffi* ‘to praise’ is turned into the prefixed *ym·hoffi* ‘praise oneself’. Here *X hun* functions as an intensifier.

- (16) *nac ym·hoffa vyth dy hun*
 NEG PV·praise_{2SG.IMPV} ever 2SG_{INTS}
 ‘do not ever praise thyself / lobe dich niemals selbst’ (Lewis 1925: 29.37
 [*Cynghorau Catwn*])

Parina (2007) criticises this analysis, arguing that the instances of *X hun* contained in the Middle Welsh text *Pedeir Keinc y Mabinogi* (PKM, ca. 26,000 words) are to be considered as intensifiers corresponding to the different types established by typological research (by König 2001 and others). Parina also maintains that this text contains no instances of full reflexives coded with *X hun(an)*.

The examination of the *ym*-verbs contained in PKM by Irslinger (2017c) yielded no examples of full reflexives like in (16) either. Instead, all *ym*-verbs in said text were found to belong to the group of middle situation types. This

¹⁰ The special reflexive pronoun *sich* only figures in the third person. In all other persons, the respective personal pronoun in dative or accusative case is used, cf. Irslinger (2014b: 171–172).

clearly shows that the corpus of PKM is too small to contain all possible expressions of reflexivity in Middle Welsh.

3 The corpus-based study

3.1 The corpora: *Rhyddiaith y 13eg Ganrif* and *Rhyddiaith Gymraeg 1300–1425*

The following study is based on the corpora *Rhyddiaith y 13eg Ganrif: Fersiwn 2.0* (Isaac et al. 2013) and *Rhyddiaith Gymraeg* (Luft, Thomas and Smith 2013),¹¹ covering together the whole period of Middle Welsh.

Rhyddiaith y 13eg Ganrif contains nearly half a million words from 27 texts preserved in 17 manuscripts. The main textual genres represented are history and law, which together, in approximately equal parts, make up about 90 percent of the corpus.¹² The remaining 10 percent includes mostly short or fragmentary texts belonging to the Mabinogion, natural history, religion, romance, and wisdom literature.

Rhyddiaith Gymraeg 1300–1425 contains some 2.8 million words from over 100 texts belonging to different genres and preserved in 54 manuscripts. The corpus contains texts belonging to all medieval genres, namely genealogy, geography, grammar, history, law, Mabinogion, medicine, natural history, religion, romance and wisdom literature.¹³

11 Texts, titles and manuscripts pages/folios are cited according to *Rhyddiaith y 13eg Ganrif* and *Rhyddiaith Gymraeg / Welsh Prose 1300–1425* unless stated otherwise. Translations are my own, unless another author is indicated.

12 The history section consists of three versions of *Brut y Brenhinoedd* from NLW MS. Peniarth 44, Llanstephan 1 and the Dingestow Court manuscript (NLW MS. 5266). Although these texts are independent translations from Latin, they are nevertheless very similar. Occasional passages with identical wordings are due to coincidence (Sims-Williams 2016: 55). The law texts from British Library Cotton Caligula A.iii, NLW MS. Peniarth 29, NLW MS. Peniarth 30, British Library Cotton Titus D.ii and British Library Additional 14931 all belong to the Iorwerth redaction. Due to the special character of this textual genre, they contain many passages with identical readings, which are also preserved in the later versions of *Rhyddiaith Gymraeg 1300–1425*.

13 Lange (2007: 81–82) states that in Old English the occurrence of intensifiers is genre-sensitive. They are found more frequently in texts closer to oral registers and directly addressing the reader, while they are rarer in scientific and formal registers. This seems to be the case in Middle Welsh as well, but it is not possible to test this hypothesis at the time being. *Rhyddiaith Gymraeg* gives separate word counts only for the manuscripts, which mostly contain texts belonging to different genres, but not for the texts themselves.

3.2 Quantitative and functional analysis of *X hun(an)*

The analysis in the present section is based on *Rhyddiaith Gymraeg 1300–1425* alone, which is more representative and balanced because of its size and textual variation. The results will therefore describe the language of the second half of the Middle Welsh period. Nevertheless, they are also valid for the earlier period covered by *Rhyddiaith y 13eg Ganrif*, as all functional types are also found there in roughly similar proportions. In addition, the *Bruts* and especially the law texts contain passages, which have identical or very similar counterparts in the later versions. *Rhyddiaith y 13eg Ganrif* will be considered in detail in section 4.

With the help of the wordlist, the *Rhyddiaith Gymraeg 1300–1425* has been searched for all occurrences of *hun*, *hunan* and *hunein* in different spellings, including misspellings. The following homographs were then excluded: *hun*, *hvn* ‘sleep’, *hun*, *hvn* ‘one, only’ (including *hun* used attributively and *yr hun* introducing a relative clause), *hvn* as an unusual spelling of the demonstratives *hwn*, *hynn* and a handful of unclear instances. The corpus yielded 4,091 instances of *X hun*¹⁴ (Table 2).

Table 2: Instances of *X hun* in *Rhyddiaith Gymraeg 1300–1425* sorted by genre.

	Instances	
	All	Different
Genealogy	–	–
Geography	21	18
Grammar	24	18
History	1548	546
Law	677	316
Mabinogion	228	128

¹⁴ In the rest of the chapter, *X hun* will be used in place of all graphic and grammatical variants, i.e. *X hun(an)*, *X hvn(an)* and plural *X hunein*, *X hvnein*, *X huneyn*, *X hvneyn*.

Table 2 (continued)

	Instances	
	All	Different
Medicine	34	26
Natural History	4	4
Religion	792	328
Romance	631	408
Wisdom	132	116
Total	4091	1908

In medieval corpora, popular texts are typically preserved in multiple copies that are more or less identical. The *Brut y Brenhinoedd* or the *Ystoria Carolo Magno: Chronicl Turpin* for example occur in fifteen manuscripts. While similar, albeit different, wordings show the range of possible expressions for a certain concept, identical passages are duplicates that would distort the results of a quantitative analysis. These duplicates have thus been eliminated from the corpus, reducing the data by more than half.

Table 3 lists all constructions with *X hun* in the corpus. Duplicates were identified according to the following criteria: If the same construction is involved or if the same constituent is intensified, passages are considered as duplicates even if they were lexically different. In (17) and (18), *e hun* follows a personal pronoun as an adnominal intensifier. Although different pronouns are involved, i.e. *ynteu* and *efo*, the passage was counted only once.

In the cases where the constructions were different or *X hun* occurred with another constituent, the passages were considered as different, even when the rest was identical. In contrast to (17) and (18), *e hun* following the verb *aeth* is an actor-oriented intensifier in (19). This passage was thus counted as a separate instance.

- (17) *ac ynteu e hun a aeth y gastell dimlyot*
 and 3_{SG.MASC} 3_{SG.MASC.INTS} PTC went_{3_{SG.PRET}} to castle Dimlyot
 ‘and he himself went to Dimlyot Castle’ (*Brut y Brenhinoedd*; Oxford Jesus College Manuscript 111, page 38^r (149): 29)

- (18) *ac efo e hun a aeth y gastel dimloec*
 and 3SG_{MASC} 3SG_{MASC.INTS} PTC went_{3SG.PRET} to castle Dimloec
 ‘and he himself went to Dimloec Castle’ (*Brut y Brenhinoedd*; NLW MS. 3035 (Mostyn 116), page 61^v: 13)
- (19) *ac ynteu a aeth e hun yg castell dimlot*
 and 3SG_{MASC} PTC went_{3SG.PRET} 3SG_{MASC.INTS} to castle Dimlot
 ‘and he went to Dimlot Castle himself’ (*Brut y Brenhinoedd*; NLW MS. Peniarth 46, page 254: 16)

There are no functional differences between the variants *X hun* and *X human* or between the use of singular and plural forms *e hun* and *e(u) hunein*. Accordingly, the passages containing these variants were considered as identical.

3.3 Constructions with *X hun*

The 1,908 instances of *X hun* contained in *Rhyddiaith Gymraeg 1300–1425* were analysed according to their function. The results are listed in Table 3. According to this analysis, *X hun* is employed mainly (i.e. in at least in 97.33% of the cases) as an intensifier in the different constructions illustrated in 3.2.1.

In 51 instances, *X hun* following a verb or verbal noun occurs in situations that comply with the definition of “full reflexives”. However, it would be mistaken to assume that *X hun* has the function of a reflexive marker in all these cases. Rather, these instances show a number of different constructions, which will be examined in detail in section 4 to determine the function of *X hun*.

3.3.1 Adnominal intensifiers

Adnominal intensifiers follow a NP or proper name (20), a NP preceded by a possessive adjective (21), a pronoun (22)¹⁵ or a prepositional pronoun (23) respectively. Both the simple NP and the construction POSS+NP may or may not be preceded by a preposition, e.g. *y henw e hvn* ‘his own name’, *o ’y henw e hvn*

¹⁵ The stressed possessive pronoun *eidaw* is used both predicatively and substantivised, cf. Evans (1964: 54–55) and *eGPC* (s.v. *eiddo*) for the respective constructions. The 17 instances of *e hun* following substantivised *eidaw* have been counted as POSS+NP in Table 3.

Table 3: Quantitative functional analysis of *X hun* in *Rhyddiaith Gymraeg 1300–1425*.

	n.c. ¹⁶	Intensifier adjoined to:						“full reflexive event”
		NP	POSS+NP	pron.	prepos. pron.	head	verb	
Genealogy	0							
Geography	18	1	4	2	4	4	2	1
Grammar	18	6			9	2	1	
History	546	1	75	213	70	88	11	79
Law	316		20	121	52	39	23	61
Mabinogion	128		18	48	12	18	8	24
Medical	26		3	7	1	12		3
Natural Hist.	4		1	1	1	1		
Religious	328		66	98	26	63	10	49
Romance	408	1	86	113	41	64	15	74
Wisdom	116		6	42	6	24	4	23
Total	1908	2	282	647	211	322	77	316
Percentage ¹⁷	100.00		14.78	33.91	11.06	16.87	4.04	16.56

‘from his own name’. In most genres, the POSS+NP type is significantly more frequent than all other types.

The examples are given within their contexts to illustrate the function of the intensifiers, i.e. structuring the respective situations according to the roles of the participants involved, which may be either central or peripheral.

(20) [*Ac yny diwet hwn pymp kenedyl yssyd yny chyuanhedu nyd amgen. normanyeit. bryttanneit. saesson. fichtieit. ac ysgottieit.*]

ac o hynny oll nyd oed gynt yn y
 and of PROX all NEG be_{3SG.IMPF} before in 3PL_{POSS}

¹⁶ n.c. = not classified: due to corruption of the manuscript, it was impossible to determine the context and thus the function of *X hun*.

¹⁷ Values were rounded off to the second decimal place.

medu o 'r mor pwy gilyd namyn bryttannieit
 possess_{VN} from DEF sea to another except Britons
eu hun.

3PL_{INTS}

[‘And today, there are five nations who inhabit it, namely the Normans, the Britons, the Saxons, the Picts, and the Scots] and of all these, in the past no one possessed it from sea to sea, but the Britons themselves.’
 (*Brut y Brenhinoedd*; BL Cotton Cleopatra MS B V part i, page 2^v: 7–9)

- (21) *ac y dodes ynteu ar y ran kymre*
 and PTCL put_{3SG.PRET} 3SG on 3SG_{MASC.POSS} part Cambria
o 'y henw e hvn.
 from 3SG_{MASC.POSS} name 3SG_{MASC.INTS}
 ‘and he called his part Cambria from his own name.’ (Parry 1937: 24 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS B V part i, page 11^v: 3])

- (22) [*Sef y rodes y auarwy y nei. llundein ac yarllaeth keint. Ac a rodes y the-
 neuan y nei y llall yarllaeth kernyw.*]
Ac ynteu e hun yn vrenhin ar gwbyl.
 and 3SG_{MASC} 3SG_{MASC.INTS} in king on whole
 [‘To Avarwy his nephew he gave London and the Earldom of Kent, and to Tenevan, his other nephew, he gave the Earldom of Cornwall,] and he himself was king over the whole.’ (Parry 1937: 70 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS B V part i, page 34^v: 23])

- (23) [*a choffau na wnathoed y vrawd ydau ef dim o'r cam.*]
namyn ef a wnathoed cam y 'v vrawd
 but 3SG_{MASC} PTCL do_{3SG.PLPF} wrong to 3SG_{MASC.POSS} brother
ac idaw e hvn.
 and to_{3SG.MASC} 3SG_{MASC.INTS}
 [‘and to remember that his brother had done him no wrong,] but that he had done wrong to his brother and to himself’ (Parry 1937: 50 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS B V part i, page 24^v: 15])

The characteristic morphology of Middle Welsh prepositional objects in (23) contrasts clearly with Old English, where intensifiers after prepositional objects are also frequent (van Gelderen 2000: 47). On the one hand, most Middle Welsh prepositions possess personalised forms originating mostly from their fusion with following personal pronouns, the so-called “prepositional pronouns” or

“inflected prepositions”; *idaw* ‘to him’, for example, is the third singular masculine of *y* ‘to’. On the other hand, the preposition and the following pronoun are separate units in Old English (example 24).

- (24) *heht hie bringan to him selfum*
 order_{3SG.PRET} her_{ACC} bring_{INF} to him_{DAT} self_{DAT}
 ‘ordered (them) to bring her to himself.’ (van Gelderen 2000: 47 [*Genesis* 2629])

3.3.2 Intensifiers as heads

Like English *X-self*, *X hun* can occur alone, without a preceding noun or pronoun, thus claiming the function of a pronoun for itself.¹⁸ This use is frequently found in comparisons after *no(c)* ‘than’, and after *kanys* ‘since’, *namyn* ‘but’ and *onyt* ‘except’, but also without any preceding word, as in (25). In all cases, the use of a personal pronoun or a personal pronoun + intensifier would be possible as well.

Almost without exception, *X hun* as a head codes the subject, but in a few cases it is found after uninflected prepositions.

- (25) [*Arglwydi heb ef pei barnewch wi oll ellwg hengyst.*]
Mu hunan a 'e lladwn ef.
 1SG_{INTS} PTCL 3SG_{MASC.INFX} kill_{1SG.IMPF.SUBJ} 3SG_{MASC}
 [‘Lords, said he, if you all would judge to release Hengist,] (I) myself would kill him.’ (lit.: ‘It’s myself, who would kill him.’) (*Brut y Brenhinoedd*; Cardiff MS. 1.362 [Hafod 1], page 63^v: 19)

3.3.3 Actor-oriented (adverbial) intensifiers

Actor-oriented intensifiers are further subdivided into two types: exclusive and inclusive. In the exclusive type (26a), the meaning of the intensifier corresponds roughly to *personally*, in the inclusive type (26b) the intensifier could be replaced by *too*, *also* (König 2001: 748).

¹⁸ Evans (1964: 89) calls these intensifiers “heads”, while Parina (2007: 394) labels the function “diskursiv [discursive]”.

- (26) a. *I have swept this court myself. Nobody helped me.*
 b. *I have myself swept this court. I know how difficult that is.*

Welsh uses identical markers for both types (Parina 2007: 393). Example (27) illustrates the actor-oriented exclusive use, which is the predominant one in the corpus.¹⁹

- (27) *ac y cladpwyf ef yn y gaer a*
 and PTCL bury_{PRET.IMPS} 3SG_{MASC} in DEF city PTCL
adeillassei e hunan yn anrydedus.
 build_{3SG.PLPF} 3SG_{MASC.INTS} PTCL honourable
 ‘and he was buried honourably in the city which he had built himself.’
 (*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 11^r: 21)

3.3.4 Additional resumptive pronouns

An additional pronoun can stand between the intensified constituent and the intensifier, cf. (28) with a prepositional pronoun and (29) with a possessive construction. This pronoun refers to the intensified constituent.

However, since the presence or absence of additional pronouns can be observed frequently in otherwise identical versions, the pragmatic effect does not seem to be very significant. The author of the Cotton Cleopatra version of the *Brut* has a strong preference for them. Overall, resumptive pronouns are relatively rare. They occur most frequently after finite verb forms, i.e. in 8.54% of all verbs followed by intensifiers.²⁰

- (28) [*A gorchymyn a oruc aganipus yr freinc ar eu heneit ac ev hanreith eu bod kyn vfydet y lyr ac yw verch.*]
ac y bythynt idaw ef e hvn.
 and PTCL be_{3PL.HAB.} to_{3SG.MASC} 3SG_{MASC} 3SG_{MASC.INTS}
 [‘And Aganippus bade the French, on their lives and their possessions, to be as obedient to Lear and to his daughter] as they would be to himself’

¹⁹ Out of 316 instances of actor-oriented intensifiers, only about 20 appear to be inclusive. In several cases, however, it was not immediately evident which use was intended. A more detailed examination would be necessary.

²⁰ NP + pronoun: 2 instances; possessive NP + pronoun: 29; pronoun + pronoun: 1; prepositional pronoun + pronoun: 20; verb + pronoun: 27. In Table 3, these numbers are contained in the counts of the respective groups.

(Parry 1937: 39–40 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 19^v: 1])

(29) [*ac yna y dodes corineus ar y ran ef . . .*]

o' y henw ef e hun kerniw.

from 3SG_{MASC.POSS} name 3SG_{MASC} 3SG_{MASC.INTS} Cornwall

[‘And then Corineus named his part . . .] Cornwall after his own name’ (Parry 1937: 23 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 11^r: 9])

4 *Ym*-verbs and simple verbs + *X hun*

4.1 Semantics

In Middle Welsh, both *ym*-verbs and simple verbs occur with *X hun*, coding “full reflexives” and similar situation types. The following section will thus give a semantic analysis of the two groups.

In this section, the nineteen examples found in *Rhyddiaith y 13eg Ganrif* will be considered as well. In the following tables, the numbers in the first column refer to these nineteen examples. Those in the second to the attestations in *Rhyddiaith Gymraeg 1300–1425*.

Table 4: Semantics and frequency of *ym*-verbs followed by *X hun*.

	Rhyddiaith y 13eg Ganrif	
<i>am</i> - ‘about’	1	<i>ymdeith</i> ‘to walk about’
	1	? <i>ymogel</i> ‘to take care’ (< **‘to watch about?’)
reciprocal	1	<i>ymguro</i> ‘to beat (one another or oneself)’, here: reciprocal
body care	1	<i>ymwisc</i> ‘to dress’
	1	<i>amgreffinnaw</i> ‘to scratch oneself’
body movement	1	<i>ymdroi</i> ‘to turn (oneself)’
	1	<i>ymdyrchafel</i> ‘to raise (oneself)’
spontaneous event	1	<i>ymdangos</i> ‘to appear, to show oneself’
	2	<i>ymagor</i> ‘to open’ (of doors)
benefactive (prototypical)	1	<i>ymwledu</i> ‘to feast’ (or reciprocal?)
	1	<i>ymborth</i> ‘to feed, sustain (oneself)’

Table 4 (continued)

Rhyddiaith y 13eg Ganrif		
benefactive (marked) / positive self-directed actions / reflexive	1	1 <i>emdiffryd</i> 'to defend oneself'
		2 <i>ymgyuoethogi</i> 'to enrich oneself'
		1 <i>ymdyrchael</i> 'to raise oneself (to kingship)'
		1 <i>ymwneuthur yn vrenhin</i> 'to make oneself king'
		1 <i>ymwneuthur yn iach</i> 'to save oneself'
		1 <i>ymwassanaethu</i> 'to serve oneself'
	4	12 <i>ymrydhau</i> 'to free oneself'
		1 <i>ymroddi</i> 'to give oneself, to submit oneself'
		<i>ymostegu</i> 'to calm oneself, to maintain silence'
self-awareness	2	<i>ymadnabot</i> 'to know oneself'
self-improvement	1	<i>ymbrofi</i> 'to prove oneself'
self-love	3	<i>ymhoffi</i> 'to praise oneself'
	1	<i>ymuoli</i> 'to praise/admire oneself'
self-criticism	1	<i>ymddiheuraw</i> 'to excuse oneself'
	2	<i>ymgeryddu</i> 'to reproach oneself, to punish oneself'
	1	<i>ymgyfyawnhau</i> 'to justify oneself'
negative self-directed actions	1	<i>ymdoddi</i> 'to consume oneself'
	1	<i>ymlycru</i> 'to corrupt oneself'
self-punishment	1	2 <i>emboeni</i> 'to punish oneself'
		<i>ymgosbi</i> 'to punish oneself'
suicide	1	1 <i>ymdihennidio</i> 'to execute oneself'
		2 <i>ymgroggi</i> 'to hang oneself'

I argued above (section 2.2) that in the combination *ym*-verb + *X hun* the prefix codes co-reference, while *X hun* functions as an actor-oriented intensifier added for role disambiguation. As stated by Gast and Siemund (2006: 365–367, 370), the intensifier blocks middle readings of polyfunctional verbal (or pronominal) middle markers, stating who is the intentional agent. This is confirmed by the fact that all instances found in the database are reflexive or belong to middle situation types, whereas reciprocal *ym*-verbs are almost absent. One example is *ymguro* 'to beat (one another or oneself)' (*Ystoriau Saint Greal*, NLW MS. Peniarth 11, page 239^v: 9), which in the passage in question is clearly reciprocal.

Table 4 lists them according to the increasing markedness of co-reference, starting with typical middle situation types, covering a number of different

positive and negative self-directed actions and ending with the highly marked verbs denoting suicide. Of course, it is sometimes difficult to determine, for the verbs located in the middle of the table, where the benefactive middle ends and the full reflexive starts. While *ymdyrchafel* ‘to raise (oneself)’ belongs to the middle situation types when denoting a body movement, it is instead benefactive and self-directed when denoting a metaphorical movement such as a rise in rank (‘to raise oneself to kingship’). *Ymroddi* ‘to give oneself, to submit oneself’ is, with its 16 attestations across different genres, the most frequent verb.

The unprefixed verbs listed in Table 5 cover the same semantic fields, i.e. positive and negative self-directed actions, including even the more detailed semantics like self-awareness, self-punishment or suicide. As with the *ym*-verbs, synonymous or nearly synonymous verbs are available for several meanings, implying that the semantic scope of both groups is actually relatively small.

Table 5: Semantics and frequency of simple verbs + *X hun* in “reflexive” situations. PRON indicates that co-reference is expressed by an infixed pronoun or a possessive adjective, rather than by *X hun*.

Rhyddiaith y 13eg Ganrif	
reflexive, “neutral”	1 1 PRON <i>bwrw X hun</i> ‘to throw oneself’ (to the ground) 1 PRON <i>ffustyaw X hun</i> ‘to beat oneself’ (of a bell) 1 PRON <i>kymunaw X hun</i> ‘to communicate oneself’ (religious) PRON <i>rhwymyo X hun</i> ‘to bind oneself’ (by contract)
benefactive / positive self-directed actions	5 8 PRON <i>amdiffyn X hun</i> ‘to defend oneself’ 1 PRON <i>cymorth X hun</i> ‘to help oneself’ 3 PRON <i>gwneuthur X hun yn iach</i> ‘to make oneself safe’ 1 PRON <i>nerthau X hun</i> ‘to help oneself’ 2 <i>gwneuthur X hun yn iach</i> ‘to make oneself safe’ 1 <i>amdiffyn X hun</i> ‘to defend oneself’ 2 <i>iachau X hun</i> ‘to save oneself’ 2 <i>gwneuthur X hun yn iach</i> ‘to make oneself safe’
self-awareness	1 <i>adnabot X hun</i> ‘to know oneself’
self-improvement	1 PRON <i>ymendio X hun</i> ‘to amend oneself’ ²¹ 1 <i>ardymheru X hun</i> ‘to moderate oneself’ 1 <i>kymedroli X hun</i> ‘to moderate oneself’

²¹ *Ymendáu* ‘to rectify, improve’, with its variants *amendio*, *emendio* and *mendio*, is not an *ym*-verb, but a borrowing from Old French *amender* ‘to correct’. It also constructed both transitively and intransitively.

Table 5 (continued)

		Rhyddiaith y 13eg Ganrif	
self-love	1	PRON	<i>ganmawl X hun</i> 'to praise oneself'
	2		<i>moli X hun</i> 'to praise oneself'
self-criticism	1	PRON	<i>angreiffto X hun</i> 'to reproach oneself'
	1	PRON	<i>angreitho X hun</i> 'to reproach oneself'
	1	PRON	<i>barnu ehun</i> 'to judge oneself'
	1	PRON	<i>galw X hun</i> 'to call oneself (a wretch)'
	1	PRON	<i>kymryt X hun</i> 'to take oneself (for a fool)'
negative self-directed actions	1	PRON	<i>gwatwaru X hun</i> 'to ridicule oneself'
	1	PRON	<i>gweled X hun</i> 'to see / consider oneself (as ugly)'
	1	PRON	<i>roddi X hun</i> 'to give oneself (in danger)'
	1	PRON	<i>taraw X hun</i> 'to strike oneself'
	1	PRON	<i>twyllaw X hun</i> 'to cheat oneself'
	1	PRON	<i>ymelldigo X hun</i> 'to curse oneself'
	1		<i>cablu X hun</i> 'to blame oneself'
1		<i>cnoi X hun</i> 'to chew up oneself'	
self-punishment	2	PRON	<i>poeni X hun</i> 'to punish oneself'
suicide	2	3	PRON <i>brathu X hun</i> 'to stab oneself'
	1	3	PRON <i>lladd X hun</i> 'to kill oneself'
	1		PRON <i>bot X hun yn y lad</i> 'to kill oneself' (lit. 'to be oneself at one's killing')
			<i>llad X hun</i> 'to kill oneself'

Table 6: Light-verb constructions with *gwneuthur* 'to do'.

		Rhyddiaith y 13eg Ganrif	
suicide	1	1	<i>gwneuthur X hun y leith</i> 'to effect oneself one's death'
		1	<i>gwneuthur y leith X hun</i> 'id.' or 'effect one's own death'
	1	1	<i>gwneuthur X hun y dihenyd</i> 'to effect oneself one's death'

Finally, both groups also contain light-verb constructions with *gwneuthur* 'to do' (Table 6). In these examples, co-reference is not marked on the verb or verb phrase, but in the possessive that precedes the associated noun – see below, (63) to (66).

There are almost no unprefixed reflexive verbs with “neutral” semantics, i.e. in which the effect of the verbal action on the agent is neither explicitly positive nor negative. This statement is however not based on their verbal semantics alone, but on the precise contexts in which the verbs occur. Thus *gweled X hun* ‘to see / consider oneself as sth.’ or *galw X hun* ‘to call oneself sth.’ are neutral in principle, but in their actual contexts they convey a negative judgement of the agent about himself. *Ffustyaw X hun* ‘to beat oneself’, on the other hand, would be a negative self-directed action in the case of a human agent, but in the only attestation found in the corpus it refers to a bell.

The verbs listed in (30) occur with and without *ym-*, as in (31) and (32). Their semantics are largely synonymous, as some of them occur in similar contexts or in parallel versions of the same text.

- (30) *ymadnabod* : *adnabot X hun* ‘to know oneself’
ymboeni : PRON *poeni X hun* ‘to punish oneself’
ymuoli : *moli X hun* ‘to praise oneself’
ymwneuthur yn iach : *gwneuthur X hun yn iach* ‘to make oneself safe’

- (31) *Na vawl dy hun yn ormod ac na chapla*
 NEG praise_{2SG.IMPV} 2SG_{REFL} too much and NEG reproach_{2SG.IMPV}
dy hun yn ormod.
 2SG_{REFL} too much
 ‘Do not praise yourself too much and do not reproach yourself too much.’
 (*Cynghorau Catwn*; NLW MS. Llanstephan 27, page 168^f: 16)

- (32) *Nac ym-uawl du hun ac nac ym-hoffa du hun.*
 NEG PV·praise_{2SG.IMPV} 2SG_{INTS} and NEG PV·admire_{2SG.IMPV} 2SG_{INTS}
 ‘Do not praise yourself and do not admire yourself.’ (*Cynghorau Catwn*;
 NLW MS. Peniarth 3 part ii, page 38: 11)

If over time one strategy of reflexive marking is replaced by another, it is to be expected that both variants co-occurred during a transitional period. It is thus not surprising that verbs can be found both with and without prefix. In contrast to this, some verbs, like those denoting different types of suicide, always occur either prefixed or unprefixed. Some of these even contradict the typological rule according to which, in languages with two different reflexive markers, the heavier marker is used for the more marked situations (Kemmer 1993: 62). In this sense, the *ym*-verbs *ymdihennidid* ‘to execute oneself’ and *ymgrogi* ‘to hang

oneself' are atypical. On the other hand, *brathu X hun* 'to stab oneself' and *lladd X hun* 'to kill oneself' always occur unprefixes.

The reason for this unexpected behaviour may be that the corresponding *ym*-verb is widely used as a reciprocal or has already been lexicalised with another meaning. The default reading of verbs denoting different kinds of killing or killing with various kinds of weapons or instruments is reciprocal (33).

- (33) *llad* 'to kill' : *ymladd* 'to fight' (< *'to kill each other')
saethu 'to shoot, to fire' : *ymsaethu* 'to fire at each other'
taraw 'to strike' : *ymdaraw* 'to strike one another'
gwan 'to stab, to kill' : *ymlwan* 'to joust, to fight in single combat'
brathu 'to stab' : **ymvrathu* 'to stab one another'

If a hypothetical **ymvrathu* were derived from transitive *brathu* 'to stab', its most likely meaning would be 'to stab one another' and not 'to stab oneself'. This is not an issue in the case of *ymgrogri* 'to hang oneself' and *ymdiheiddio* 'to execute oneself', as mutual hanging or executing is not possible.²²

Another lexicalised *ym*-verb is *ymwelet* 'to meet each other', so that *PRON gwelet X hun* translates Latin *se uidens* (34).

- (34) *gwelet* 'to see' : *ymwelet* 'to meet each other'

In the following case, the situation is even more complex, as the adjective *iach* 'healthy, well, whole' is the basis of four different verbs, one of which is an *ym*-verb (35). Although the semantics of *ymiachau* 'to bid farewell' are reciprocal, they cannot be derived from the underlying adjective or the corresponding unprefixes verbs (36), because in that case the meaning should be 'to heal each other'. The meaning 'to bid farewell' is rather based on the concept of 'leaving

²² An interesting typological parallel can be found in Modern Greek verbs with the meaning 'to kill oneself, to commit suicide'. Besides the compound αυτοκτονώ (active) 'to commit suicide', in which co-reference is expressed by the first constituent αυτο- 'self-', there are a number of other verbs with 'middle morphology', i.e. their inflection as medio-passives signals that the agent performs the action on him- or herself. Several verbs have additional meanings typical for other middle situation types e.g. spontaneous events like 'to perish', or intransitive 'to smash': σκοτώνω (active) 'to kill': σκοτώνομαι (middle) 'to kill oneself; to die, to perish, to struggle', κρεμάω (active) 'to hang': κρεμέμαι (middle) 'to hang oneself', απαγχονίζω (active) 'to hang': απαγχονίζομαι (middle) 'to hang oneself', τσακίζω (active) 'to break, to squeeze': τσακίζομαι (middle) 'to smash, to struggle, to kill oneself'.

each other in a healthy condition’ or ‘wishing each other health’. The lexicalised semantics of *ymiachau* thus seem to prevent a reflexive interpretation. On the contrary, the reflexive *iachau X hun* ‘to save oneself’, based on the unprefix verb, displays the expected semantics.

- (35) adjective *iach* ‘healthy, well, whole’
 transitive *iachaf, iachu* ‘to heal, cure’
 transitive & *iachâf, iachâu* ‘to make whole(some), heal, cure; save’
 intransitive
 reflexive *iachau X hun* ‘to save oneself’
 reciprocal *ymiachau* ‘to bid farewell’

- (36) *ac heb ohir kymryt y bererin ffonn ae*
 and without delay take_{3SG.PRET} DEF pilgrim staff and-3SG_{MASC.POSS}
balmidyden. a ymiachau ae dylvyth a
 palm.branch and bid.farewell_{VN} with-3SG_{MASC.POSS} family PTCL
oruc. ac yr mor yd aeth.
 do_{3SG.PRET} and DEF sea PTCL go_{3SG.PRET}
 ‘and without delay he took his pilgrim staff and his palm branch, and he bade farewell to his family, and went to the sea.’ (*Ystoria Bown de Hamtwn*; NLW MS. Peniarth 5, page 148^v)

In Middle Welsh, *ym*-verbs are usually either reflexive or reciprocal but rarely both at the same time. This is different in German and French, where *sich* and *se* often mark both categories. In Welsh, it was only as late as the sixteenth century that some reciprocal verbs started to be used also as reflexives. For instance, *ymddiddan* ‘to speak with each other, to converse’ acquired the additional meaning ‘to amuse oneself’ (Irslinger 2017c: 119). Another example is *ymadnabod* (reciprocal ‘to know each other’), which in *Cynghorau Catwn* occurs as the equivalent of *adnabot X hun* ‘to know oneself’ (37). In addition to this single reflexive use, there are several attestations of reciprocal *ymadnabot* ‘to know each other’.

- (37) PRON *adnabot X hun* ‘to know oneself’ = *ymadnabot* ‘to know oneself’
ymadnabot ‘to know each other’

4.2 Syntax

From the lists above it becomes clear that in most cases the equivalent of an intransitive *ym*-verb with reflexive function is not the unprefix verb + *X hun*, but rather the construction “pronoun + unprefix verb + *X hun*”, whereby the pronoun codes the direct object of the transitive verb.

The pronoun is ambiguous with regard to co-reference in the third person. Thus the reader or listener has to infer from the context that the third person plural pronoun *eu* is used co-referentially in (38) and (39), but not in (40). As a consequence, intensifiers are frequently added for reference disambiguation.

- (38) [A *gwedy gwelet o antigonus . . . yr aerua honno. neilltuaw a oruc*]
a' y oreugwyr gyd ac ef. y geisiaw
 and 3SG_{MASC.POSS} best.men with 3SG_{MASC} to try_{VN}
ev hamdiffin.
 3PL_{POSS} defend_{VN}
 [‘And after Antigonus . . . had seen this slaughter, he drew aside,] and his leading men with him, to try to defend themselves.’ (Parry 1937: 12 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 5^v: 27])

- (39) [A *gwedy nachaffant hynny. wynt a ervynnassant cannyat y adeiliat caer onadunt ev hun. kyulet achroen ech.*]
y geisiaw ev hamdiffin rac ev gelynnyon.
 to try_{VN} 3PL_{POSS} defend_{VN} from 3PL_{POSS} enemies
 [‘And when they did not / get that, they asked for permission to build a fortress of their own, as broad as an ox-hide,] to try to defend themselves from their enemies.’ (Parry 1937: 12 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 55^v: 19])

- (40) [A *gwedy eu bod tridieu yn ymlat ar kestyl o bop ryw vod. ar gwyr. y mewn yn ymlad ac wynt yn wraul ac yn llaurus.*]
anvon a orugant ar brutus y erchi idaw dyuot eu
 send_{VN} PTCL do_{3SG•PRET} to Brutus to ask_{VN} to_{3SG.MASC} come_{VN} 3PL_{POSS}
hamdiffyn.
 defend_{VN}
 [*cans ny ellynt wy ymderbyneit ac wynt rac meynt y nyueroed allan.*]
 [‘And after they had fought against the castles for three days in every sort of way, and the men within had fought them bravely and laboriously,] they sent to Brutus to ask him to come to defend them, [for, because of

the great numbers outside, they could not resist them.’] (Parry 1937: 12 [Brut y Brenhinoedd; BL Cotton Cleopatra MS. B V part i, page 6^r: 10–11])

Pronouns coding coreferential or non-coreferential direct objects occur in two different constructions, i.e. with a finite verb or with a verbal noun:

- an infixed pronoun denoting the object precedes a finite transitive verb (41) (Evans 1964: 55);
- a possessive adjective precedes a noun (42).

This second construction is identical to the one discussed above in (21), except that the noun is replaced by a verbal noun. Both variants, i.e. POSS+VN and POSS+NOUN, occur in the parallel versions of *Saith Doethion Rhufain*, (42) and (43). On the formal level, the intensifier following a POSS+VN construction is, of course, adnominal. Nevertheless, these constructions, which outnumber those with finite verbs by far, will be discussed together with reflexive finite verbs.

Both the infixed or independent pronoun and the possessive marker agree with the subject (which is co-referent with the object) and the intensifier with regard to person, number and gender.

- (41) *Yna ef a’ e trewis e hun*
 Then 3SG_{MASC} PTCL 3SG_{MASC.INFX} strike_{3SG.PRET} 3SG_{MASC.INTS}
a’ e yluin y dan benn y vronn.
 with 3SG_{MASC.POSS} beak under 3SG_{MASC.POSS} breast
 ‘Then it [a bird] struck (it) itself with its beak under its breast.’ (*Ystoriau Saint Greal*; NLW MS. Peniarth 11, page 63^v: 24)

- (42) [*a chyndrwc yd aeth arnaw ef hynny.*]
a’ e vrathu e hun a wnaeth y dan
 and 3SG_{MASC.POSS} stab_{VN} 3SG_{MASC.INTS} PTCL do_{3SG.PRES} under
y vron a’ e gyllell
 3SG_{MASC.POSS} breast with 3SG_{MASC.POSS} knife
 [*yny dygwyd yn varw y’r llawr.*]
 [‘And he took it so ill, that] he stabbed (him) himself under his breast with his knife, [until he fell dead to the ground.’] (*Saith Doethion Rhufain*; Oxford Jesus College MS. 111, page 131^r (541): 34)

- (43) *a' e vrath e hun a wnaeth am benn*
 and 3SG_{MASC.POSS} stabbing 3SG_{MASC.INTS} PTCL do_{3SG.PRET} around top
y vron
 3SG_{MASC.POSS} breast
 'and he stabbed (him) himself under his breast with his knife.' (*Saith Doethion Rhufain*; Oxford Jesus College MS. 20, page 56^{av}: 1)

The very same constructions code pronominal direct objects of transitive verbs, cf. (44) with infixed pronouns and finite verbs and (40) above with the POSS+VN construction.

- (44) [*Ac ewythyr ydaw ef ehvn adylyhei gwledychu gwedy custennyn: ac ef a ryuelawd a hwnnw.*]
ac a' y delhiis ac a' y
 and PTCL 3SG_{MASC.INFX} capture_{3SG.PRET} and PTCL 3SG_{MASC.INFX}
rodes yng karchar.
 put_{3SG.PRET} in prison
 ['And his uncle should have ruled after Constantine; and he fought with him] and captured him and put him in prison.' (Parry 1937: 104 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 96^v: 22–23])

The use of ambiguous pronouns plus disambiguating intensifiers is thus essentially the same as in Old English, as in (2) to (4) above. This strategy is still the predominant one in Middle Welsh (Table 7).

Both constructions are found in Modern Welsh as variants, cf. (45a) and (46a) with additional co-referential pronouns versus (45b) and (46b) without them.²³ Contrary to the development in English, the older strategy has not yet completely vanished in Welsh. Despite this, the pronominal constructions do not figure in all grammars of Modern Welsh.

- (45) a. *Fe i gwelodd ei hunan yn y drych.*
 PTCL 3SG_{FEM.INFX} see_{PRET} 3SG_{FEM.INTS} in DEF mirror
 'She saw herself in the mirror.' (Thomas 1996: 269)
 b. *Gwelodd ei hunan yn y drych.*
 see_{PRET} 3SG_{FEM.REFL} in DEF mirror
 'She saw herself in the mirror.' (Thomas 1996: 269)

²³ See Thomas (1996: 269), Borsley, Tallerman, and Willis (2007: 222), Poppe (2009: 254, footnote 7).

- (46) a. *Rwy 'n gallu fy ngweld fy hun*
 be_{1SG PRES} PRED can_{INV} 1SG_{POSS} see_{VN} 1SG_{INTS}
yn y drych.
 in DEF mirror
 'I can see myself in the mirror.' (Poppe 2009: 254, footnote 7)
- b. *Rwy 'n gallu gweld fy hun yn y drych.*
 be_{1SG PRES} PRED can_{INV} see_{VN} 1SG_{REFL} in DEF mirror
 'I can see myself in the mirror.' (Poppe 2009: 254, footnote 7)

Table 7 gives the distribution of *ym*-verbs and simple verbs followed by *X hun* coding “full reflexive” events from both corpora (cf. Tables 4, 5 and 6 above).

X hun can only be analysed as a reflexive marker in as few as 14 cases, and even some of these are controversial. One example is contained in *Rhyddiaith y 13eg Ganrif*, the others are found in *Rhyddiaith Gymraeg*. Accordingly, these 13 cases of reflexive *X hun* constitute 0.68 % of the 1,908 instances of *X hun* found in this corpus. Although this number validates the dating of the beginning of the use of *X hun* as a reflexive marker in the second part of the Middle Welsh period, it is certainly insufficient to justify Evans' (1964: 89) labelling of *X hun* as the Middle Welsh “reflexive pronoun”.

Table 7: Distribution of *ym*-verbs and simple verbs followed by *X hun* coding “full reflexive” events.

Genre	Total	light-verb		ym-verb		simple verb	
		co-ref. not marked on VP		pron. + finite verb		poss. + vn	finite verb
		<i>X hun</i> = intensifier				<i>X hun</i> = reflexive marker	
Genealogy							
Geography	2		1			1	
Grammar							
History	34	3	17	2	13		
Law	9		9				
Mabinogion							

Table 7 (continued)

Genre	Total	light-verb	ym-verb	simple verb		
		co-ref. not marked on VP		pron. + finite verb	poss. + vn	finite verb
			<i>X hun</i> = intensifier			<i>X hun</i> = reflexive marker
Medicine						
Nat. Hist.	1		1			
Religion	34	2	12		14	4
Romance	20		6	3	9	2
Wisdom	23		12	3	3	5
				8	39	12
Total	123	5	57	47		14
Percentage	100,00	4,07	46,34	38,21		11,38

4.3 *X hun* as a reflexive marker in the corpora

Table 8 lists the 14 instances of *X hun* as a reflexive marker together with additional information on the manuscripts that contain them, their dates according to *Rhyddiaith Gymraeg 1300–1425*²⁴ based on Huws (2000: 58–64) and the forms in which they are attested.

4.3.1 Distribution and date

The attestations range from the beginning to the end of the Middle Welsh period and are found both in earlier and later manuscripts. Reflexive *X hun* is slightly more frequent in fourteenth and fifteenth-century manuscripts (mainly the *Red Book of Hergest* and the *Red Book of Talgarth*), but it does not seem that older textual versions were systematically modernised during the process of copying. It is of course possible that the conservative written registers preserve features that had already largely been abandoned in the spoken language.

²⁴ See <http://www.rhyddiaithganoloesol.caerdydd.ac.uk/en/manuscripts.php>

Table 8: Attestations of reflexive *X hun* according to manuscripts (N = negator, IMPV = imperative 2SG).

Text	Manuscript	Date	Form	Verb
Bown de Hamtwn	Pen. 5 Jesus 111	1350 c. 1375–1425	N- 1SG	<i>moli X hun</i> ‘to praise oneself’
Credo Athana-sius, Introduction	Pen. 5	1350	VN	<i>adeilat X hun</i> ‘to edify oneself’
Cronicl Turpin	Pen. 5 Pen. 9 Jesus 111	1350 1300–1350 c. 1375–1425	3SG	<i>amdifffyn X hun</i> ‘to defend oneself’
Cyngh. Catwn	Pen. 3 pg. ii	c. 1275–1325	IMPV	<i>kymedroli X hun</i> ‘to moderate oneself’
Cyngh. Catwn	Llanst. 27	c. 1375–1425	IMPV	<i>adnabot X hun</i> ‘to know oneself’
Cyngh. Catwn	Llanst. 27	c. 1375–1425	IMPV	<i>ardymheru X hun</i> ‘to moderate oneself’
Cyngh. Catwn	Llanst. 27	c. 1375–1425	N- IMPV	<i>cablu X hun</i> ‘to blame oneself’
Cyngh. Catwn	Llanst. 27	c. 1375–1425	IMPV	<i>moli X hun</i> ‘to praise oneself’
Delw'r Byd	Jesus 111	c. 1375–1425	3PL	<i>cnoi X hun</i> ‘to chew up oneself’
Gwyrtheu Mair	Pen. 14	1250–1300	N- 2SG	<i>llad X hun</i> ‘to kill oneself’
Y Groglith	Pen. 7	c. 1275–1325	N- 3SG	<i>gwneuthur X hun yn iach</i> ‘to make oneself safe’
Y Groglith	Shrewsb. 11	c. 1375–1425	IMPV	<i>gwneuthur X hun yn iach</i> ‘to make oneself safe’
Y Groglith	Pen. 5 Llanst. 27	1350 c. 1375–1425	IMPV	<i>iachau X hun</i> ‘to save oneself’
Y Groglith	Pen. 5	1350	VN	<i>iachau X hun</i> ‘to save oneself’

Only *Y Groglith* shows some variation on the same passage, which renders Matthew 27:42 in (47). In the light-verb construction *gwneuthur X hun yn iach* ‘to make oneself safe’, *X hun* is used with a reflexive function in the oldest manuscript, Peniarth 7, in (48). On the other hand, the two younger versions make use of pronouns, as in (49) and (50). Peniarth 5 uses the synonymous verb *iachau* with reflexive *X hun*, in (51). Lastly, *Efengyl Nicodemus* adds another sentence expressing the same content again with a light-verb construction + intensifier, featuring the prefixed verb *ywneuthur* (52).

(47) [*alios salvos fecit*]

se ipsum non potest saluum facere

REFL_{ACC} INTS_{ACC.SG} NEG can_{3SG.PRES} safe_{ACC.SG} make_{INF}

‘He made others safe;] (him) himself he cannot make safe.’ (Weber 2007 [Matthew 27: 42; *Biblia Sacra Vulgata*])²⁵

(48) *ereill a wna ef yn yach ac ny*
others PTCL make_{3SG.PRES} 3SG_{MASC} PRED safe and NEG

wna e hvn

make_{3SG.PRES} 3SG_{MASC.REFL}

‘He makes safe others, and he doesn’t make himself (safe).’ (*Y Groglith*; NLW MS. Peniarth 7, page 58^v (215): 2)

(49) [*Ereill heb wy a wnaei ef yn iach.*]

ac ny dichawn y wneuthur e hun.

and NEG be.able_{3SG.PRES} 3SG_{MASC.POSS} make_{VN} 3SG_{MASC.INTS}

‘Others, they said, he saved] and he is not able to save (him) himself.’ (*Y Groglith*; NLW MS. Llanstephan 27, page 105^v: 17)

(50) *Ereill a wna yn iach ac ny ’s*
others PTCL make_{3SG.PRES} PRED safe and NEG 3SG_{MASC.INF}

gwna e hun

make_{3SG.PRES} 3SG_{MASC.INTS}

‘He makes safe others, and he doesn’t make (him) himself (safe).’ (*Y Groglith*; Shrewsbury MS. 11, page 113: 16)

²⁵ See also <https://www.bibelwissenschaft.de/startseite/wissenschaftliche-bibelausgaben/vulgata/>

- (51) [
- Ereill a wna ef yn iach.*
-]

ac ny eill iachau e hun.

and NEG can_{3SG.PRES} save_{VN} 3SG_{MASC.REFL}

‘[He saves others], and cannot save himself.’ (*Y Groglith*; NLW MS. Peniarth 5, page 7^v: 20)

- (52) [
- Ereill a wnaei ef yn iach ac ny dichawn y wneuthur e hun.*
-]

Ym-wnaet yn iach e hun.

PV·do_{3SG.IMP} PTCL safe 3SG_{MASC.INTS}

‘[He made safe others, and he is not able make himself (safe)] He shall make himself safe.’ (*Efengyl Nicodemus*; NLW MS. Peniarth 5, page 32^r: 14)

Texts with reflexive *X hun* usually also contain instances of the pronominal constructions, unless they are very short and thus do not possess many reflexive verbs altogether. The only exception is *Cynghorau Catwn*, which is the text with the highest number of reflexive *X hun* in the corpus.

The earliest attestation is (53) from *Gwyrtheu Mair* in Peniarth 14, which Huws (2000: 58) dates to the second half of the thirteenth century. GPC gives 1250 as a date for the text, i.e. the beginning of this period. Evans (1964: 89) points out that the expected form with an infix pronoun would be **ony'th ledy du hun*.

The same text has two other reflexive constructions, one with the POSS+VN construction (54) and one with an *ym*-verb (55). The latter is replaced by a POSS+VN construction in the later version in Llanstephan 27 (56).

- (53) [
- na elly caffael yechyt am e pechaut ry wnaethost*
-]

ony ledy duhun

unless kill_{2SG.PRES} 3SG_{MASC.REFL}

‘[you cannot get redemption from the sin you have done,]’ ‘unless you kill yourself’ (*Gwyrtheu Mair*; NLW MS. Peniarth 14, Jones 1941: 24)

- (54) [
- Llawer hep ef a wnaeth o drwc*
-]

ac en diwethaf e lad e hun.

and in last 3SG_{MASC.POSS} kill_{VN} 3SG_{MASC.INTS}

‘[He did, said he, a lot of evil,]’ ‘and in the end he killed himself’ (Jones 1941: 25 [*Gwyrtheu Mair*; NLW MS. Peniarth 14])

- (55) *Ac ena e dechreuws e vicedonus em·boeni*
 and then PTCL begin_{3SG.PRET} DEF vicedominus PV·punish_{VN}
ehun
 3SG_{MASC.REFL}
 ‘and then the vicedominus started to punish himself’ (Jones 1939: 148
 [*Gwyrtheu Mair*; NLW MS. Peniarth 14])
- (56) *Ac yna o newyd y dechreuawd teophilus y*
 and then anew PTCL begin_{3SG.PRET} Teophilus 3SG_{MASC.POSS}
boeni e hun
 punish_{VN} 3SG_{MASC.INTS}
 ‘and then Teophilus began to punish himself anew’ (*Gwyrtheu Mair*; NLW
 MS. Llanstephan 27, page 176^f: 10)

This isolated example is followed by the instances contained in NLW MS. Peniarth 5 (*White Book of Rhydderch*). Of these, Evans (1964: 89) cites (57) from *Credo Athanasius* following Lewis’ analysis of the passage. According to Lewis (1930: 193), this text was translated in the second half of the 13th century. *Adeilat e hun* is found in the introduction, which was not part of the Latin text, but was drafted by the Welsh translator. While admitting that the omission of *y* could be a scribal mistake, Lewis prefers to consider *adeilat ehun* as an early example of the reflexive use of *X hun*. He argues that this use, which had become very common by 1615, had to have started long before then (Lewis 1930: 195).

On the contrary, GPC (s.v. *hun*², section b) considers it as a scribal mistake and lists the example as a POSS+VN construction (58).

- (57) *Pob cristaun weithonn a dyly adeilat e hun*
 every Christian now PTCL must_{3SG.PRES} build_{VN} 3SG_{MASC.REFL}
 [*truy weithredoed da yn temyl y Duv a hymny yn gyuuch ac y carhaedo truy*
gret a gobeith a charyat teyrnas gvlad nef.]
 ‘Every Christian now has to build himself’ [‘through good works into a
 temple to God and this so high that he will achieve through belief and
 hope and love the kingdom of heaven.’] (*Credo Athanasius* [Introduction];
 NLW MS. Peniarth 5, page 48^v: 13)
- (58) *Pob cristaun . . . a dyly y [drll.]²⁶ adeilat ehun*
 (NLW MS. Peniarth 5, page 14g. B v. 196)

²⁶ drll. = *darllener, darlleniad* ‘read(ing), version’.

Another interesting case is that of *amddiffyn X hun*²⁷ in (59), which has exact parallels in NLW MS. Peniarth 9, page 1^v: 4 and Oxford Jesus College MS. 111, page 95^v (400): 9. The reflexive use of *e hun* in (59) is at variance with 13 instances of the POSS+VN construction as in (60) from *Brut y Brenhinoedd*, *Brut y Tywysogion*, *Ystoria Carolo Magno: Rhamant Otfel* and *Ystoriau Saint Greal*.

- (59) [*Canys rolond a dugassei gantaw trossawl troydic hir.*]
ac a hwnnw yd amddiffynnwys e hun educher.
 and with DIST PTCL defend_{3SG+PRET} 3SG_{MASC.REFL} till evening
 [‘for Rolond had brought with him a long twisted bar,] and with that he
 defended himself until the evening’ (Williams 1892: 463 [*Ystoria Carolo
 Magno: Chronicl Turpin*; NLW MS. Peniarth 5, page 74^r (63): 34])
- (60) *ym·rodi a wnaethant y eu hamdiffyn e hunein*
 PV·submit_{VN} PTCL do_{3PL.PRET} to 3PL_{POSS} defend_{VN} 3PL_{INTS}
o hynny allan
 from DIST on
 ‘they submitted themselves to defend themselves from then on’ (*Brut y
 Brenhinoedd*; Oxford Jesus College MS. 111, page 39^r (154): 27)

4.3.2 Change through linguistic convergence?

Strikingly, no instances of reflexive *X hun* are found in “native” texts like the *Mabinogion* or the laws, but all of them occur in translations or adaptations from Latin or Old French. One could speculate that the change in Middle Welsh was at least partly triggered by contact influence, but it is hard to find any evidence to sustain this claim. This may be due to the following reasons.

In some cases, both the Latin texts and the corresponding Welsh translations were extremely popular, so that it is impossible to determine which version underlies a translated text. Later versions may not necessarily rely on the Latin original, but rather on other translations.

But even in cases where the source is clear, the Welsh translators frequently rendered the content of a passage in their own words rather than producing verbatim translations.

²⁷ The *am-* in *amddiffyn* (sometimes *ymddiffyn*) retains the original prepositional meaning ‘about, at all sides’ (Vendryes 1927: 50). *Amdiffyn* is constructed mostly transitively and thus differs from *ym-*verbs containing the grammaticalised prefix, which are predominantly intransitive.

As the example of *Y Groglith* has shown, different versions use different constructions, all of which are well rooted in the language. The reflexive use of *X hun* does not seem to be triggered by the underlying Latin (or Greek) text.

The Cordeilla passage in the Cotton Cleopatra *Brut* is another example of the independence of the Welsh version (62), which contains two verbs denoting suicide that do not figure in the Latin text (61). Both of them use the POSS-VN construction.

- (61) [*Eam quoque ad ultimum captam in carcerem posuerunt*
ubi ob amissionem regni dolore obducta
 where by loss_{ACC} kingdom_{GEN} grief_{ABL} overwhelmed_{ABL.SG.FEM}
sese interfecit.²⁸
 REFL_{ACC} kill_{3SG.PRET}
 ‘[Finally they captured and imprisoned her] where overwhelmed by grief at the loss of her kingdom she killed herself.’ (Reeve and Wright 2007: 45 § 32 [Geoffrey of Monmouth’s *Historia regum Britanniae*])

- (62) [*A gwedy medyliaw ohoney am y hen deilyngdawd ry gollassei. ac nat oed obeith idi ymatkyuot ohynny.] o diruawr dolur hynny y gwnaeth hy hun y lleith. nyt amgen nogyd y brathu hy hun a chillell adan y bronnn yny gollas y heneid. ac yna y barnwyd mae dybrytta agheu y dyn y llad e hun*
 ‘And after thinking over her former dignity which she had lost, and she had no hope of raising herself out of it, out of exceeding grief over it **she did / effected herself her death** – that is, **she stabbed herself** with a knife under the breast so that she lost her life. And at that time it was considered the most ignominious death for a person **to kill himself.**’ (Parry 1937: 41 [*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 20^r: 14–16])²⁹

Latin *sese interfecit* has the reduplicated and thus emphatic reflexive marker *sese*, but no intensifier. The Welsh translator chose a light-verb construction, for which the analysis in (63) seems probable, especially in view of the similar light-verb constructions in (64) to (66).

²⁸ The First Variant Version (ed. Wright 1988: 27) has *sese interemit* ‘she killed herself’.

²⁹ Parry (1937: 41) reads *hun* ‘sleep’ and translates *y gwnaeth hy hun y lleith* ‘she slept the sleep of death’, assuming a metaphorical or euphemistic expression for committing suicide. Although *hun* ‘sleep’ is occasionally used this way in Middle Welsh, Parry’s analysis seems unconvincing in view of (64) and (65), which mention the instrument with which the act was carried out, and (66), which might be a POSS-NP construction.

Example (66) is most likely to be read as a POSS-NP construction ‘effecting his own dead’, but, as the object of a verb occasionally stands between the finite verb and the intensifier, it is not excluded that this passage corresponds to the others with a slightly modified syntax.

- (63) *y gwnaeth hy hun y lleith*
 PTCL do_{3SG.PRET} 3SG_{FEM.INTS} 3SG_{FEM.POSS} death
 ‘she effected personally her death’, lit. ‘she did herself her death’ (*Brut y Brenhinoedd*; BL Cotton Cleopatra MS. B V part i, page 20^r: 14–16)

- (64) [*A phan giglev pilatus hynny*]
y gwnaeth e hun y dihenyd a 'e gyllell.
 PTCL do_{3SG.PRET} 3SG_{MASC.INTS} 3SG_{MASC.POSS} death with 3SG_{MASC.POSS} knife
 [‘And when Pilatus heard this,] he effected his own death with his knife.’
 (*Ystoria Bilatus*; NLW MS. Peniarth 5, page 11^r: 14)

- (65) *y gorvc e hvn y leas a 'y gyllell*
 PTCL do_{3SG.PRET} 3SG_{MASC.INTS} 3SG_{MASC.POSS} death with 3SG_{MASC.POSS} knife
 ‘he effected his own death with his knife.’ (*Ystoria Bilatus*; NLW MS. Peniarth 7, page 63^v (236): 20)

- (66) [*Pan gigleu archelaus mab herot hynny; digallonni a oruc. a gossot y wayw yn y daear a mynet ar y vlaen*]
a gvneuthur y leith e hun.
 and do_{VN} 3SG_{MASC.POSS} death 3SG_{MASC.INTS}
 [‘When Archelaus the son of Herod heard this, he lost his heart and putting his lance on the ground and going on its point’] ‘effecting his own dead’ / ‘and effecting himself his death’ (*Ystoria Titus*; NLW MS. Peniarth 5, page 37^r: 48)³⁰

The Latin influence on *Gwyrtheu Mair* is more difficult to assess. The Latin text is transmitted in several slightly different versions. Example (68) expresses the order to kill oneself with the simple reflexive pronoun *te* followed by the intensifier *ipsum*. The emphatic pronoun *temet* in (67) is already present in Classical

³⁰ Cf. Ehrmann and Pleše (2011: 12.546) for the Latin text: *Herodes amputavit lanceam suam et fixit in terram et iactavit se super et mortuus est*. ‘Herod broke off his spear, fixed it in the ground, and threw himself over it and died.’

Latin. In Vulgar, Late and Middle Latin, pronouns enlarged by *-met* become increasingly frequent and were often fused together with the intensifier,³¹ like in the example.

Although the Welsh version is an independent renarration (69), the similarity of the Latin and Welsh verb phrases is striking, especially because the Welsh author was probably aware of the parallel structures of Latin *temet-ipse* and Middle Welsh *du-hun*. It is unlikely, however, that the Welsh author would choose to calque the Latin reflexive strategy after having significantly altered the whole passage.

- (67) [*Scias quum pro malis operibus quae gessisti. iam non potes salutem consequi nisi feceris quae dixero tibi. Abscide primum tua genitalia membra et deinde **interfice temetipsum.***

and then kill_{2SG.IMPV} 2SG_{REFL}-2SG_{INTS}

[‘Know that for the bad deeds you have done, you cannot obtain redemption unless you will do what I will say to you. First, cut off your genitals] and then kill yourself.’ (Neuhaus 1886: 38 [*The Pilgrim Girardus*; BL Cotton Cleopatra MS. C X])

- (68) *deinde **interime te ipsum***

then kill_{2SG.IMPV} 2SG_{REFL} 2SG_{INTS}

‘Then kill yourself!’ (Neuhaus 1886: 38 [*The Pilgrim Girardus*; BL Arundel MS. 346])

- (69) [*na elly caffael yechyt am e pechaut ry wnaethost*

ony **ledy duhun**

unless kill_{2SG.PRES} 2SG_{INTS}

[‘you cannot get redemption from the sin you have done,] unless you kill yourself’ (Jones 1941: 24 [*Gwyrtheu Mair*; NLW MS. Peniarth 14])

A second instance of emphatic reflexive + intensifier in the Latin text has no correspondence in the Welsh version at all (70, 71). Later on, the Latin reflexive verb is rendered by the POSS-VN construction (72, 73).

31 Cf. Väänänen (1981: 123), Puddu (2005: 206–223).

- (70) [At ille putans ueraciter eum sanctum esse Jacobum qui talia iuberet. arrepto ferro membra uirilia abscidit. ac postea per guttur suum ferrum trahens.]

semetipsum ad mortem uulnerauit.

REFL_{ACC}-INTS_{ACC} to death_{ACC} wound_{3SG.PERF}

['And he [Girardus] believed, that it was Saint Jacob who ordered this, and, seizing a knife, cut off his genitals. And after this, drawing the knife against his throat,] he hurt himself deadly.' (Neuhaus 1886: 38 [*The Pilgrim Girardus*; BL Cotton Cleopatra MS. C X])

- (71) *Sef a oruc enteu o debygu en wir*
 this.is PTCL do_{3SG.PRET} 3SG_{MASC} of think_{VN} PRED true
panyv yago ebostol oed ef gwneithur a orchymynassei
 that.is Jacob apostle be_{3SG.IMPF} 3SG_{MASC} do_{VN} PTCL order_{3SG.PLPF}
a marw vu.
 and dead be_{3SG.PRET}

'This is what he did, (as he was) really thinking that it was the apostle Jacob, he did what he had ordered and died.' (lit.: 'and he was dead.')

(Jones 1941: 24 [*Gwyrtheu Mair*; NLW MS. Peniarth 14])

- (72) *et quod ad extremum se peremisset.*

and that at end_{ACC} REFL kill_{3SG.PERF}

'and finally he had killed himself.' (Neuhaus 1886: 39 [*The Pilgrim Girardus*; BL Cotton Cleopatra MS. C X])

- (73) *ac en diwethaf e lad e hun.*

and in last 3SG_{MASC.POSS} kill_{VN} 3SG_{MASC.INTS}

'and in the end he killed himself' (Jones 1941: 25 [*Gwyrtheu Mair*; NLW MS. Peniarth 14])

In the following case, the correspondence between the Old French source (74) and its Middle Welsh translation (75) is rather close. However, while the idiom 'to praise someone to the value of one glove' does not seem to occur elsewhere in Middle Welsh, the verb *moli X hun* 'to praise oneself' is found also in the Llanstephan version of the *Cynghorau Catwn* (ex. 78). Theoretically, the Old French pronoun *me* could have triggered a pronominal construction in Middle Welsh, but instead *fu hunan* is used with reflexive function.

- (74) *jeo ne me preyse le vailant de un gant.*
 I NEG 1SG_{REFL} praise_{1SG.PRES} the_{MASC} worth of one glove
 ‘I do not praise myself to the value of one glove.’ (Stimming 1899: 68, l. 1797 [*Boeve de Haumtone*])
- (75) [*a ffan ymladom ony ladaf i dy benn di yr mawr a’m cledeu.*]
ny volaf fu hunan werth vn uanec.
 NEG praise_{1SG.PRES} 1SG_{REFL} worth one glove
 [‘and when we fight, if I do not cut off thy head, thou great fellow, with my sword,] I will not praise myself to the value of one glove.’ (Williams 1892: 539 [*Ystoria Bown de Hamtwn*; NLW MS. Peniarth 5, page 134^v (301): 22])

The following passage from *Imago mundi* by Honorius Augustodunensis contains a combination of the reflexive marker *se* and the intensifier *ipse* (76). To render the Latin *seipsos . . . corrodentes* ‘chewing up themselves’, the author of the Welsh translation *Delw y Byd* gives two apparently synonymous versions (77). Both verbs seem to occur only once and were thus most likely specifically created for this passage. Interestingly, the two verbs that the author decided to use (the *ym*-verb *ymdoddi* based on *toddi* ‘to melt’ and the transitive verb *chnoi* ‘to bite, to chew’) are both constructed reflexively. Even though these very verbs were custom-made to render the Latin version, both types of verbs already existed in Welsh. The *ym*-strategy was still productive at the time when the use of reflexive *X hun* started to spread.

- (76) [. . . *praesertim cum me non mihi soli, sed toti mundo genitum intelligam, omittens invidos tabescentes, non me*],
sed seipsos livido corde corrodentes
 but REFL_{ACC}-INTS_{ACC.PL} jealous_{ABL.SG.NEUT} heart_{ABL} chewing_{ACC.PL.MASC}
 [‘. . .above all, I understand not only my own birth, but the birth of the whole world, leaving aside grieving individuals, who are chewing up not me,] but themselves with a jealous heart. . .’ (Flint 1983: 48–49 [*Imago mundi*])
- (77) [*Ac yn bennaf oll pryt na dyallwyf i vvg geni y my hun. mwy noc y ’r holl vyt gan ysgaelussaw y dnyyon kyghorunnus.*]
ac a ym·dodant e hunein ac a gnoant e hunein
 and PTCL PV·melt_{3PL.PRES} 3PL_{INTS} and PTCL chew_{3PL.PRES} 3PL_{REFL}

o gallon gyghorynnvs
of heart jealous

[‘And above all, since I do not only understand my own birth, but of the whole world, neglecting jealous people] who gnaw themselves and who chew themselves up with a jealous heart.’ (*Delw’r Byd*; Oxford Jesus College 111, page 243^r (976): 14)

In conclusion, it seems that for the translation of reflexives, Middle Welsh authors resorted to a repertoire of constructions also found in non-translated texts and did not try to imitate³² the Latin or Old French structures.³³

4.3.3 Formal aspects

In no less than six cases, reflexive *X hun* occurs with a second singular imperative verb, cf. (78) to (81). All instances come from two texts only: four from the *Cynghorau Catwn* and two from *Y Grogllith*.

- (78) *Na vawl dy hun yn ormod ac na chapla*
NEG praise_{2SG.IMPV} 2SG_{REFL} too much and NEG reproach_{2SG.IMPV}
dy hun yn much
2SG_{REFL} too *ormod*.
‘Do neither praise nor reproach yourself too much.’ (*Cynghorau Catwn*;
NLW MS. Llanstephan 27, page 168^r: 16)

- (79) [*Pan gymhellow dolor di yn irloned rac kared dy weissyon.*]³⁴
kymedrola dy hvn hyt pan ellych arbet y rei
moderate_{2SG.IMPV.2SG} 2SG_{REFL} so that can_{2SG.SUBJ} forgive_{VN} DEF ones

³² Cf. Winford (2003: 63–65) on “structural convergence”, i.e. imitation of the syntactic structures of the contact language with the lexical means of one’s own language. A similar model is “replica grammaticalization”, developed by Heine and Kuteva (2003: 539).

³³ A different development took place in Breton, where the French influence was much stronger. The Breton prefix *em-* was equated with French *se* and became, combined with a pronoun, part of the preverbal reflexive and reciprocal marker Modern Breton *en em*, cf. Irslinger (2014b: 187, 199).

³⁴ Cf. Duff (1954: 602) for the Latin text: *Seruorum culpa cum te dolor urguet in iram, ipse tibi moderare, tuis ut parcere possis*. ‘If pain drives you in anger because of the fault of your servants, moderate yourself, so that you can forgive the ones belonging to you/your people’.

teu di.

2SG_{POSS} 2SG

[‘If pain drives you in anger because of the sin of your servants,] moderate yourself, so that you can forgive the ones belonging to you’ (*Cynghorau Catwn*; NLW MS. Peniarth 3 part ii, page 37:3)

- (80) *os crist wyt ti gwna di hunan yn iach*
 if Christ be_{2SG.PRES} 2SG do_{2SG.IMPV} 2SG_{REFL} PRED safe
 ‘if you are Christ, save yourself’ (lit. ‘make yourself safe’) (*Y Groglith*; Shrewsbury MS. 11, page 114:2)

- (81) [*Hwm a distryw temyl duw. ac ympen y tridieu a ’e hadeila.*]
iachaa dy hun.
 save_{2SG.IMPV} 2SG_{REFL}
 [‘The one who destroyed the temple of God and rebuilt it after three days’]
save yourself.’ (*Y Groglith*; NLW MS. Peniarth 5, page 7^v: 18)

In (82) and (83) from the *Cynghorau Catwn*, the function of the pronoun *di* following the verb is unclear. Objects of imperative verbs are invariably expressed by independent pronouns, not by infixed pronouns (82). In addition, independent pronouns occasionally code the objects of other verb forms (83) (Evans 1964: 49–50). Postverbal *di* could thus be the object of the verb expressed by the independent pronoun, while the disambiguating intensifier indicates its co-reference with the subject coded in the verbal ending.

Nevertheless, it seems more likely for *di* to code the subject and thus refer to the person expressed by the verb. *Cynghorau Catwn* contains three instances of this use with a transitive non-reflexive verb, like in (84).

- (82) *Ardymhera di du hvn o ’r gwin.*
 moderate_{2SG.IMPV} 2SG 2SG_{REFL} from DEF wine
 ‘Moderate (you) yourself from wine.’ (*Cynghorau Catwn*; NLW MS. Llanstephan 27, page 31: 3)

- (83) *Kanys ot atnabydy di dy hun doeth wyt.*
 for if know_{2SG.HAB} 2SG 2SG_{REFL} wise be_{2SG.PRES}
 ‘For if you know (you) yourself, you are wise.’ (*Cynghorau Catwn*; NLW MS. Llanstephan 27, page 32: 20)

- (84) *na chappla di arall am y bei*
 NEG blame_{2SG.IMPV.2SG} 2SG another for DEF mistake
a vo arnat ti dy hun.
 PTCL be_{3SG.PRES.SUBJ} on_{2SG} 2SG 2SG_{INTS}
 ‘Do (you) not blame another for the mistake that is on yourself.’
 (Cynghorau Catwn; NLW MS. Llanstephan 27, page 165^v: 15)

4.3.4 Transition from one system to another

To move from the old Middle Welsh system of reflexive marking to the new one, two simultaneous steps are necessary:

- loss of the object pronoun or the prefixed *ym-*
- reanalysis of *X hun* as object of the verb

This development is illustrated in (85) for the different structural types:

- (85) PRON + finite verb $e_{\text{REFL}} \text{ trewis } e \text{ hun}_{\text{INT}} \rightarrow \text{ trewis } e \text{ hun}_{\text{REFL}}$
 POSS-VN construction $y_{\text{REFL}} \text{ brathu } hi \text{ hun}_{\text{INT}} \rightarrow \text{ brathu } hi \text{ hun}_{\text{REFL}}$
ym-verb $ym_{\text{REFL}} \text{ boeni } e \text{ hun}_{\text{INT}} \rightarrow \text{ poeni } e \text{ hun}_{\text{REFL}}$

The reanalysis of Middle Welsh *X hun* seems thus natural enough, especially since Middle Welsh *X hun* already occurs as a head in a pronoun-like function coding the subject or after a preposition. It is however more difficult to explain why the preverbal pronouns and *ym-* prefixes were lost.

One context in which this could have happened, are the imperative constructions discussed in 4.3.3, where infix pronouns preceding the verb are not possible. The hypothetical phrase in (86) contains an imperative verb followed by an emphasising subject pronoun and an object pronoun + intensifier. The sequence of two second person singular pronouns with different functions is not attested and seems to be ungrammatical like English *himself himself*, i.e. the sequence of reflexive and actor-oriented intensifier in (87) from Gast and Siemund (2006: 360). The Middle Welsh object pronoun was dropped then, leading to expressions that are actually attested, both with and without an emphasising subject pronoun (88, 89).

- (86) **ardymhera di ti du hvn*
 moderate_{2SG.IMPV} 2SG_{SUBJ} 2SG_{OBJ} 2SG_{INTS}
 ‘Moderate (*you) (*yourself_{REFL}) yourself_{INTS} / Mäßige du dich selbst!’

(87) *He killed* (**himself*_{REFL}) *himself*_{INTS}
 Er tötete sich selbst

(88) *ardymhera di du hvn*
 moderate_{2SG.IMPV} 2SG_{SUBJ} 2SG_{REFL}
 ‘Moderate (*you) yourself / Mäßige du dich!’

(89) *kymedrola dy hvn*
 moderate_{2SG.IMPV} 2SG_{REFL}
 ‘Moderate yourself / Mäßige dich!’

While this could be a starting point for the reanalysis of *X hun*, one wonders whether these pragmatically marked reflexive imperative clauses were frequent enough to trigger the change of system. This objection is reinforced by the fact that all instances contained in *Rhyddiaith Gymraeg 1300–1425* cluster in two single texts.

The small sample of 14 instances of reflexive *X hun* in the Middle Welsh corpora is thus not enough to formulate a strong hypothesis concerning the trigger of the change. More insights can probably be gained from the analysis of Early Modern texts, where the *X hun* reflexives become more frequent.

5 Conclusions

The quantitative study based on *Rhyddiaith Gymraeg 1300–1425* showed that the alleged Middle Welsh ‘reflexive pronoun’ *X hun* functions, in fact, as an intensifier in 99.32 % of cases. Only as few as 14 instances from both corpora showed its use as a reflexive, which then became widespread in the modern language. In Middle Welsh, ‘full reflexive’ events are coded by the prefix *ym-* or by an infixed pronoun. Since both strategies are ambiguous, intensifiers are added for referent disambiguation and role disambiguation. The two strategies are about equally frequent and, to some degree, interchangeable. In some cases, a strategy may be blocked because of lexical or syntactic constraints.

Although Evans (1964: 89) was aware that the reflexive use of *X hun* was only at its initial stages in Middle Welsh, he probably would be surprised to find out that the entire Middle Welsh corpus does not provide many more instances than the three examples that he cited in his *Grammar of Middle Welsh*.

At present, it is impossible to determine what brought about such change in the Middle Welsh system, however, linguistic convergence with Latin or Old French can certainly be excluded in light of the instances discussed above.

Finally, hypotheses on linguistic convergence between Welsh and English regarding the expression of reflexivity will have to take into consideration the scarcity of reflexive *X hun* in the Welsh corpus before 1425.

Acknowledgement: This study was carried out within the project *Detransitivity in the Brittonic languages: Reflexivity, reciprocity and middle voice constructions* funded by the German Research Council.

Joseph F. Eska and Benjamin Bruch

11 Prolegomena to the diachrony of Cornish syntax

1 Prelude

Little scholarship has been conducted upon Cornish syntax, including the configuration of the affirmative root clause.¹ In this preliminary study, we examine the configuration of this clause-type in Cornish from its earliest records through Late Cornish and attempt to establish what can be said about it, or, at times, what possibilities should be considered when nothing definitive may be said.

2 Old Cornish

We follow Schrijver (2011: 2–5) in the view that the earliest neo-Brittonic did not diverge into discrete languages until the eighth century and that what have been termed Old Cornish and Old Breton did not diverge from each other prior to the eleventh century. Until that time, they formed a unitary Old Southwest Brittonic. Under such an analysis, there are, in fact, no attested Old Cornish verbal sequences.

2.1 VSO in Old neo-Brittonic

On the basis of comparison with Old Welsh and Old Southwest Brittonic, we presume that Old Cornish was VSO on the way to becoming V2. Cf. the following VSO clauses:

- (1) a. Old Welsh
rodesit elcu guetig equs . . .
give_{3SG.PRET} Elcu then horse
'Elcu then gave a horse . . . ' (Jenkins and Owen 1984, cited after Watkins 1987 [The "Surexit" Memorandum])

¹ There is no controversy over the fact that negative root clauses and all embedded clauses, through all periods of the attestation of Cornish, were V1. A constituent is permitted to appear before the negator in negative clauses, but they look like other V1 clauses in that the verb is not third person singular but is conjugated.

b. Old Southwest Brittonic

dadarued epac(dou) . . . int rid ou mod

occur_{3SG.PRES} epacts PTCL free 3PL_{POSS} manner

‘Adviennent les épactes sans obstacle [The epacts occur without hindrance].’ (Fleuriot 1964b: 412 [Bibliothèque Municipale d’Angers, MS. 477])

On the basis of compound or negated verbal forms such as Old Welsh *imm-it cel* (Lambert 2003: 113, gloss 85; Schrijver 2011: 49) ‘it conceals itself’ and Old Southwest Brittonic *ni-s-guilom* (Fleuriot 1964b: 262 [Bibliothèque Municipale d’Angers, MS. 477]) ‘nous ne la voyions pas [we would not see it]’, which contain object agreement affixes that continue pronominal morphemes – which are known to occur in the lower left periphery of the clause – we understand the verb in these VSO clauses to occupy Fin and to move there synchronically because Fin bears uninterpretable φ -features (viz. person, number), i.e. [u φ], that trigger movement through T into the left periphery.²

2.2 V2 in Old neo-Brittonic

There is also evidence for V2 root clauses in Old neo-Brittonic. Borsley, Tallerman, and Willis (2007: 290) cite one token from Old Welsh and Fleuriot (1964b: 413) several from Old Southwest Brittonic:³

(2) a. Old Welsh

[*Gur dicones remedaut elbid*] *a ’n guorit*

man make_{3SG.PRET} wonder world PTCL 1PL_{INFX} redeem_{3SG.PRET}

‘The man who created the wonder of the world redeems us.’ (Williams 1980: verse 5^{a-b} [Juvencus Englynion])

b. Old Southwest Brittonic

[*do(u) cuntraid*]. . . *a int im pop un mis*

two neap.tide PTCL be_{3PL.PRES} in each one month

‘Deux marées de morte-eau . . . sont dans chacun mois [There are two neap-tides every month].’ (Fleuriot 1964b: 413 [Bibliothèque Municipale d’Angers, MS. 477])

² An uninterpretable feature must be checked for a clause to be interpretable. In this instance, the φ -features on the verb check the uninterpretable φ -features in Fin by movement of the verb into Fin. For further discussion, see Svenonius (2007).

³ See Fleuriot (1964a: 151) for the full text of the token cited. All glossed examples cited after other authors adopt their respective glossing.

It appears likely that these V2 clauses developed out of cleft-type constructions (Manning 2001; Willis 2009: 146–147). From the synchronic point of view, such clauses, like V1 clauses, bear [u ϕ] features on Fin that draw the verb into the left periphery, but, unlike V1 clauses, also an Edge Feature that causes an XP to move into SpecFinP, which then may move higher into the left periphery.

3 Middle Cornish

There is considerable diversity of opinion about the configuration of affirmative root clauses in Middle Cornish as reflected in the title of Mark Kille’s Harvard University B.A. Honours thesis, “*What thing is next I don’t quite know*”: *An analysis of variation in word order and subject-verb agreement in Middle Cornish* (1995).⁴ Lewis (1946: 47), followed by Kille (1995: 5), maintains that both SVO and V1 are unmarked configurations in Middle Cornish. Though he notes that a variety of constituents can precede the verb – and also observes that V1 occurs, and, importantly, that some forms of *bos* ‘be’ require a V1 configuration – Williams (2011: 336) suggests “that Middle Cornish is in essence an SVO language”. George (1991) compiles all of the surface configurations attested in the play *Beunans Meriasek* (BMer., composed ca 1500, edited in Stokes 1872), but is satisfied to conclude by listing only the most common ones.

3.1 Surface V2 in Middle Welsh and Middle Breton

It is clear that affirmative root clauses in Middle Welsh and Middle Breton are V2, i.e. the preverbal XP is not restricted to the Subject, but may be also an Object or Adverb(ial). The post-verbal position of the Subject in (3b) and (4c) is diagnostic of the V2 character of these clauses. The following tokens are cited after Borsley, Tallerman, and Willis (2007: 287–290):⁵

⁴ The translated quotation is from BMer. (line 107): *pandryv nessa ny won fest*.

⁵ For unequivocal demonstrations that Middle Welsh affirmative root clauses bear V2 configuration, see Willis (1998) and Meelen (2016). Middle Breton has not been the focus of similar studies, but see Schafer (1995) and Borsley and Kathol (2000) for Modern Breton.

(3) Middle Welsh

a. Subject-initial

[*Riuedi Mawr o sswydwyr*] *a gyuodassant y uynydd . . .*
 numbers large of officials PTCL rise_{3PL.PRET} up
 ‘Large numbers of officials got up . . .’ (PKM 16.18–19)

b. Object-initial

Ac [ystryw] a wnaeth y Gwydyl
 and trick PTCL make_{3SG.PRET} DEF Irish
 ‘And the Irish played a trick.’ (PKM 44.11)

c. Adverb(ial)-initial

[*Yn Harlech*] *y bydwch seith mlyned ar ginyaw . . .*
 in Harlech PTCL be_{2PL.FUT} seven years at dinner
 ‘In Harlech you will be at dinner for seven years . . .’ (PKM 45.2–3)

(4) Middle Breton

a. Subject-initial

[*Cesar*] *a respontas deze . . .*
 Caesar PTCL reply_{3SG.PRET} to_{3PL}
 ‘Caesar replied to them. . .’ (Ernault 1887a: 82 § 12 [La vie de sainte Catherine])

b. Object-initial

hac [an holl doueouse] . . . a meur an oll
 and DEF all gods=PROX PTCL have_{1SG.PRES} DEF all
dispriset. . .

renounce_{PST-PTCPL}

‘. . .and I have renounced all those gods. . .’ (Ernault 1887a: 80 § 8)

c. Adverb(ial)P-initial

hac [en continant] ez aparissas an eal dezy
 and immediately PTCL appear_{3SG.PRET} DEF angel to_{3SG.FEM}
 ‘. . .and immediately the angel appeared to her.’ (Ernault 1887a: 84 § 13)

3.2 Surface V2 in Middle Cornish

Recent theoretically oriented scholarship, e.g. Borsley, Tallerman, and Willis (2007: 291), notes that identical structures are found in Middle Cornish, some tokens of which follow:

(5) a. Subject-initial

[ny] a 'th wor the pen an gluas
 1PL PTCL 2SG_{INFX} put_{3SG.PRES} to Land's End
 'We will bring you to Land's End.' (BMer. 1. 594)⁶

b. Object-initial

[guyr] a gousaf vy
 truth PTCL speak_{1SG.PRES} 1SG
 'I speak truth.' (Norris 1859, 2: 1. 909)

c. Adverb(ial)-initial

[ragon] y pesys y das
 for_{1PL} PTCL beseech_{3SG.PRET} 3SG_{MASC.POSS} father
 'For us he beseeched his father.' (Stokes 1860–1861: 6, stanza 9^a, [The Passion])

3.3 Surface V3* in Middle neo-Brittonic⁷

All of the Middle neo-Brittonic languages allow more than a single constituent to occur before the verb, though, by and large, only one of these constituents may be an argument,⁸ e.g.:

(6) a. Middle Welsh

ac [ar hynny] [at Uath uab Mathonwy] yd aethant wy
 and on DIST to Math uab Mathonwy PTCL go_{3PL.PRET} 3PL
 'And thereupon they went to Math uab Mathonwy.' (PKM 68.15–16)

⁶ Though we cite the standard edition and translation by Stokes (1872), we occasionally silently adopt improvements in the text and translation by Ray Edwards in Syed and Edwards (1996).

⁷ The asterisk indicates that the verb appears in third or later position in the clause.

⁸ The reason for this constraint requires further research. Under the view that V2 structures in the Brittonic languages developed out of cleft-type structures (see section 2.2), a preliminary hypothesis may be that since only a single argument can appear before the relative pronoun in a cleft structure, as affirmative root clauses were interpreted as V2, there would not have been any evidence for a language learner that more than a single argument could appear in the left periphery.

b. Middle Breton

[*breman*] [a crenn] [me] a gourchemen dit
 now PTCL plainly 1SG PTCL ask_{3SG.PRES} t_{O2SG}
 ‘Now, I plainly ask of you.’ (Ernault 1887b: 256 l. 240 [Vie de sainte
 Nonn])

c. Middle Cornish

ha [my] [lemmen] a 'th vygeth
 and 1SG now PTCL 2SG_{INFX} baptise_{3SG.PRES}
 ‘And now I will baptise you.’ (BMer. l. 941)

These structures do not undermine the V2 analysis, for the important characteristics of the configuration are that the [uφ] features on Fin draw the verb into the lower left periphery and that its Edge Feature causes an XP to move into SpecFinP.

3.4 The architecture of the left periphery

Since Rizzi (1997), it has become clear that the left periphery of the root clause is highly articulated (see further Poletto 2002; Benincà and Poletto 2004; Rizzi 2004, 2013; Haegeman 2012, *inter alios*). Under this analysis, the *communis opinio* understands the hierarchical architecture of the left periphery to be:

(7) [_{FrameP} [_{ForceP} [_{TopP} [_{FocP} [_{FinP} . . .]]]]]

Within this framework, FrameP is the locus of scene-setting locative and temporal adverb(ial)s and hanging topics, ForceP is the locus of markers of illocutionary force and clause-typing, TopP is the locus of topical XPs, FocP is the locus of focussed XPs, and FinP expresses the finiteness or non-finiteness of the clause.

However, in a more refined analysis, Frascarelli and Hinterhölzl (2007: esp. 88) propose that there are, in fact, three different types of topics, each of which is projected separately within the left periphery. In this regard, Hinterhölzl and Petrova (2010: 320–321) write:

- (a) ABOUTNESS TOPIC: “what the sentence is about” (Reinhart, 1981; Lambrecht, 1994), “what is a matter of standing and current interest or concern” (Strawson, 1964);
 (b) CONTRASTIVE TOPIC: an element that induces alternatives which have no impact on the focus value and creates oppositional pairs with respect to other topics (Kuno, 1976; Büring, 1999);

(c) FAMILIAR TOPIC: a given, D-linked constituent, which is typically destressed and realised in a pronominal form (Pesetsky, 1987)⁹

The hierarchical architecture of the left periphery in (7) is, thus, expanded to:

- (8) [FrameP [ForceP [AbTopP [ContrTopP [FocP [FamTopP [FinP . . .]]]]]]]]

Poletto (2002) was the first to propose that the locus of the V2 phenomenon can be either FinP + ForceP or FinP alone. In the former, the verb and initial XP move through FinP to ForceP, as in (9), thus severely restricting the number of constituents that can appear before the verb.

- (9) [FrameP [ForceP **XP** [Force **V**] [AbTopP [ContrTopP [FocP [FamTopP [FinP ~~XP~~ [Fin **V**] . . .]]]]]]]]

In the latter, however, the verb and XP remain in FinP, as in (10), and FrameP, ForceP, AbTopP, ContrTopP, FocP, and FamTopP may all host constituents that precede the verb.

- (10) [FrameP [ForceP [AbTopP [ContrTopP [FocP [FamTopP [FinP **XP** [Fin **V**] . . .]]]]]]]]

We may look to medieval Romance for an illustration. Wolfe (2016, 2018) discusses this microvariation and demonstrates that later Old French is a Force V2 language and restricts the number of constituents that can precede the verb. In (11), a frame-setting clause appears in SpecFrameP (adapted from Wolfe 2018: 69):

- (11) *Et quant il est apareilliez, il prent ses*
 and when he be_{3SG.PRES} appear_{PTCPL} he take_{3SG.PRES} his
armes et monte . . .]
 weapons and ride_{3SG.PRES}
 ‘When he appeared, he took his weapons and rode . . .’ (Pauphilet 1923:
 l. 129 [*La queste del Saint Graal*])

In Wolfe’s sample, there are but two tokens of V4 out of 632 clauses (0.32 %). On the other hand, in Wolfe’s sample of 622 clauses in Old Occitan, not only does V3 occur more often than in later Old French,¹⁰ but V4 occurs in 8.04 % of clauses, as well as V5 in 1.29 % and V6 in 0.64 %, none of which appear in Old

⁹ With regard to the givenness/accessibility characteristic of *familiar topics*, cf. also Chafe (1987).

¹⁰ Old Occitan 29.74%, Old French 24.53%.

French. Thus, V5 clauses, such as (12), are possible in Old Occitan, but not in later Old French (cited after Wolfe 2018: 68):

- (12) *E* [*per aisso*], [*illi*] [*adones*], [*am gran confusion*] *comandet*
 and for this she therefore at great confusion command_{3SG.PRET}
a totas
 to all
 ‘Because of this, amongst great confusion, she commanded everyone to . . .’ (Albanès 1879: 96 § 41 [*La vie de sainte Douceline*])

It is this ‘relaxed’ instantiation of the V2 phenomenon that we find in Middle neo-Brittonic languages. A definitive V6 token is cited by Borsley, Tallerman, and Willis (2007: 293) after Poppe (1991: 178), with an analysis in the present framework indicated:

- (13) *Ac* [_{FrameP} *o ’r dywed*] [_{ForceP} *gan wuyhaf grym a llafvr*]
 and of DEF end with greatest power and toil
 [_{AbTopP} *gwedy kaffael o ’r Brytanyeyt penn e mynyd*],
 after get_{VN} of DEF Britons top DEF mountain
 [_{FocP} *en e lle*] [_{FamTopP} *wynt*] *a lle*] *dangossasant . . .*
 in DEF place 3PL PTCL place show_{3PL.PRET}
 ‘And in the end with the greatest power and toil once the Britons had gained the top of the mountain in that place they showed . . .’ (Roberts 1971: lines 795–797 [*Brut y brenhinedd*])

3.5 Preverbal Object DP + pronominal Subject

George (1991: 216) calls attention to tokens of an Object DP + pronominal Subject + affirmative particle *a* + verb construction and labels it “a valid one” – which we interpret to mean that he believes it to be generated by the grammar – upon the basis of the fact that there are 29 tokens of it in *Beunans Meriasek*, five of which, he states, “are [not] dependent on the rhyme”. He regards it as a Cornish innovation (George 1990: 229–230, 239–240). Such constructions are found not only in *Beunans Meriasek*, but in earlier texts, as well, e.g.:

- (14) a. *ag* [ol ʒe vɔʒ] [hy] a wra
 and all 2SG_{POSS} will 3SG_{FEM} PTCL do_{3SG.PRES}
 ‘And she will do all of your will.’ (Toorians 1991: 1. 20 [*The Middle Cornish Charter*])
- b. *ha* [henna] sur [my] a greys
 and DIST surely 1SG PTCL believe_{3SG.PRES}
 ‘... and I surely believe that.’ (Norris 1859, 1: l. 1263 [*Origo Mundi*])

We understand the pronominal Subjects in the Middle Breton clause in (6b) and the Middle Welsh clause in (13) clearly to be hosted in SpecFamTopP. For a clear Middle Cornish example, consider the following passage, in which Teudar addresses St Kea:

- (15) *Mars o Christ Du mar rajak,* ‘If Christ was God so gracious,
pew o e das? who was his father?
Te pen boba lagajak, You goggle-eyed head of a clown,
ro gorthyb vas. give a satisfactory answer.
Bith war! Na fal! Watch out! Don’t fail!’
 (Thomas and Williams 2007: lines 208–212
 [*Bewnans Ke*])

Second person singular deixis having been established, the succeeding line is:

- (16) [*Anotho*] [*te*] *re gowsys*
 of_{3SG.MASC} 2SG PERF speak_{3SG.PRET}
 ‘You have spoken of him.’ (Thomas and Williams 2007: l. 213 [*Bewnans Ke*])

There can be little doubt but that the pronominal Subject is hosted by SpecFamTopP.¹¹

3.6 More on preverbal Object DP + pronominal Subject

Though there clearly are available Specifier positions to host an Object DP within the left periphery while a pronominal Subject is hosted by SpecFamTopP, we are not certain that such is the correct analysis, for, as mentioned in section 3.3, a

¹¹ For a discussion of this construction in Old English and Old High German, see Walkden (2015).

constraint exists against more than one argument appearing in the left periphery. It may be that all tokens of this construction are not generated by the grammar, but represent an over-determination of it by the necessity to enable a rhyme (here marked by underlining: single and double, dashed and wavy).¹² Consider the surrounding context of the two tokens of this construction cited in (14):

- (17) a. *lemen y3 torn my as re*
ha war en grey3 my an te
nag vs y far
an bar3ma 3e Pons tamar
{ad} my ad pes warty by3 da
ag ol 3e vo3 hy a wra
rag flog yw ha gensy so3,
ha gassy 3e gafus y bo3.
- ‘Now I give her into your hand,
and upon the . . ., I swear it,
there is not her equal
from here to the Tamar bridge.
I beg you, be good to her,
and she will do all your will,
for she is a child, and . . .
And allow her to have her will.’
(Toorians 1991: lines 15–22 [*The Middle Cornish Charter*])¹³

¹² We recall Jakobson’s (1923: 16 = 1979: 15) “teoriju organizovannogo nasilija poëtičeskoj formy nad jazykom [theory of organised violence of poetic form over (natural) language]”. For an example, consider the twelfth stanza of the early Irish poem *Fo réir Choluimb céin ad·fias* [As long as I speak, (may I be) obedient to Columb]:

- Do-ell Éirinn, indel cor,*
cechaing noib nemed mbled,
brisis tola, tindis for,
fairge al druim dánae fer.
- ‘He turned away from Ireland, having made covenants (?),
he traversed in ships the whales’ sanctuary,
he broke desires, he was illuminated (?),
A brave man over the ridge of the sea.’ (Kelly 1973:
§ 12)

In this poem of 4 | 3 heptasyllabic lines, one finds linking alliteration between the last word of a line and the first word of the succeeding line, line-internal alliteration, and rhyme. As observed by Watkins (1995: 121), the words in the final line of this stanza occur virtually in the reverse of unmarked order, viz. *Fer dánae al druim fairge*, in order to enable the 4 | 3 scansion and retain the alliterative patterns:

- | | | | | | |
|------------------------------------|-----------|----------------------|--|------------------------------|--------------------|
| <i>fairge</i> | <i>al</i> | <i>druim</i> | | <i>dánae</i> | <i>fer</i> |
| sea _{GEN} | over | ridge _{DAT} | | brave _{NOM.SG.MASC} | man _{NOM} |
| ‘a brave man over the sea’s ridge’ | | | | | |

¹³ The text is difficult in places. We provide the reference to the most recent published edition and translation, but provide the reading of Bruch (2005: 335) and his unpublished translation when either are uncertain.

- b. *me a vyn mos the'n temple*
ha dev ena a worthye
kepar del goth thy'mmo vy
ef yv arluth nef ha'n beys
ha henna sur my a greys
a luen colon pur theffry
- 'I will go to the temple,
 And will worship God there,
 As it is incumbent on me
 He is Lord of heaven and earth,
 And that I surely believe
 With full heart, very earnestly.'
 (Norris 1859, 1: lines 1259–1264)

It is clear from these passages that the preverbal Object DP + pronominal Subject construction exists to enable the rhyme.

3.6.1 Apparent exceptions

There are apparent exceptions to the hypothesis that preverbal Object DP + pronominal Subject constructions are always poetic overdeterminations, but we are uncertain whether they ought to be considered as authentic instances of Middle Cornish clausal configuration.¹⁴ We provide an example. Consider the following stanza:

- (18) *Arluth henna ny a ra*
desempys duen alema
aspyans pup ay quartron
me agis gyd rum ena
pur uskis bys in cambron
- 'Lord, that we will do.
 Straightway let us go hence.
 Let every one spy from his corner.
 I will guide you, by my soul,
 Very quickly, as far as Camborne.'
 (BMer. lines 978–982)

The first verbal clause of this stanza is glossed as:

- (19) [*henna*] [*ny*] *a ra*
 DIST 1PL PTCL do_{3SG.PRES}
 'We will do that.' (BMer. l. 978)

Clearly the configuration of this line, as can be seen in (18), enables the rhyme, but one may note that the configuration **ny a ra henna* would also enable the rhyme, while yielding a configuration with only one argument before the verb,

¹⁴ See the appendix for a discussion of the five tokens that George states are not the result of poetic overdetermination.

hence the position of *henna* must be serving a pragmatic purpose and is not merely the result of poetic overdetermination. Indeed, this is correct, but there is more to be said. Consider the immediately preceding stanza addressed by Teudar to the torturers:

- | | |
|--|--|
| (20) <i>Meryasek ythyv gelwys</i>
<i>in crist yma ov cresy</i>
<i>genogh why bethens sesijs</i>
<i>gruegh y tormontya besy</i>
<i>crist mar ny veth denehys</i>
<i>pegh then horsen trewesy</i>
<i>genogh kynfove lethys</i>
<i>me agis menten defry.</i> | ‘Meriasek is he called:
In Christ he believes.
By you let him be seized
Do ye torment him.
If Christ be not denied,
A thrust to the doleful whoreson!
Though he be slain by you,
I will maintain you certainly.’
(BMer. lines 970–977) |
|--|--|

In context, then, it is clear that *henna* in (19) is a response to the directions in the passage in (20), and, thus, that it, in fact, has been moved into the left periphery to occupy SpecAbTopP. But we must also look at the position of the pronominal Subject in this line and, indeed, determine why it is present at all. Were *ny* not in preverbal position, we might expect the clause to appear as **Arluth, henna a ren* with a conjugated verb, but not only would such a line not rhyme, it would also be a syllable short. In *Beunans Meriasek*, a pronominal Subject, though not required after a conjugated verb, could be included in order to make rhyme or syllable-count, but **Arluth, henna a ren ny* would not provide the necessary rhyme in this token. In this clause, it is not that the Object DP has been displaced by poetic overdetermination, but that the grammatically unnecessary pronominal Subject has been inserted in order to enable the correct syllable-count.¹⁵

3.7 Preverbal Subject DP + Object DP constructions

In this section, we merely observe that, just as the preverbal Object DP + pronominal Subject construction exists to enable the rhyme and/or syllable-count, constructions with both Subject DP and Object DP before the verb, in either order, occur for the same reason, e.g.:

¹⁵ Indeed, chevilles are very commonly employed in Middle Cornish verse in order to achieve the necessary number of syllables in a line.

- (21) a. [dew æn] [cris] a æanvonas
 two man Christ PTCL send_{3SG.PRET}
 æe berna boys ha dewas
 to buy_{VN} food and drink
 ‘Christ sent two men to buy food and drink.’ (Stokes 1860–1861: 16, stanza 42^a [*The Passion*])
- b. [mab marya] [mur a beyn]
 son Mary great of pain
 a woævy y ’n vr na
 PTCL endure_{3SG.IMPF} in DEF hour DIST
 ‘The son of Mary endured great pain in that hour.’ (Stokes 1860–1861: 18, stanza 54^a [*The Passion*])

A good illustration of the extent to which displacement can occur in order to enable necessary rhyme and syllable-count is the following, the first four lines of an ABABABAB stanza:

- (22) [pur wyr] [certan] [an den ma]
 very true certain DEF man PROX
 [lyes den] re wruk treyle
 many man PERF do_{3SG.PRET} turn_{VN}
 [agan laha] [ef] yma¹⁶
 1PL_{POSS} law 3SG_{MASC} be_{3SG.PRES}
 [pup vr] ow contradye
 every hour PROG contradict_{VN}
 ‘Truly, this man certainly has converted many men. He is always opposing our law.’ (Norris 1859, 1: lines 2423–2426 [*Passio Christi*])

3.8 Some comments on variation in Middle Cornish texts

The number of clauses that deviate from affirmative root clause V2 in the Middle Cornish corpus varies amongst texts. Deviation is fairly common in the poem

¹⁶ Note that *yma* is one of the few verbs that usually requires V1 configuration even in affirmative root clauses. The occurrence of an adverb(ial) or participle to the left of *yma* is not unusual, but a DP in that position is very unusual. Note, furthermore, that, in Breton, *emañ* is employed only after an adverb(ial) or participle, while Subject + V + Complement requires *a zo*; this construction does not have an equivalent in Cornish, however.

Pascon agan Arluth (ca. 1400), but considerably less frequent in the two saints' plays, *Beunans Meriasek* and *Bewnans Ke*, which were probably written around a century later. The *Ordinalia* cycle of mystery plays, which are roughly contemporaneous in date of composition with *Pascon agan Arluth*, do not exhibit nearly as many tokens of such divergent affirmative root clauses as the poem. This is particularly telling when we consider that the central play of the *Ordinalia* cycle, *Passio Christi*, contains some lines which are also found in *Pascon agan Arluth*, although scholarly opinion is divided as to whether the play or the poem is the older text – or original source of the lines in question.¹⁷ It is worth considering, however, that the configurational differences amongst the texts may relate to the fact that *Pascon agan Arluth* is a poem, while the *Ordinalia*, *Beunans Meriasek*, and *Bewnans Ke* are plays.

The medieval Cornish dramas seem clearly to have been written as texts to be performed aloud by actors. Evidence from historical records and in the manuscripts of the plays themselves suggests that they were actually staged (Joyce and Newlyn 1999: 541–558; Bakere 2009: 213). But, as a poem, *Pascon agan Arluth* may have been written as a more purely “literary” text, to be read privately or even silently, rather than recited or performed for an audience. It is, therefore, possible that the language and number of configurational liberties found in it represent a more “literary” register of Cornish, while the plays stay closer to the norms of the spoken language, though displaying divergent clausal configuration when demanded by rhyme and syllable-count.

Another factor is the type of verse form that *Pascon agan Arluth* employs. Whereas the other works of Middle Cornish verse are written in a wide variety of stanza forms, most of which require only three or four pairs of rhyming lines – such as AABCCB or ABABCDDC, in which there are two A lines, two B lines, and two C (and D) lines – *Pascon agan Arluth* is written almost exclusively in eight-line stanzas rhymed ABABABAB, in which the poet is obliged to supply two sets of four rhyming words, i.e. four A rhymes and four B rhymes.¹⁸ It may be that these much stricter rhyming requirements are what spurred the author of *Pascon agan Arluth* – who, unlike the Middle Welsh or Irish bards, was probably not a professional or

¹⁷ Nance (1949: 368), Murdoch (1981: 823–826), Williams (2006: 66), and George (2010: 493) believe that *Pascon agan Arluth* is the older text. Fowler (1961: 104–111) takes the opposite view. Bruch's current opinion tends to side with Fowler upon the basis of metrical evidence.

¹⁸ Bruch (2009: 90–91) remarks upon the difficulty of finding the necessary rhyming words for an ABABABAB stanza in Middle Cornish, and suggests that this may be a reason why this rhyming pattern is employed less over time, occurring less commonly in *Beunans Meriasek* than in the *Ordinalia*, and not being present at all in the surviving text of *Bewnans Ke*.

highly trained poet – to scramble the constituent order of his affirmative root clauses quite freely simply to get an appropriate syllable at the end of each line. It seems as though some of the most divergent affirmative root clause configurations in the *Ordinalia* also occur in stanzas with ABABABAB rhyming stanzas such as in (22), suggesting that it was the need to provide a particular end rhyme, more than any other consideration, that prompted poets to depart so widely from the V2 norm that likely was typical of spoken Middle Cornish.

4 Late Cornish

General opinion is that affirmative root clauses in Late Cornish, presumably under the influence of English, had become SVO (e.g. Jenner 1904: 158; Lyon and Pengilly 1987; Wmffre 1998: 62–63; Brown 2001: 248–249; Gendall 2004: 98–100, 140; Williams 2011: 336). This may be, but the textual sources are often not of high trustworthiness.

4.1 The *Tregear homilies*: A transitional text

Williams (2011: 336) states that the *Tregear homilies*, dated to ca. 1558, are characterised by some Middle Cornish features and some Late Cornish features.¹⁹ He states that they are written “in fairly colloquial prose” (Williams 2011: 336); indeed, in his judgement, Tregear’s “morphology and syntax are perfect” (Williams 2011: 338). However, a close examination of the text, a translation of Edmund Bonner’s *A profitable and necessary doctrine, with certayne homelyes adioyned therevnto* (1555), indicates that Tregear’s translation closely follows the clausal configuration of his English exemplar, violating a variety of features of Cornish syntax. We note, furthermore, that Tregear deletes, enlarges, or paraphrases portions of Bonner’s text. See the appendix for an illustration of his translation practice, several points concerning which we discuss in the following subsection.

¹⁹ For example, Tregear’s language generally evinces the Late Cornish unrounding of /œ/ so as to merge with /ɛ/ and shows frequent use of periphrastic constructions in preference to conjugated verbs, but does not provide any trace of the Late Cornish pre-occlusion of nasals and unrounding of /y/ to /i/.

4.1.1 Comments upon John Tregear's translation practice

The *Tregear homilies* appear likely to be a poor, overly hasty, and unpolished translation of Bonner, in which Tregear, though a native speaker of Cornish, tends to preserve the configuration of constituents in his English exemplar whenever possible (perhaps because he was translating one phrase at a time). It is clear that he was sufficiently influenced by English lexis and syntax to the extent that he introduces English words and configurational syntax, e.g. Adj + N relative ordering, not only in untranslated borrowings such as *sufficiant cawse* 'sufficient cause' in folio 1^r, l. 5, but also in his own additions to the text, such as *perfect colonow* 'perfect hearts' in folio 1^r, l. 3, in which the noun is Cornish. This may be evidence that, for educated bilinguals such as Tregear, there was a kind of "priests' Cornish" equivalent to the Breton *brezhoneg beleg* 'priests' Breton', which incorporated elements of French lexis and syntax (see Williams 2006: 189).

Though Tregear gives the impression that he wishes to stay faithful to English configurational syntax wherever possible, there are instances in which his knowledge of Cornish sometimes encourages him to make a different choice. Examples of this include tokens in which he substitutes Cornish V1 for English SVO in negative root clauses and embedded clauses, e.g.:

- (23) a. *ny rug eff leverall in pegh, mas i*
 NEG do_{3SG.PRET} 3SG_{MASC} say_{VN} in sin but in
'n plural number, in pehosow.
 DEF plural number in sins
 'He saieth not sinne, but in the plural number, sinnes.' (folio 8^v, l.14)
- b. *lymmyn pan rug du . . . creatya ha gull den*
 now when do_{3SG.PRET} god create_{VN} and make_{VN} man
 'Now when god had . . . cre[a]ted man . . .' (folio 2^r, lines 18–19)

There are also clausal tokens in which he breaks up long English sentences or reorders the constituents of those sentences so as to produce coherent Cornish sentences that are closer to the syntax of the traditional texts (including, perhaps, a preference for V2 configuration in affirmative root clauses). Tregear is not reluctant to place two DPs before the verb in a root clause, at least in instances in which his English exemplar does, so '[onely one kynde of fruyte] [he] charged hym . . .' becomes [*Saw vn kynda a frut*] [*an tas dew*] *a chargias mabden . . .* 'only one kind of fruit, God the Father charged mankind' in folio 4^r, lines 2–3. Since two DPs can occur before the verb in Middle Cornish verse texts, perhaps such phrasing sounded acceptable to Tregear as a Cornish

speaker. On the other hand, Tregear renders the following clause from folio 4^r, l. 3, in which he wishes to construct an embedded negative clause meaning ‘that he might not meddle with or touch it [i.e. the fruit],’ as *na rella myllya na tuchia worta* with V1 (Aux [S] V PP) configuration, not S Aux V PP as it would be in English.²⁰

Tregear is perhaps most useful as a source of negative evidence concerning the syntax of Cornish, since there appear to be clauses in which he chooses (or is forced) to avoid duplicating the English clausal configuration, even though the latter seems to be his default approach to translating Bonner’s prose. Like Breton clergy, he seems quite comfortable with producing overdetermined DPs such as *an frut an wethan* for Bonner’s ‘the fruyte of the tree,’ for which **frut an wethan* is expected. But when faced with ‘The Prophette, Dauid . . . alledgeth,’ Tregear produces progressive *yma an profet dauid ow allegia* ‘The Prophet David is alleging,’ since this is how he usually expresses the simple present tense in Cornish, employing *yma* + verbal noun and the V1 configuration normal for this verb (Williams 2016: 120–124) or the progressive particle *ow* + verbal noun.²¹

In our view, the *Tregear homilies* cannot be employed to establish much about Middle or Late Cornish affirmative root clause configuration.

4.2 The writings of the Boson family

Three members of the Boson family left Late Cornish texts from ca 1660 to ca 1730. They were not native speakers of the language, however, so one must be suspicious of English influence in the texts that they produced. In the folktale *Dzhûan Tshei an Her* ‘John of Chyannor’, said to have been written by Nicholas Boson,²² one does not find diagnostic V2 structures such as Object DP/Adverb-(ial)P + affirmative particle + verb + Subject DP, but there are a number of tokens in which an Object DP and a pronominal Subject precede the affirmative particle *a* and verb, e.g.:²³

²⁰ Of course, the English clause that he is translating, *vterly to refrayne from [eating the fruit]*, is not itself negative or V1 or even an embedded clause, as it lacks a finite verb (we note that Cornish has no equivalent to such a negated infinitive).

²¹ Unlike Breton, in which it is permissible to formulate an SVO sentence of the type *D. a zo* + verbal noun, Cornish does not have a means of constructing an SVO root clause employing Subject DP + *a yw* + verbal noun; cf. n. 16.

²² Printed by Lhuyd (1707: 251–253) in an idiosyncratic orthography.

²³ We preserve Lhuyd’s orthography in these tokens.

- (24) a. [*Kibmiaz têg*] [*ev*] *a kymeraz . . .*
 leave fair 3SG PTCL take_{3SG.PRET}
 ‘He took fair leave . . .’ (Padel 1975: 15 § 3)
- b. *Ha [an mona] [an dzhei] a gavaz; ha [’n*
 and DEF money 3PL PTCL find_{3SG.PRET} and DEF
bara] [dzhei] a dhabraz
 bread 3PL PTCL eat_{3SG.PRET}
 ‘And they found the money; and they ate the bread.’ (Padel 1975: 19 § 46)

It seems to us that clauses such as these may well be the result of L1 interference from English, but they do not provide secure evidence that, as in English, the clausal configuration was SVO with the possibility of fronting another constituent for topicalisation, not least because of the presence of the affirmative root clause particle.

4.3 The Bible translations of Wella Rowe

The Bible translations of Wella Rowe,²⁴ which date from ca 1690, are thought to represent some of the latest surviving works written or translated by a native speaker of Cornish. As translations, one must be cautious about his Cornish replicating the configuration of his English exemplar. One passage from Genesis 3:14, however, may suggest that Cornish was moving towards SVO:

- (25) *War tha doer chee ra moaze, ha douste chee ra*
 upon 2SG_{POSS} belly 2SG do_{2SG.PRES} go_{VN} and dust 2SG do_{2SG.PRES}
debre oll deethyow tha vownyas
 eat_{VN} all days 2SG_{POSS} life
 ‘Vpon thy belly shalt thou goe, and dust shalt thou eate, all the dayes of thy life.’ (Cornish: Loth 1902: 180; trans. KJV Genesis 3:14)

This passage contains a token of OSV configuration in the Cornish text, *douste chee ra debre*, which probably is intended to parallel the configuration of the English exemplar with topicalisation, *dust shalt thou eat*. Interestingly, the English text shows two tokens of V2 configuration, *vpon thy belly shalt thou goe* and *dust shalt thou eate*, in which *shalt* is in second position following an initial

²⁴ Comprised of Genesis 3, the ten commandments, and Matthew 2 and 4.

PP or DP, respectively; in both tokens, however, Rowe's Cornish translation has altered the configuration to SVO: *War tha doer chee ra moaze* (PP S Aux V) and *douste chee ra debre* (O S Aux V). This may well comprise the best – though hardly conclusive – evidence we have for Late Cornish moving towards becoming SVO in affirmative root clauses.²⁵

4.4 William Bodinar's letter to Daines Barrington

William Bodinar's letter to Daines Barrington, written 3 July 1776, is often credited as the last text in the traditional Cornish corpus. It is worth noting, however, that Bodinar is not considered to have been a native speaker of Cornish because he, as described in the letter, learnt the language as a boy from older fishermen during expeditions out to sea:²⁶

- (26) *me rig deskey Cornoack termen me vee mawe*
 1SG do_{3SG.PRET} learn_{VN} Cornish time 1SG be_{3SG.PRET} boy
 'I learnt Cornish when I was a boy.' (Pool and Padel 1975–1976: 234.7)

Bodinar employs S Aux V O configuration in *me rig desky Cornoack* 'I did learn (= learnt)²⁷ Cornish', which is consistent with both English SVO and Middle Cornish V2, but note that he employs SVO configuration in the embedded clause *termen me vee mawe* – though this is, perhaps, owing to that fact that he is a native speaker of English, and so may not be diagnostic.²⁸ In his discussion of this clause in Pool and Padel (1975–1976: 236), Padel cites a comparable token of *termen* employed as a complementiser in a letter written by John Boson in 1710 in which the configuration is also SV in the embedded clause – but, of course, Boson also was not a native speaker of Cornish.

²⁵ The anonymous reviewer asks whether we think that in prospective SVO Late Cornish, as in English, the Subject occupies SpecTP and there is no V→T movement. Unfortunately, upon the basis of such slight information, such a determination cannot be made.

²⁶ We note, however, that Pool and Padel (1975–1976: 235) comment that Bodinar's "Cornish is authentic – better than that of John Boson some sixty years earlier."

²⁷ Aux V in place of a conjugated verb is typical of Tregear and Late Cornish.

²⁸ We observe that Bodinar's letter contains 12 lines of Cornish, none of which include the affirmative particle *a*, which could suggest that it had become phonologically null in the variety that Bodinar acquired, leading him to have constructed an SVO grammar. Such a small sample is hardly probative, however.

4.5 Concluding remarks about Late Cornish

The analysis of Late Cornish syntax is inherently problematic. Aside from the *Tregear homilies*, a text in transition from Middle to Late Cornish, which, in our view, shows considerable interference from English, the corpus is very small and composed of translations by a native speaker and texts by non-native speakers. As mentioned in section 4.3, Wella Rowe's use of XP S Aux V in place of the XP V S O attested in two instances in his English exemplar appears to be legitimate evidence in favour of Cornish moving towards SVO in affirmative root clauses, but such crumbs are little upon which to hang a definitive judgement.

5 Future work

The preliminary remarks presented herein have sketched the broad outlines of the diachrony of the configuration of the affirmative root clause in Cornish. Clearly, much more work remains to be done. The next step for us, now underway, is to create parsed corpora for *Passio Christi*, the central play of the *Ordinalia* (ca 1400), and *Beunans Meriasek* (ca 1504) to make hard data available for Middle Cornish. The prospect for progress on Late Cornish – barring the discovery of more texts – appears doubtful. An edition of the *Tregear homilies* is required before its syntactic structures can be systematically investigated.

Appendix I Remarks on George's “valid” tokens of preverbal Object DP + pronominal Subject constructions

George (1991: 216) notes that the preverbal Object DP + pronominal Subject construction in the clauses at BMer. lines 1807–1808, 1888–1889, 3224–3225, 4340–4342, and 4515–5156 are not required to enable the rhyme. The Subject and Verb in each token occur in the first half of a line and, thus, do not participate in the end-rhyme pattern of the stanza. In each token, it is possible to posit a grammatical line that would preserve V2, as well as the syllable-count of the phrase.

– The five tokens

The relevant lines follow:

- (27) a. *ha thyso age hanov*
 and to_{2SG} 3PL_{POSS} name
me a leuer pur ylyn
 1SG PTCL say_{3SG.PRES} very fair
 ‘And to thee their names
 I will tell very fairly.’ (BMer. lines 1807–1808)
- b. *ha the borse mes a 'th ascra*
 and 2SG_{POSS} purse out of 2SG_{POSS} bosom
me a 'm beth ha 'th margh uskis
 1SG PTCL 1SG_{INFX} have_{3SG.PRES} and 2SG_{POSS} horse swift
 ‘And thy purse out of thy bosom
 I will have, and thy swift horse.’ (BMer. line 1888)
- c. *v lon bowyn dufunys*
 five steer beef mince_{PST-PTCPL}
y a depse in ij deth
 3PL PTCL eat_{3SG.COND} in two day
 ‘Five beef steers minced
 They would eat in two days.’ (BMer. lines 3224–3225)
- d. *the volnogeth*
 2SG_{POSS} will
par del deleth
 even as be.fitting_{3SG.PRES}
ny a ra snell
 1PL PTCL do_{3SG.PRES} quickly
 ‘Thy will
 ‘Even as is meet
 We will do swiftly.’ (BMer. lines 4340–4342)
- e. *dadder the lues huny*
 good to many one
eff a ruk ze ihesu gras
 3SG_{MASC} PTCL do_{3SG.PRET} to Jesus thanks
 ‘Goodness to many a one
 He did, to Jesus thanks.’ (BMer. lines 4515–4516)

– **Commentary**

All five tokens can be recomposed as V2 clauses that maintain the correct syllable count with the Object DP in the left periphery by employing a conjugated verb, as in (28a, c–d), and/or placing the pronominal subject in post-verbal position, as in (28b–e):

- (28) a. BMer. line 1808: *me a leuer* → *a lauaraf*
 b. BMer. line 1889: *me a'm beth* → *a'm beth vy*
 c. BMer. line 3225: *y a depse* → *a thepsens y*
 d. BMer. line 4342: *ny a ra* → *a ren ny*
 e. BMer. line 4516: *eff a ruk* → *a ruk eff*

In (28b–e), the V2 alternative requires the use of a postposed pronominal Subject in order to maintain the necessary syllable-count.²⁹ We note that George (1991: 228–229) discusses clauses with an Object DP in the left periphery, but the tokens that he cites are a negative clause and a *wh*-question and thus not relevant. It may be that, while the preverbal Object DP + pronominal Subject construction in earlier Middle Cornish texts existed to enable rhyme and syllable-count, as in (14) and (17), the author of *Beunans Meriasek*, writing roughly a century later, simply preferred not to employ V2 structures with only the Object DP in the left periphery. Clearly, further research making use of multiple texts is required, which we intend to take up in the future.

We note that *me a leuer* in (28a) is a separate case, since *a lauaraf* with conjugated verb provides the necessary syllable-count itself, i.e. without an overt post-posed pronominal Subject. We call attention to the fact, however, that *me a leuer* occurs earlier in the stanza, as illustrated in (29), so perhaps the author chose to employ the same phrasing, as parallelism is a well-known feature of literary language (Fabb 1997: 137–164 *et passim*).

- | | |
|---|--|
| (29) <i>Nyns o an rena dewov</i>
<i>me a leuer costentyn</i>
<i>ij abostel caradov</i>
<i>y o ze crist cuff colyn</i>
<i>Myr age ymach heb wov</i>
<i>mars yns y havel certyn</i>
<i>ha thyso age hanov</i>
<i>me a leuer pur ylyn</i> | ‘Those were not gods,
I say, Constantine.
Two beloved apostles
They were to Christ the dear heart.
Behold their images without a lie
Whether they are like them certainly,
And to thee their names
I will tell very fairly.’
(BMer. lines 1801–1808) |
|---|--|

²⁹ In (28b), the post posed pronominal *vy* in fact cross-references the object agreement affix.

Appendix II Illustration of John Tregear's translation practice

Grey highlighted text = Bonner's English not translated by Tregear

Bold text = Bonner's English altered or replaced with an equivalent phrase by Tregear

SMALL CAPITAL text = additional words or phrases added by Tregear

Single underscore = English word that had already been borrowed directly into Cornish or that was left untranslated by Tregear

Double underscore = English word (or word previously borrowed from English into Cornish) added by Tregear in his Cornish translation

– *Tregear Homilies, folio 1^r, lines 1–5:*

The Prophette, Daud in **his** fore score and nintenth
*IMA an profet dait i 'n peswar vgangs ha nownsag*³⁰
 be_{3SG.PRES} DEF prophet David in DEF four score and 19

psalme, exhorting all people to **syng** prayse
psalme, |² ow exortya oll AN bobyll the ry prayse HAG HONOR
psalm PROG exhorting_{VN}³¹ all DEF people to give_{VN} praise and honor

to **almighty** god, and to serue him in gladnes, and
the du |³ HA th 'y servya in lowendar, ha
 to God to 3SG_{MASC.POSS} serve_{VN} in gladness and

reioyse in **his** sight,
GANS PERFECT COLONOW THE reiosya |⁴ in sight AGAN CREATOR HA
 with perfect³² hearts to reioyce_{VN} in sight 1PL_{POSS} creator and

³⁰ Note that Tregear here employs the cardinal, rather than the ordinal, numeral.

³¹ Tregear has converted this phrase, which uses a present participle in English, to a Cornish clause with a progressive construction (i.e. a periphrastic present as general present tense; cf. colloquial Modern Welsh). He has also broken what is one sentence in Bonner's text into two sentences in Cornish (each using the periphrastic present tense), with the break coming in line 4.

³² Note the use of the preposed adjective, as per English. This is not unusual when Tregear employs English adjectives to modify a noun, even when the noun itself is translated into Cornish. Note also that the adj. *perfect* does not appear in Bonner's English text. It is an addition by Tregear, who seems to enjoy embellishing and expanding upon Bonner's text as he translates. A comparable example of an untranslated English Adj + N phrase is *sufficient cawse* in line 5.

allegedgeth thys
 REDEMAR.³³ yma AN PROFET DAUID ow |⁵ *allegia helma*
redeemer be_{3SG.PRES} DEF prophet David PROG allege_{VN} PROX

as a sufficient cause **thereof.**
kepar ⟨ha⟩ *dell*³⁴ EW-A *sufficiant cawse* AGAN REDEMPCION
 as (as) (as) be_{3SG.PRES}-3SG_{MASC} sufficient cause 1PL_{POSS} redemption

– *Tregear Homilies, folio 4^r, lines 2–6:*

onely³⁵ one kynde of fruyte **he** charged
saw vn kynda a frut AN TAS DU |³ a *chargias*
 only 1 kind of fruit DEF Father God PTCL charge_{3SG.PRET}

hym vtterlye to refrayne from, on
 MABDEN NA RELLA MYLLYA NA TUCHIA WORTA war
 mankind NEG_{SUB} do_{3SG.IMPF.SUBJ} meddle_{VN} nor touch_{VN} at_{3SG.MASC} upon

payne of death, (and) that **not of** the body
 |⁴ *bayne* MERWALL a vernans henn o MERNANS an corfe
 pain dying_{VN} of death DIST be_{3SG.IMPF} death DEF body

alone, but of the soule also) **which** was the fruyte of
 ha 'n ena |⁵ *inweth henna o a frut a*
 and DEF soul also DIST be_{3SG.IMPF} DEF fruit of

33 Note that the Cornish repeats the preposition the ‘to’ of this infinitive construction, although to is omitted in the English exemplar.

34 *kepar ha* (as originally written) is equivalent to English ‘as’ + DP, e.g. ‘as a sufficient cause,’ but the usual way to say ‘as a cause’ would be *avel* + DP, not *kepar ha* + DP, which usually means ‘just like . . .’ or ‘even as . . .’ Even more surprisingly, Tregear has altered it here to *kepar dell*, which is equivalent to English ‘as’ + V, which requires him to introduce a new verb *ewa* ‘it is’ and add a subordinate clause.

35 Or ‘the sufficient cause of our redemption,’ depending upon how we interpret this ambiguous Cornish phrase. Presumably, Tregear felt it necessary to replace ‘thereof’ with a phrase that specifies ‘of our redemption,’ even though this seems to change the meaning of the text from ‘cause to sing praise to almighty God’, which is how we understand Bonner’s English text.

the tree called in scripture the tree of knowledge
 'n wethan (*gylwys in scripture*) |⁶ an wethan a wothfes
 DEF tree call_{PST-PTCPL} in scripture DEF tree of know_{VN}

of good, and euyt.
 a 'N da ha 'N drog.
 of DEF good and DEF evil

References

[All URLs cited are current as of 11/5/2020.]

Abbreviations

BL	London, British Library
Blathm.	<i>Poems of Blathmac</i> (Carney 1964: 1–88)
BMer.	<i>Beunans Meriasek</i> (Stokes 1872)
CELT	<i>Corpus of Electronic Texts</i> (Färber 2012)
CIH	<i>Corpus Iuris Hibernici</i> (Binchy 1978)
CorPH	<i>Corpus Palaeohibernicum</i> (Stifter et al. 2015–)
DIL	<i>Dictionary of the Irish language</i> (Quin 1913–1975)
eDIL	<i>Electronic dictionary of the Irish language</i> (Toner et al. 2019)
eGPC	The electronic version of <i>Geiriadur Prifysgol Cymru</i> (<i>Geiriadur Prifysgol Cymru</i> online)
ESV	<i>The Holy Bible, English standard version</i> , new classic reference edition (2011). Wheaton, IL: Crossway
FST	Finite-state transducer
GOI	<i>A grammar of Old Irish</i> (Thurneysen 1946)
GPC	<i>Geiriadur Prifysgol Cymru</i> (Thomas et al. 1950–2002)
IrGospThomas	<i>The Irish Gospel of Thomas</i> (Carney 1964: 90–105)
KJV	Authorized <i>King James Version</i> of the Bible (1611)
LEIA	<i>Lexique étymologique de l'irlandais ancien</i> (Vendryes et al. 1959–)
LL	<i>Book of Leinster</i> (Best et al. 1954–1983)
LU	<i>Lebor na hUidre</i> (Best and Bergin 1929)
MBT	Memory-based tagger
MD	<i>Metrical Dindsenchas</i> (Gwynn 1991 [1903–1935])
ML	<i>The Milan Glosses</i> (Thes. 1: 7–483), also edited in Griffith and Stifter (2013)
MS	Manuscript
NLP	Natural Language Processing
NLW	Aberystwyth, National Library of Wales
PKM	<i>Pedeir Keinc y Mabinogi</i> (Williams 1951)
RIA	Dublin, Royal Irish Academy
Sg.	<i>The St Gall Priscian Glosses</i> (Thes. 2: 49–224), also edited in Bauer (2015) and Bauer, Hofman and Moran (2018)
SR	<i>Saltair na Rann</i> (Stokes 1883)
Thes.	<i>Thesaurus Palaeohibernicus</i> (Stokes and Strachan 1901–1910)
TLH	<i>Thesaurus Linguae Hibernicae</i> (Kelly and Fogarty 2006–2011)
Tur.	<i>The Turin Glosses</i> (Thes. 1: 484–494), also edited in Lash (2018)
Wb.	<i>The Würzburg Glosses</i> (Thes. 2: 499–712)

Corpora, Lexicons, and Dictionaries

- Barrett, Siobhán. 2018a. The Poems of Blathmac database. In David Stifter, Siobhán Barrett, Bernhard Bauer, Ellen Ganly, Aaron Griffith, Tianbo Ji, Truc Ha Nguyen, Elliott Lash, Godstime Osarobo, Fangzhe Qiu & Nora White. 2015–present, *Corpus Palaeohibernicum*. Maynooth: Maynooth University. <https://chronhib.maynoothuniversity.ie>.
- Bauer, Bernhard. 2015. *The online database of the Old Irish Priscian Glosses*. <http://www.univie.ac.at/indogermanistik/priscian/>.
- Bauer, Bernhard, Rijklof Hofman & Pádraig Moran. 2018. *St Gall Priscian Glosses, version 2.0*. <http://www.stgallpriscian.ie>.
- Bech, Kristin & Kristine Eide. 2014. *The ISWOC corpus*. University of Oslo: Department of Literature, Area Studies and European Languages. <http://iswoc.github.io>.
- Celano, Giuseppe G. A., Gregory Crane & Bridget Almas. 2014. *The Ancient Greek and Latin dependency treebank, version 2.0*. https://github.com/PerseusDL/treebank_data.
- Dinneen, Patrick S. 1927. *Foclóir Gaedhilge agus Béarla: An Irish-English dictionary*. Dublin: Irish Texts Society.
- Färber, Beatrix (ed.). 2012. *CELT: Corpus of Electronic Texts*. <http://celt.ucc.ie/>.
- Fleuriot, Léon. 1964a. *Dictionnaire des gloses en vieux breton*. Paris: C. Klincksieck.
- Galves, Charlotte & Helena Britto. 2002. *The Tycho Brahe corpus of historical Portuguese*. University of Campinas: Department of Linguistics. <http://www.tycho.iel.unicamp.br/~tycho/>.
- Glare, Peter G. W. 2012 [1968–1982]. *Oxford Latin dictionary*, 2nd edn. Oxford: Clarendon Press.
- Geiriadur Prifysgol Cymru* (GPC) Online. 2014. University of Wales Centre for Advanced Welsh41 & Celtic Studies. <http://www.geiriadur.ac.uk>.
- Griffith, Aaron & David Stifter. 2013. *Dictionary and database of the Old Irish glosses in the Milan MS Ambr. C301 inf*. https://www.univie.ac.at/indogermanistik/milan_glosses/.
- Isaac, Graham, Simon Rodway, Silva Nurmio, Kit Kappahan & Patrick Sims-Williams. 2013. *Rhyddiaith Gymraeg o Lawysgrifau'r 13eg Ganrif, fersiwn 2.0*, Aberystwyth: Prifysgol Aberystwyth. <https://web.archive.org/web/20141223043511/http://cadair.aber.ac.uk/dspace/handle/2160/11163> (original URL: <http://cadair.aber.ac.uk/dspace/handle/2160/11163>).
- Kavanagh, Séamus (author), Dagmar S. Wodtke (ed.). 2001. *A lexicon of the Old Irish glosses in the Würzburg manuscript of the Epistles of St. Paul*. Vienna: Österreichische Akademie der Wissenschaften.
- Kelly, Patricia & Hugh Fogarty. 2006–2011. *Thesaurus Linguae Hibernicae*. <https://www.ucd.ie/th/index.html>.
- King, Dennis, Elliott Lash & Liz Gabay. 2006. In *Dúil Bérlai: Deilbhíocht an Bhriathair*. Online implementation by Caoimhín P. Ó Donnáile. <https://www2.smo.uhi.ac.uk/sengoidelc/duil-belrai/foirmeacha/>.
- Kroch, Anthony & Ann Taylor. 2000. *The Penn-Helsinki parsed corpus of Middle English (PPCME2)*. <https://www.ling.upenn.edu/hist-corpora/>.
- Kroch, Anthony, Beatrice Santorini & Ariel Diertani. 2016. *The Penn parsed corpus of modern British English (PPCMBE2)*. <https://www.ling.upenn.edu/hist-corpora/>.
- Kroch, Anthony, Beatrice Santorini & Lauren Delfs. 2004. *The Penn-Helsinki parsed corpus of early Modern English (PPCEME)*. <https://www.ling.upenn.edu/hist-corpora/>.
- Kroonen, Guus. 2013. *Etymological dictionary of Proto-Germanic*. Leiden/Boston: Brill.

- Lash, Elliott. 2014a. *The parsed Old and Middle Irish corpus (POMIC), version 0.1*. Dublin: Dublin Institute for Advanced Studies, Online Publications. <https://www.dias.ie/celt/celt-publications-2/celt-the-parsed-old-and-middle-irish-corpus-pomic/>.
- Lash, Elliott. 2018. The Minor Glosses database. In David Stifter, Siobhán Barrett, Bernhard Bauer, Ellen Ganly, Aaron Griffith, Tianbo Ji, Truc Ha Nguyen, Elliott Lash, Godstime Osarobo, Fangzhe Qiu & Nora White. 2015–present. *Corpus Palaeohibernicum*. Maynooth: Maynooth University. <https://chronhib.maynoothuniversity.ie>.
- Luft, Diana, Peter W. Thomas & Mark D. Smith (eds.). 2013. *Rhyddiaith Gymraeg 1300–1425*. <http://www.rhyddiaithganoloesol.caerdydd.ac.uk>.
- Matasović, Ranko. 2009. *Etymological dictionary of Proto-Celtic*. Leiden: Brill.
- Ó Dónaill, Niall (ed.). 1977. *Foclóir Gaeilge-Béarla*. Baile Átha Cliath: Oifig an tSólaíthair.
- Qiu, Fangzhe. 2019. The Annals of Ulster database. Stifter In David Stifter, Siobhán Barrett, Bernhard Bauer, Ellen Ganly, Aaron Griffith, Tianbo Ji, Truc Ha Nguyen, Elliott Lash, Godstime Osarobo, Fangzhe Qiu & Nora White. 2015–present. *Corpus Palaeohibernicum*. Maynooth: Maynooth University. <https://chronhib.maynoothuniversity.ie>.
- Quin, Ernest G. (ed.). 1983 [1913–1975]. *Dictionary of the Irish language: Based mainly on Old and Middle Irish materials*, compact edn. Dublin: Royal Irish Academy.
- Scannell, Kevin. 2018. *Droichead DIL [DIL Bridge]* <https://cadhan.com/droichead/>.
- Stifter, David, Siobhán Barrett, Bernhard Bauer, Ellen Ganly, Aaron Griffith, Tianbo Ji, Truc Ha Nguyen, Elliott Lash, Godstime Osarobo, Fangzhe Qiu & Nora White. 2015–present. *Corpus Palaeohibernicum*. Maynooth: Maynooth University. <https://chronhib.maynoothuniversity.ie>.
- Thomas, Richard J., Gareth A. Bevan & Patrick J. Donovan. 1950–2002. *Geiriadur Prifysgol Cymru*. Caerdydd: Gwasg Prifysgol Cymru.
- Toner, Gregory, Máire Ní Mhaonaigh, Sharon Arbuthnot, Marie-Luise Theuerkauf & Dagmar Wodtke. 2019. *Electronic dictionary of the Irish language*, 2019 edn. www.dil.ie.
- Vendryes, Joseph, Édouard Bachellery & Pierre-Yves Lambert. 1959–. *Lexique étymologique de l'irlandais ancien*. Dublin: Dublin Institute for Advanced Studies.
- Wallenberg, Joel C., Anton Karl Ingason, Einar Freyr Sigurðsson & Eiríkur Rögnvaldsson. 2011. *Icelandic parsed historical corpus (IcePaHC), version 0.9*. [https://linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus_\(IcePaHC\)](https://linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus_(IcePaHC)).
- Willis, David & Ingo Mittendorf. 2004. *Corpus Hanesyddol yr Iaith Gymraeg 1500–1850*. <https://web.archive.org/web/20190902191133/http://people.ds.cam.ac.uk/dwew2/hcwl/menu.htm> (original URL: <http://people.ds.cam.ac.uk/dwew2/hcwl/menu.htm>).

Text Editions

- Ahlqvist, Anders (ed. and trans.). 2018. The Milan poem. In Anders Ahlqvist & Pamela O'Neill (eds.), *Fir Fesso: A festschrift for Neil McLeod*, 13–29. Sydney: University of Sydney.
- Albanès, Joseph-Hyacinthe (ed. and trans.). 1879. *La vie de sainte Douceline, fondatrice des beguines de Marseille*. Marseille: Étienne Camoin.
- Atkinson, Robert (ed. and trans.). 1887. *The passions and homilies from Leabhar Breac*. Dublin: Royal Irish Academy.
- Bakere, Jane A. 2009 [1980]. *The Cornish Ordinalia: A critical study*, 2nd edn. Hayle: Kesva an Taves Kernewek.

- Bergin, Osborn & Richard I. Best (eds.). 1934–1938. Tochmarc Étaíne. *Ériu* 12. 137–196.
- Best, Richard I. (ed.). 1936. *The commentary on the psalms with glosses in Old-Irish preserved in the Ambrosian Library (MS. C 301 inf.): Collotype facsimile, with introduction*. Dublin: Royal Irish Academy.
- Best, Richard I. & Osborn Bergin (eds.). 1929. *Lebor na hUidre*. Dublin: Royal Irish Academy.
- Best, Richard I., Osborn Bergin, Michael A. O'Brien & Anne O'Sullivan (eds.). 1954–1983. *The Book of Leinster: Formerly Lebar na Núachongbála*, 6 vols. Dublin: Dublin Institute for Advanced Studies.
- Bieler, Ludwig (ed. and trans.). 1979. *The Patrician texts in the Book of Armagh*. Dublin: Dublin Institute for Advanced Studies.
- Binchy, Daniel A. (ed.). 1941. *Críth Gablach*. Dublin: Dublin Institute for Advanced Studies.
- Binchy, Daniel A. (ed. and trans.). 1955. Irish law tracts re-edited: I. *Coibnes Uisci Thairidne*. *Ériu* 17. 52–85.
- Binchy, Daniel A. (ed. and trans.). 1962. The Old-Irish table of penitential commutations. *Ériu* 19. 47–72.
- Binchy, Daniel A. (ed. and trans.). 1966. Bretha Déin Chécht. *Ériu* 20. 1–66.
- Binchy, Daniel A. (ed.). 1978. *Corpus Iuris Hibernici*, 6 vols. Dublin: Dublin Institute for Advanced Studies.
- Breatnach, Liam (ed. and trans.). 2017a. *Córus Bésgnai: An Old Irish law tract on the church and society* (Early Irish Law Series 7). Dublin: Dublin Institute for Advanced Studies.
- Breatnach, Liam (ed. and trans.). 2017b. The *Trefocal Tract*. In Gordon Ó Riain (ed.), *Dá Dtrian Feasa Fiafraighidh: Essays on the Irish grammatical and metrical tradition*, 1–65. Dublin: Dublin Institute for Advanced Studies.
- Bromwich, Rachel & Daniel S. Evans (eds. and trans.). 1992. *Culhwch ac Olwen: An edition and study of the oldest Arthurian tale*. Cardiff: University of Wales Press.
- Byrne, Mary E. (ed.). 1908. Airec Menman Uraird Meic Cosse. *Anecdota from Irish Manuscripts* 2. 42–76.
- Carney, James (ed. and trans.). 1964. *The poems of Blathmac son of Cú Brettan together with the Irish Gospel of Thomas and a poem on the Virgin Mary* (Irish Texts Society 47). Dublin: Irish Texts Society.
- Charles-Edwards, Thomas & Fergus Kelly (eds. and trans.). 1983. *Bechbretha* (Early Irish Law Series 1). Dublin: Dublin Institute for Advanced Studies.
- Crawford, Samuel J. (ed.). 1922. *The Old English version of the Heptateuch: Aelfric's treatise on the Old and New Testament and his preface to Genesis*. London: Milford.
- Duff, John W. (ed.). 1954. *Minor Latin poets*. London: Heinemann.
- Ehrman, Bart D. & Zlatko Pleše (eds.). 2011. *The apocryphal gospels: Texts and translations*. Oxford: Oxford University Press.
- Ernault, Émile (ed. and trans.). 1887a. La vVie de sainte Catherine: Texte moyen breton. *Revue celtique* 8. 76–95.
- Ernault, Émile (ed. and trans.). 1887b. Vie de sainte Nonn. *Revue celtique* 8. 230–301, 405–491.
- Flint, Valerie I. J. (ed.). 1983. Honorius Augustodunensis: Imago mundi. *Archives d'histoire doctrinale et littéraire du Moyen Âge* 49. 7–153.
- Goetinck, Glenys W. (ed.). 1976. *Historia Peredur vab Efracw*. Caerdydd: Gwasg Prifysgol Cymru.
- Greene, David (ed.). 1955. *Fingal Rónáin and other stories* (Mediaeval and Modern Irish Series 16). Dublin: Dublin Institute for Advanced Studies.

- Gwynn, Edward (ed. and trans.). 1991 [1903–1935]. *The Metrical Dindshenchas*, 5 vols (Todd Lecture Series 8–12). Dublin: Dublin Institute for Advanced Studies.
- Gwynn, Edward (ed. and trans.). 1914. An Irish penitential. *Ériu* 7. 121–195.
- Hertz, Martin. 2009 [1855–1858]. *Prisciani grammatici Caesariensis Institutionum grammaticarum libri XVIII*, 2 vols (Grammatici Latini, ex recensione Henrici Keilii 1–2). Cambridge: Cambridge University Press.
- Hofman, Rijcklof (ed.). 1996. *The Sankt Gall Priscian commentary, part I*, 2 vols (Studien und Texte zur Keltologie 1). Münster: Nodus.
- Jenkins, Dafydd & Morfydd Owen. 1984. The Welsh marginalia in the Lichfield Gospels, part II: The “Surexit” memorandum. *Cambridge medieval Celtic studies* 7. 91–120.
- Jones, Gwenan (ed.). 1939. Gwyrthyeu y Wynvydedic Veir, part I. *Bulletin of the Board of Celtic Studies* 9. 144–148, 334–341.
- Jones, Gwenan (ed.). 1941. Gwyrthyeu y Wynvydedic Veir, part II. *Bulletin of the Board of Celtic Studies* 10. 21–33.
- Kelly, Fergus (ed. and trans.). 1973. A poem in praise of Columb Cille. *Ériu* 24. 1–34.
- Lambert, Pierre-Yves. 2003. The Old Welsh glosses on weights and measures. In Paul Russell (ed.). *Yr hen iaith*, 103–134. Aberystwyth: Celtic Studies Publications.
- Lewis, Henry (ed.). 1925. Cynghorau Catwn. *Bulletin of the Board of Celtic Studies* 2. 16–25.
- Lewis, Henry (ed.). 1930. Credo Athanasius Sant. *Bulletin of the Board of Celtic Studies* 5. 193–203.
- Loth, Joseph. 1902. Études corniques: 2. Textes inédits en cornique moderne. *Revue celtique* 23. 173–200.
- MacNeill, John. 1913. Poems by Flann Mainistrech on the Dynasties of Ailech, Mide and Brega. *Archivium Hibernicum* 2. 37–99.
- McManus, Damian & Eoghan Ó Raghallaigh. 2010. *A Bardic miscellany: Five hundred bardic poems from manuscripts in Irish and British libraries* (Trinity Irish Studies 2). Dublin: Department of Irish, Trinity College Dublin.
- Meid, Wolfgang (ed.). 1974. *Táin Bó Fraích*, 2nd edn. Dublin: Dublin Institute for Advanced Studies.
- Meyer, Kuno (ed. and trans.). 1892. *Aislinge Meic Conglinne—The Vision of MacConglinne: A Middle-Irish wonder tale*. London: David Nutt.
- Meyer, Kuno (ed. and trans.). 1894. *Hibernica Minora: Being a fragment of an Old-Irish treatise* (Anecdota Oxoniensia 8). Oxford: Clarendon Press.
- Meyer, Kuno (ed. and trans.). 1899. Gein Branduib maic Echach ocus Aedáin maic Gabráin inso síis. *Zeitschrift für celtische Philologie* 2. 134–137.
- Meyer, Kuno (ed. and trans.). 1903. Stories and songs from Irish manuscripts: King Eochaid has horse’s ears. *Otia Merseiana* 3. 46–54.
- Meyer, Kuno (ed.). 1903. Eine altirische Homilie. *Zeitschrift für celtische Philologie* 4. 241–243.
- Meyer, Kuno (ed.). 1912. Mitteilungen aus irischen Handschriften: Die Helden von Emain Macha. *Zeitschrift für celtische Philologie* 8. 217–218.
- Mulchrone, Kathleen (ed. and trans.). 1939. *Bethu Phátraic: The tripartite Life of Patrick*. Dublin: Royal Irish Academy.
- Murdoch, Brian. 1981. Pascon agan Arluth: The literary position of the Cornish poem of the Passion. *Studi medievali* 22. 821–836.
- Murphy, Gerard (ed. and trans.). 1956. *Early Irish lyrics: Eighth to twelfth century*. Oxford: Clarendon Press.
- Nance, R. Morton. 1949. A Cornish poem restored. *Old Cornwall* 4. 368–371.

- Neuhaus, Carl (ed.). 1886. *Die lateinischen Vorlagen zu den alt-französischen Adgar'schen Marien-Legenden*. Heilbronn: Gebr. Henninger.
- Norris, Edwin (ed. and trans.). 1859. *The ancient Cornish drama*. Oxford: Oxford University Press.
- Ó Cuív, Brian (ed.). 1966. Some items from Irish tradition. *Éigse* 11. 167–187.
- Ó Murchadha, Diarmuid (ed. and trans.). 2009. *Lige Guill. The grave of Goll: A Fenian poem from the Book of Leinster* (Irish Texts Society 62). London: Irish Texts Society.
- Ó Néill, Pádraig (ed.). 2012. *Exegetica: Psalterium Suthantoniense* (Corpus Christianorum: Continuatio Mediaevalis 240, Scriptores Celtigenae 6). Turnhout: Brepols.
- O'Brien, Michael A. (ed. and trans.). 1952. A Middle-Irish poem on the birth of Áedán mac Gabráin and Brandub mac Echach. *Ériu* 16. 157–170.
- O'Donoghue, Tadhg. 1921. Advice to a prince. *Ériu* 9. 43–54.
- O'Donovan, John (ed. and trans.). 1856. *Annala Ríoghachta Éireann: Annals of the Kingdom of Ireland by the Four Masters, from the earliest period to the year 1616*. Dublin: Hodges, Smith & Co.
- O'Rahilly, Cecile (ed. and trans.). 1967. *Táin Bó Cúailnge from the Book of Leinster*. Dublin: Dublin Institute for Advanced Studies.
- Oskamp, Hans P. A. (ed. and trans.). 1970. *The voyage of Máel Dúin: A study in early Irish voyage literature*. Groningen: Wolters-Noordhoff.
- Padel, Oliver J. (ed. and trans.). 1975. *The Cornish writings of the Bosun (i.e. Boson) family: Nicholas, Thomas and John Bosun (i.e. Boson), of Newlyn, circa 1660–1730*. Redruth: Institute of Cornish Studies.
- Parry, John J. (ed. and trans.). 1937. *Brut y Brenhinedd: Cotton Cleopatra version*. Cambridge, MA: The Medieval Academy of America.
- Pauphilet, Albert (ed.). 1923. *La quête del Saint Graal*. Paris: E. Champion.
- Pósdör, Dóra. 1999. *Twelve poems Attributed to Flann Mainistrech from the Book of Leinster*. Dublin: Trinity College Dublin dissertation.
- Pool, Peter A. S. & Oliver J. Padel. 1975–1976. William Bodinar's letter, 1776. *Journal of the Royal Institution of Cornwall* 7. 231–236.
- Qiu, Fangzhe. 2018. The first judgment in Ireland. In Anders Ahlqvist & Pamela O'Neill (eds.), *Fír Fesso: A Festschrift for Neil McLeod*, 185–201. Sydney: The University of Sydney.
- Reeve, Michael D. (ed.) & Neil Wright (trans.). 2007. *Geoffrey of Monmouth: The history of the kings of Britain; An edition and translation of De gestis Britonum (Historia regum Britanniae)* (Arthurian Studies 69). Woodbridge: Boydell Press.
- Richards, Melville (ed.). 1948. *Breudwyt Ronabwy*. Caerdydd: Gwasg Prifysgol Cymru.
- Roberts, Brynley F. (ed.). 1971. *Brut y brenhinedd: Llanstephan MS.1 version*. Dublin: Dublin Institute for Advanced Studies.
- Roberts, Brynley F. (ed.). 1975. *Cyfranc Lludd a Llefelys*. Dublin: Dublin Institute for Advanced Studies.
- Sims-Williams, Patrick. 2016. The Welsh versions of Geoffrey of Monmouth's "History of the Kings of Britain". In Axel Harlos & Neele Harlos (eds.), *Adapting texts and styles in a Celtic context*, 53–74. Münster: Nodus Publikationen.
- Smith, Peter. 2001. Mide maigen Clainne Cuind: A medieval poem on the kings of Mide. *Peritia* 15. 108–144.
- Smith, Peter. 2007. *Three historical poems ascribed to Gilla Cómáin*. Münster: Nodus Publikationen.

- Stam, Nike. 2010. *Aided Chúanach mac Cailchíne: Aggressive Tribes and Aggressive Trees; A Critical Edition*. Utrecht: Utrecht University MA thesis. <https://dspace.library.uu.nl/handle/1874/179333>
- Stern, Ludwig C. 1910. *Epistolae beati Pauli glosatae glosa interlineali: Irisch-lateinischer Codex der Würzburger Universitätsbibliothek*. Halle an der Saale: Niemeyer.
- Stimming, Albert (ed.). 1899. *Der anglonormannische Boeve de Haumtone*. Halle: Max Niemeyer.
- Stokes, Whitley (ed. and trans.). 1860–1861. *The Passion: A Middle-Cornish poem*. Appendix to *Transactions of the Philological Society* for 1860–1861. 1–100.
- Stokes, Whitley (ed. and trans.). 1872. *Beunans Meriasek: The Life of Saint Meriasek, bishop and confessor*. London: Trübner and Co.
- Stokes, Whitley (ed.). 1883. *The Saltair na Rann: A collection of early Middle Irish poems*. Oxford: Clarendon Press.
- Stokes, Whitley (ed. and trans.). 1887. *The tripartite Life of Patrick with other documents relating to that saint*. London: Stationery Office.
- Stokes, Whitley (ed. and trans.). 1891. Adamnan's second vision. *Revue celtique* 12. 420–443.
- Stokes, Whitley (ed. and trans.). 1904. Tidings of the resurrection. *Revue celtique* 25. 234–259.
- Stokes, Whitley (ed. and trans.). 1905. *Féilire Óengusso Céili Dé: The martyrology of Oengus the Culdee*. London: Henry Bradshaw Society.
- Stokes, Whitley & John Strachan (eds.). 1901–1910. *Thesaurus palaeohibernicus: A collection of Old-Irish glosses, scholia, prose, and verse*, 2 vols with suppl. Cambridge: Cambridge University Press.
- Strachan, John. 1907. An Old-Irish homily. *Ériu* 3. 1–10.
- Sweet, Henry (ed.). 1883. *King Alfred's Orosius, vol. 1: Old-English and Latin original*. London: Trübner.
- Syed, Keith (trans.), & Ray Edwards (ed.). 1996. *Beunans Meriasek-Bywnans Meryasek*. Sutton Coldfield: Kesva an Taves Kernewek.
- Thomas, Graham & Nicholas Williams (eds. and trans.). 2007. *Bewnans Ke: The Life of Saint Kea*. Exeter: University of Exeter Press.
- Thomson, Robert L. (ed.). 1968. *Owein, or Chwedyl Iarlles y Ffynnwain*. Dublin: Dublin Institute for Advanced Studies.
- Thomson, Robert L. (ed.). 1986. *Pwyll Pendeuic Dyuet*. Dublin: Dublin Institute for Advanced Studies.
- Thomson, Robert L. (ed.). 1997. *Ystoria Gereint Uab Erbyn*. Dublin: Dublin Institute for Advanced Studies.
- Thurneysen, Rudolf (ed.). 1949. *Old Irish reader*. Dublin: Dublin Institute for Advanced Studies.
- Toorians, Luran (ed. and trans.). 1991. *The Middle Cornish charter endorsement: The making of a marriage in medieval Cornwall*. Innsbruck: Institut für Sprachwissenschaft der Universität.
- Ua Nualláin, Tomás P. 1904. The quarrel about the loaf. *Ériu* 1. 128–137.
- Weber, Robert (ed.). 2007. *Biblia sacra iuxta Vulgatam versionem*, 5th edn, prepared by Roger Gryson. Stuttgart: Deutsche Bibelgesellschaft.
- White, Nora (ed. and trans.). 2006. *Compert Mongáin and three other early Mongán tales*. Maynooth: Department of Old Irish, National University of Ireland, Maynooth.
- Williams, Ifor (ed.). 1908. *Breuddwyd Maxen*. Bangor: Jarvis & Foster.
- Williams, Ifor (ed.). 1951. *Pedeir Keinc y Mabinogi*, 2nd edn. Caerdydd: Gwasg Prifysgol Cymru.

- Williams, Ifor. 1980. The Juvenius poems. In Rachel Bromwich (ed. with notes). *The beginnings of Welsh poetry: Studies by Sir Ifor Williams*, 2nd edn, 89–121. Cardiff: University of Wales Press.
- Williams, Robert (ed. and trans.). 1892. *Selections from the Hengwrt manuscripts preserved in the Peniarth Library*, vol. 2. London: Bernard Quaritch.
- Wright, Neil (ed.). 1988. *The historia regum Britannie of Geoffrey of Monmouth, vol. 2: The first variant version: A critical edition*. Cambridge: Boydell & Brewer.
- Zupitza, Julius (ed.). 1966. *Aelfrics Grammatik und Glossar*. Hildesheim: Weidmann.

Secondary Literature

- Abdi, Hervé & Lynne J. Williams. 2010. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (4). 433–459.
- Adesam, Yvonne & Gerlof Bouma. 2016. Old Swedish part-of-speech tagging between variation and external knowledge. In Nils Reiter, Beatrice Alex & Kalliopi A. Zervanou (eds.). *Association for Computational Linguistics: SIGHUM Workshop on language technology for cultural heritage, social sciences, and humanities (LaTeCH 2016) 10, Berlin, August 11, 2016*. 32–42.
- Ahlqvist, Anders. 1988. Remarks on the question of dialects in Old Irish. In Jacek Fisiak (ed.). *Historical dialectology: Regional and social*, 23–38. Berlin, New York & Amsterdam: Mouton de Gruyter.
- Ahlqvist, Anders. 1994. Litriú na Gaeilge. In Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach (eds.), *Stair na Gaeilge in ómós do P[h]ádraig Ó Fiannachta*, 23–59. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- Ahlqvist, Anders. 2014. Remarks on Old-Irish verbs and predicates: In eulogy of Pieter Seuren as historian of linguistics. <https://web.archive.org/web/20160420221219/> (original URL: <http://www.mpi.nl/people/seuren-pieter/symposium/>).
- Alberti, Chris, Daniel Andor, Ivan Bogatyy, Michael Collins, Dan Gillick, Lingpeng Kong, Terry Koo, Ji Ma, Mark Omernick, Slav Petrov, Chayut Thanapirom, Zora Tung & David Weiss. 2017. SyntaxNet models for the CoNLL 2017 shared task. <https://arxiv.org/abs/1703.04929>; *arXiv: 1703.04929 [cs.CL]*.
- Backhaus, Norbert. 1990. The structure of the list of *Remscéla Tána bó Cualngi* in the *Book of Leinster*. *Cambridge Medieval Celtic Studies* 19. 19–26.
- Bamman, David & Gregory Crane. 2011. The Ancient Greek and Latin dependency treebanks. In Caroline Sporleder, Antal van den Bosch & Kalliopi Zervanou (eds.), *Language technology for cultural heritage*, 79–98. Berlin: Springer.
- Beesley, Kenneth R. & Lauri Karttunen. 2003. *Finite state morphology*. Stanford, CA: Center for the Study of Language and Information (CSLI) Publications.
- Benincà, Paola & Cecilia Poletto. 2004. Topic, focus, and V2. In Luigi Rizzi (ed.), *The structure of CP and IP* (Cartography of Syntactic Structures 2), 52–75. Oxford: Oxford University Press.
- Berdičevskis, Aleksandr, Hanne Eckhoff & Tatiana Gavrilova. 2016. The beginning of a beautiful friendship: Rule-based and statistical analysis of Middle Russian. In Vladimir Selegey, A. V. Baytin, V. I. Belikov, I. M. Boguslavsky, B. V. Dobrov, D. O. Dobrovol'skij,

- L. L. Iomdin, E. Hovy, I. M. Kobozeva, E. B. Kozerenko, M. A. Krongauz, N. I. Laufer, N. V. Loukachevich, D. McCarthy, P. Nakov, J. Nivre, G. S. Osipov, A. Ch. Piperski, V. Raskin, S. A. Sharoff, T.E. Yanko & L. M. Zakharov (eds.), *Dialogue (2016) 15: Computational linguistics and intellectual technologies, Moscow, Russian State University for Humanities, 1–6 June 2016*, 99–111. <http://www.dialog-21.ru/en/digest/2016/titlepage/>.
- Bezanson, Jeff Stefan Karpinski, Viral B. Shah & Alan Edelman. 2012. Julia: A fast dynamic language for technical computing. *CoRR abs/1209.5145*. <http://arxiv.org/abs/1209.5145>; *arXiv:1209.5145v1 [cs.PL]*.
- de Bhaldraithe, Tomás. 1981. *Innéacs Nua-Ghaeilge don 'Dictionary of the Irish language'*. Dublin: Royal Irish Academy.
- Bhreathnach, Edel. 2002. Two contributors to the Book of Leinster: Bishop Finn of Kildare and Gilla na Náem Úa Duinn. In Michael Richter & Jean-Michel Picard (eds.), *Ogma: Essays in Celtic studies in honour of Próinséas Ní Chatháin*, 105–111. Dublin: Four Courts Press.
- Birnbaum, David J. & Hanne Martine Eckhoff. 2018. Machine-assisted multilingual word alignment of the Old Church Slavonic Codex Suprasliensis. In Stephen M. Dickey & Mark Richard Lauersdorf (eds.), *V zeleni drželi zeleni breg: Studies in Honor of Marc L. Greenberg*, 1–14. Bloomington: Slavica Publishers.
- Bisagni, Jacopo. 2012. The origins of the preterite of the Old Irish copula and substantive verb: an overview and new ideas. *Journal of Celtic Linguistics* 14. 1–29.
- Bloch-Trojnar, Maria & Mark Ó Fionnáin (eds.). 2019. *Centres and peripheries in Celtic linguistics*. Berlin: Peter Lang Verlag.
- Bock, Albert & Benjamin Bruch. 2008. *An outline of the standard written form of Cornish*. Cornwall: The Cornish Language Partnership.
- Bock, Albert & Benjamin Bruch. 2010. Nucleus length and vocalic alternation in Cornish diphthongs. *Die Sprache* 48. 34–43.
- Bock, Albert & Benjamin Bruch. 2012. New perspectives on vocalic alternation in Cornish. *Keltische Forschungen* 5. 55–97.
- Boling, Bruce D. 1972. Some problems of the phonology and morphology of the Old Irish verb. *Ériu* 23. 73–101.
- Borsley, Robert D. & Andreas Kathol. 2000. Breton as a V2 language. *Linguistics* 38. 665–710.
- Borsley, Robert D., Margaret Tallerman & David W. Willis. 2007. *The syntax of Welsh*. Cambridge: Cambridge University Press.
- Boyle, Elizabeth & Deborah Hayden. 2014. Introduction: Authority and adaptation in medieval Ireland. In Elizabeth Boyle & Deborah Hayden (eds.), *Authorities and adaptations: The reworking and transmission of textual sources in medieval Ireland*, xvii–xlvii. Dublin: Dublin Institute for Advanced Studies.
- Brants, Thorsten. 2000. TnT: a statistical part-of-speech tagger. In Sergei Nirenburg (ed.), *ANLP '00: Applied natural language processing 6*, 224–231. Stroudsburg, PA: Association for Computational Linguistics.
- Bray, Tim, Jean Paoli, Charles M. Sperberg-McQueen, Eve Maler & François Yergeau (eds.). 2008. *Extensible markup language (XML) 1.0*, 5th edn. <http://www.w3.org/TR/2008/REC-xml-20081126/>.
- Breatnach, Liam. 1983. *Varia IV: 3. Do-midethar in the meaning 'guesses, solves a riddle'*. *Ériu* 34. 195.

- Breatnach, Liam. 1994a. An Mheán-Ghaeilge. In Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach (eds.), *Stair na Gaeilge in ómós do P[h]ádraig Ó Fiannachta*, 221–333. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- Breatnach, Liam. 1994b. Varia II: 2. Prepositions with added vowel in relative compound verbs. *Ériu* 45. 197–198.
- Breatnach, Liam. 1996. Poets and poetry. In Kim McCone & Katharine Simms (eds.), *Progress in medieval Irish studies*, 65–77. Maynooth: Department of Old Irish, Saint Patrick's College, Maynooth.
- Breatnach, Liam. 2003. On words ending in a stressed vowel in Early Irish. *Ériu* 53. 133–142.
- Breatnach, Liam. 2005. *A companion to the Corpus Iuris Hibernici* (Early Irish Law Series 5). Dublin: Dublin Institute for Advanced Studies.
- Breatnach, Liam. 2015. Lebor na hUidre: Some linguistic aspects. In Ruairí Ó hUiginn (ed.), *Lebor na hUidre* (Codices Hibernenses Eximii 1), 53–77. Dublin: Royal Irish Academy.
- Breatnach, Liam (author), Elisa Roma (trans.). 2013. *Il medioirlandese-Middle Irish*. NY: Edwin Mellen Press.
- Breatnach, Liam, Ruairí Ó hUiginn, Damian McManus & Katherine Simms (eds.). 2015. *International congress of Celtic studies 14, Maynooth University, 1–5 August 2011*. Dublin: Dublin Institute for Advanced Studies.
- Bresnan, Joan. 2001. *Lexical-functional syntax*. Oxford: Blackwell.
- Brinton, Laurel J. & Elizabeth C. Traugott. 2005. *Lexicalization and language change* (Research Surveys in Linguistics), 2nd edn. Cambridge: Cambridge University Press.
- Bronner, Dagmar. 2016. Nasalierung im Buch von Armagh: Überlegungen zu altirischen Schreibkonventionen. *Zeitschrift für celtische Philologie* 63. 29–48.
- Brown, Wella. 2001. *A grammar of modern Cornish*, 3rd edn. Saltash: Kesva an Taves Kernewek.
- Bruch, Benjamin. 2005. “*Du gveras a. b. c / An pen can henna yv d*”: *Cornish verse forms and the evolution of Cornish prosody, c. 1350–1611*. Cambridge, MA: Harvard University dissertation.
- Bruch, Benjamin. 2009. Medieval Cornish versification: An overview. *Keltische Forschungen* 4. 55–126.
- Budassi, Marco & Elisa Roma. 2018. On the origin of the absolute vs. conjunct opposition in Insular Celtic. *Indogermanische Forschungen* 123. 293–338.
- Büring, Daniel. 1999. Topic. In Peter Bosch & Rob van der Sandt (eds.), *Focus: Linguistic, cognitive, and computational perspectives*, 142–165. Cambridge: Cambridge University Press.
- Carnie, Andrew (ed.). 2011. *Formal approaches to Celtic linguistics*. Newcastle upon Tyne: Cambridge Scholar Publishing.
- Celano, Giuseppe G. A., Gregory Crane & Saeed Majidi. 2016. Part of speech tagging for Ancient Greek. *Open Linguistics* 2016 (2). 393–399.
- Chafe, Wallace. 1987. Cognitive constraints on information flow. In Russell S. Tomlin (ed.), *Coherence and grounding in discourse*, 21–51. Amsterdam: John Benjamins.
- Charles-Edwards, Thomas M. 1971. Wb. 28 c 14 and the ‘exclusive’ use of the equative in Old Irish. *Ériu* 22. 188–189.
- Cowgill, Warren. 1967. On the fate of *w in Old Irish. *Language* 43. 129–138.
- Cowgill, Warren. 1975. The origin of the Insular Celtic conjunct and absolute verbal inflexions. In Helmut Rix (ed.), *Flexion und Wortbildung*, 40–70. Wiesbaden: Reichert.

- Daelemans, Walter & Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge: Cambridge University Press.
- Daelemans, Walter, Jakub Zavrel, Antal van den Bosch & Ko van der Sloot. 2010. *Mbt: memory-based tagger, version 3* (ILK Technical Report Series 10-04).
- Davies, Robert Rees. 1966. The twilight of Welsh law: 1284–1536. *History* 51 (172). 143–164.
- Davies, Sioned. 1998. Written text as performance: The implications for Middle Welsh prose narratives. In Huw Pryce (ed.), *Literacy in medieval Celtic societies*, 133–148. Cambridge: Cambridge University Press.
- Dereza, Oksana. 2018. Lemmatization for ancient languages: rules or neural networks? In Dmitry Ustalov, Andrey Filchenkov, Lidia Pivovarova & Jan Žižka (eds.), *Artificial intelligence and natural Language (AINL-2018) 7, St. Petersburg, Russia, 17–19 October 2018* (Communications in Computer and Information Science 930), 35–47. Cham, Switzerland: Springer.
- Doyle, Adrian, John Philip McCrae & Clodagh Downey. 2019. A character-level LSTM network model for tokenizing the Old Irish text of the Würzburg Glosses on the Pauline Epistles. *Proceedings of the Celtic Language Technology Workshop 2019, co-located with MT Summit XVII, Dublin City University, Dublin, 19–23 August 2019*. 70–79. <https://www.aclweb.org/anthology/W19-6910.pdf>.
- Duncan, Elizabeth. 2012. A Reassessment of the script and make-up of Lebor na Nuachongbála. *Zeitschrift für celtische Philologie* 59. 22–66.
- Eckhoff, Hanne M. & Aleksandr Berdičevskis. 2015. Linguistics vs. digital editions: The Tromsø Old Russian and OCS Treebank. *Scripta & e-Scripta* 14–15. 9–25.
- Eckhoff, Hanne M. & Aleksandr Berdičevskis. 2016. From diachronic treebank to dictionary resource: the Varangian Rus project. In Tinatin Margalitadze and George Meladze (eds.), *European association for lexicography (EURALEX) 17*, 335–339. Tbilisi: Ivane Javakishvili Tbilisi State University.
- Eckhoff, Hanne, Kristin Bech, Gerlof Bouma, Kristine Eide, Dag Haug, Odd Einar Haugen & Marius Jøhndal. 2018. The PROIEL treebank family: a standard for early attestations of Indo-European languages. *Language Resources and Evaluation* 52 (1). 29–65.
- Evans, David E. 1967. *Gaulish personal names*. Oxford: Clarendon Press.
- Eska, Joseph F. 2003. The distribution of the Old Irish personal object affixes and forward reconstruction. In Karlene Jones-Bley, Martin E. Huld, Angela della Volpe & Miriam Robbins Dexter (eds.), *UCLA Indo-European conference 14, 8–9 November 2002*, 25–36. Washington, DC: Institute for the Study of Man.
- Eska, Joseph F. 2010. Where have all the object pronouns gone? The growth of object agreement in earlier Celtic. *Zeitschrift für celtische Philologie* 57. 25–47.
- Evans, Simon D. 1964. *A grammar of Middle Welsh*. Dublin: Dublin Institute for Advanced Studies.
- Fabb, Nigel. 1997. *Linguistics and literature: Language in the verbal arts of the world*. Oxford: Blackwell.
- Feuth, Els. 1982. Two segments or one? Nasalized voiced plosives in Old Irish. *Zeitschrift für celtische Philologie* 39. 88–95.
- Filppula, Markku, Juhani Klemola & Heli Paulasto. 2008. *English and Celtic in contact*. London: Routledge.
- Fleuriot, Léon. 1964b. *Le vieux Breton: Éléments d'une grammaire*. Paris: C. Klincksieck.
- Fomin, Maxim & Gregory Toner. 2006. Digitizing a dictionary of medieval Irish: The eDIL project. *Literary and Linguistic Computing* 21(1). 83–90.

- Fowler David C. 1961. The date of the Cornish Ordinalia. *Mediaeval studies* 23. 91–125.
- Fransen, Theodorus. 2019. *Past, present and future: Computational approaches to mapping historical Irish cognate verb forms*. Dublin: Trinity College Dublin dissertation. <http://hdl.handle.net/2262/89498>.
- Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 87–116. Amsterdam: John Benajmins.
- García-Castillero, Carlos. 1999. *La formación del tema de presente primario osco-umbro*. Vitoria: Universidad del País Vasco dissertation.
- García-Castillero, Carlos. 2001. Zusammenfall von Kasusformen im indogermanischen Personalpronomen. *Historische Sprachforschung* 114. 259–284.
- García-Castillero, Carlos. 2012. The paradigm of clause types in Old Irish (The morphological encoding of illocutionary force). In Harold Craig Melchert (ed.), *The Indo-European verb: Society for Indo-European studies, Los Angeles, 13–15 September 2010*, 61–72. Wiesbaden: Reichert.
- García-Castillero, Carlos. 2013. Las reglas del juego: Notas para una noción de ley morfológica. In Ricardo Gómez, Joaquín Gorrochategui, Joseba A. Lakarra & Céline Mounole (eds.), *Luis Michelena Chair–Koldo Mitxelena Katedraren–Cátedra Luis Michelena 3*, 153–166. Vitoria-Gasteiz: UPV/EHuren Argitalpen Zerbitzua.
- García-Castillero, Carlos. 2014. De-adjectival preverbs in the Old Irish verbal complex: a synchronic and diachronic study. *Zeitschrift für celtische Philologie* 61. 57–100.
- García-Castillero, Carlos. 2015. The type *táinic* in the Old Irish glosses: affix ordering, frequency and phonotactics. *Transactions of the Philological Society* 113 (1). 76–104.
- García-Castillero, Carlos. 2017. The split *for*, a gecko morpheme in the Old Irish verbal complex. In Ivo Hajnal, Daniel Kölligan, and Katharina Zipser (eds.), *Miscellanea Indogermanica: Festschrift für José Luis García Ramón*, 195–206. Innsbruck: Innsbrucker Beiträge zur Sprachwissenschaft.
- García-Castillero, Carlos. 2018. On morphological internalization: The origin of the Old Irish oblique relative conjunct particle *-(s)a^N*. *Diachronica* 35 (1). 88–124.
- García-Castillero, Carlos. 2020 *Clause typing in the Old Irish verbal complex: Morphological expression, paradigmatic organization, and syntactic implications*. Berlin: De Gruyter.
- Gast, Volker & Peter Siemund. 2006. Rethinking the relationship between SELF-intensifiers and reflexives. *Linguistics* 44 (2). 343–381.
- Gelderen, Elly van. 2000. *A history of English reflexive pronouns: Person, self, and interpretability*. Amsterdam/Philadelphia: John Benajmins.
- Gelderen, Elly van. 2019. Reflexive pronouns in the Lindisfarne glosses. *NOWELE: North-Western European Language Evolution* 72 (2). 220–244.
- Gendall, Richard. 2004. *A student's grammar of Late and Modern Cornish*, 2nd edn. Liskeard: Richard Gendall.
- George, Ken. 1990. A comparison of word order in Middle Breton and Middle Cornish. In Martin J. Ball, James Fife, Erich Poppe & Jenny Rowland (eds.), *Celtic linguistics-leithyddiaeth Geltaidd: Readings in the Brythonic languages; Festschrift for T. Arwyn Watkins*, 225–240. Amsterdam: John Benajmins.
- George, Ken. 1991. Notes on word order in *Beunans Meriasek*. In James Fife & Erich Poppe (eds.), *Studies in Brythonic word order*, 205–250. Amsterdam: John Benajmins.
- George, Ken. 2010. Cornish. In Martin J. Ball & Nicole Müller (eds.), *The Celtic languages* (Routledge Language Family Descriptions), 2nd edn, 488–535. London: Routledge.

- Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 5–41. Amsterdam: John Benjamins.
- Greene, David. 1958. The analytic forms of the verb in Irish. *Ériu* 18. 108–112.
- Greene, David. 1966. *The Irish language*. Dublin: Three Candles.
- Greene, David. 1971. Varia II: 2. Ir. úathad, óthad: W. odid. *Ériu* 22. 178–180.
- Greene, David. 1973. Synthetic and analytic: a reconsideration. *Ériu* 24. 121–133.
- Greene, David. 1976. The diphthongs of Old Irish. *Ériu* 27. 26–45.
- Griffith, Aaron. 2008. The animacy hierarchy and the distribution of the *notae augentes* in Old Irish. *Ériu* 58. 55–75.
- Griffith, Aaron. 2016a. On the Old Irish third palatalization and the 3sg. present of the copula. *Ériu* 66. 39–62.
- Griffith, Aaron. 2016b. Pre-verbal *ceta* ‘first’ in the glosses (and some thoughts on the origin of the *notae augentes*). Paper presented at the Workshop on Computational Linguistics and the Dating of early Irish texts, Maynooth University, 15, December 2016.
- Griffith, Aaron & David Stifter. 2014. New and corrected ms. readings in the Milan glosses. *Études celtiques* 40. 53–84.
- Griffith, Aaron, David Stifter & Greg Toner. 2018. Early Irish lexicography: A research survey. *Kratylos* 63. 1–28.
- Haegeman, Liliane. 2012. *Adverbial clauses, main clause phenomena, and composition of the left periphery*. Oxford: Oxford University Press.
- Harbert, Wayne. 2007. *The Germanic languages*. Cambridge: Cambridge University Press.
- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 2 (Handbücher zur Sprach- und Kommunikationswissenschaft 20.2), 1492–1510. Berlin: de Gruyter.
- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 2, 211–242. Mahwah, NJ: Lawrence Erlbaum.
- Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.). 2001. *Language typology and language universals: An international handbook*, 2 vols (Handbücher zur Sprach- und Kommunikationswissenschaft 20.1, 20.2). Berlin: de Gruyter.
- Haug, Dag T. T. 2015. Treebanks in historical linguistic research. In Carlotta Viti (ed.), *Perspectives on historical syntax*, 185–202. Amsterdam: John Benjamins.
- Haug, Dag T. T. & Marius L. Jøhndal. 2008. Creating a parallel treebank of the Old Indo-European Bible translations. In Caroline Sporleder, Kiril Ribarov, Anta van den Bosch, Milena P. Dobрева, Matthew James Driscoll, Claire Grover, Piroska Lendvai, Anke Luedeling & Marco Passarotti (eds.), *Language technology for cultural heritage data workshop (LaTeCH-2008) 2*, 27–34. Marrakech: European Language Resources Association.
- Haug, Dag T. T., Hanne M. Eckhoff & Eirik Welø. 2014. The theoretical foundations of givenness annotation. In Kristin Bech & Kristine Eide (eds.), *Information structure and syntactic change in Germanic and Romance languages*, 17–52. Amsterdam: Benjamins.

- Haug, Dag T. T., Hanne M. Eckhoff, Marek Majer & Eirik Welo. 2009. Breaking down and putting back together: Analysis and synthesis of New Testament Greek. *Journal of Greek Linguistics* 9 (1). 56–92.
- Haug, Dag T. T., Marius L. Jøhndal, Hanne M. Eckhoff, Eirik Welo, Mari J. B. Hertenberg & Angelika Müth. 2009. Computational and linguistic issues in designing a syntactically annotated parallel corpus of Indo-European languages. *Traitement Automatique des Langues* 50 (2). 17–45.
- Heine, Bernd & Tanya Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language* 27 (3). 529–572.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diwald (eds.), *New reflections on grammaticalization*, 83–101. Amsterdam and Philadelphia: John Benjamins.
- Hinterhölzl, Roland & Svetlana Petrova. 2010. From V1 to V2 in West Germanic. *Lingua* 120. 315–328.
- Hulden, Mans. 2009. Foma: a finite-state toolkit and library. *European chapter of the Association for Computational Linguistics: Demonstration Session (EACL-2009) 12*, 29–32. <http://www.aclweb.org/anthology/E09-2008.pdf>.
- Huws, Daniel. 2000. *Medieval Welsh manuscripts*. Cardiff: University of Wales Press.
- Inglese, Guglielmo, Maria Molina & Hanne Eckhoff. 2018. Incorporating Hittite into PROIEL: A pilot project. In Andrew U. Frank, Christine Ivanovic, Francesco Mambrini, Marco Passarotti & Caroline Sporleder (eds.), *Workshop on Corpus-Based Research in the Humanities 2*, 95–104. Vienna: Technische Universität Wien.
- Irslinger, Britta. 2014a. The gender of abstract noun suffixes in the Brittonic languages. In Sergio Neri & Roland Schumann (eds.), *Studies on the collective and feminine in Indo-European from a diachronic and typological perspective*, 57–113. Leiden/Boston: Brill.
- Irslinger, Britta. 2014b. Intensifiers and reflexives in SAE, Insular Celtic and English. *Indogermanische Forschungen* 119. 159–206.
- Irslinger, Britta. 2017a. Medb ‘the Intoxicating One’? (Re-)constructing the past through etymology. In Mícheál Ó Mainnín & Gregory Toner (eds.), *Ulidia 4*, 38–94. Dublin: Four Courts Press.
- Irslinger, Britta. 2017b. Geographies of identity: Celtic philology and the search for origins in Ireland and Germany. In Thomas Mohnicke & Joachim Grage (eds.), *Geographies and imagination: Philological research on northern Europe in the 19th century*, 175–219. Cambridge: Cambridge Scholars.
- Irslinger, Britta. 2017c. Detransitive strategies in Middle Welsh: The preverbal marker *ym-*. In Erich Poppe, Karin Stüber & Paul Widmer (eds.), *Referential properties and their impact on the syntax of Insular Celtic languages*, 101–143. Münster: Nodus.
- Irslinger, Britta. 2019. More tales of two copulas: The copula systems of Western European languages from a typological and diachronic perspective. In Ronald I. Kim (ed.), *Diachronic perspectives on suppletion*, 27–76. Hamburg: Baar-Verlag.
- Isaac, Graham. 2003. Prospects in Old Irish Syntax. *Zeitschrift für celtische Philologie* 53. 181–197.
- Isaac, Graham. 2007. A new conjecture on the origins of absolute and conjunct flexion. *Ériu* 57. 49–60.
- Jakobson, Roman. 1923. *O češskom stixe: Preimuščestvenno v sopostavlenii s ruskim*. Berlin: OPoJaz-MLK.

- Jakobson, Roman. 1979. *Selected writings: 5. On verse, its masters and explorers*. The Hague: Mouton.
- Jasanoff, Jay. 1997. An Italo-Celtic isogloss: the 3pl. mediopassive in *-ntro. In Douglas Q. Adams (ed.), *Festschrift for Eric P. Hamp*, vol. 1, 146–161. Washington D.C.: Institute for the Study of Man.
- Jaynes, Edwin T. 1957a. Information theory and statistical mechanics, part I. *Physical Review* 106 (4). 620–630.
- Jaynes, Edwin T. 1957b. Information theory and statistical mechanics, part II. *Physical Review* 108 (2). 171–190.
- Jenner, Henry. 1904. *A handbook of the Cornish language*. London: David Nutt.
- Joyce, Sally L. & Evelyn S. Newlyn. 1999. Cornwall. In Sally L. Joyce & Evelyn S. Newlyn (eds.), *Records of early English drama: Dorset/Cornwall*, 369–610. Toronto: University of Toronto Press.
- Juola, Patrick. 2008. Authorship attribution. *Foundations and Trends in Information Retrieval* 1 (3), 233–334.
- Juola, Patrick & R. Harald Baayen. 2005. A controlled-corpus experiment in authorship identification by cross-entropy. *Literary and Linguistic Computing* 20. Supplement, 59–67.
- Jurafsky, Daniel & James H. Martin. 2009. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*, 2nd edn. Upper Saddle River, NJ: Pearson Prentice Hall/Pearson education international.
- Jurish, Bryan. 2010. Comparing canonicalizations of historical German text. In Jeffrey Heinz, Lynne Cahill, and Richard Wicentowski (eds.), *Association for Computational Linguistics Special Interest Group on Computational Morphology and Phonology (ACL-SIGMORPHON-2010) 11, Uppsala, 15 July, 2010*. 72–77. Association for Computational Linguistics. <https://www.aclweb.org/anthology/W10-2209.pdf>.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. Cambridge, MA: MIT Press.
- Kaufman, Leonard & Peter J. Rousseeuw. 1987. Clustering by means of medoids. In Yadolah Dodge (ed.), *Statistical data analysis: based on the L1-Norm and related methods*, 405–416. Amsterdam: Holland.
- Keenan, Edward L. 2002. Explaining the creation of reflexive pronouns in English. In Donka Minkova & Robert Stockwell (eds.), *Studies in the history of the English language. A millennial perspective*, 325–354. Berlin: Mouton de Gruyter.
- Kelly, Fergus. 1988. *A guide to Early Irish law* (Early Irish Law Series 3). Dublin: Dublin Institute for Advanced Studies.
- Kelly, Fergus. 1997. *Early Irish farming* (Early Irish Law Series 4). Dublin: Dublin Institute for Advanced Studies.
- Kelly, Patricia. 1999. Remarks on the omission of the particle *no* in Early Irish verbs. In Peter Anreiter & Erzsébet Jerem (eds.), *Studia Celtica et Indogermanica: Festschrift für Wolfgang Meid*, 157–161. Budapest: Archaeolingua.
- Kemmer, Suzanne. 1993. *The middle voice*. Amsterdam/Philadelphia: John Benjamins.
- Kille, Mark. 1995. “What thing is next I don’t quite know”: An analysis of variation in word order and subject-verb agreement in Middle Cornish. Cambridge, MA: Harvard University B.A. Honours thesis.
- Komen, Erwin. R. 2009. *CorpusStudio*. Nijmegen: Radboud University.

- Komen, Erwin. R. 2011. *Cesax: coreference editor for syntactically annotated XML corpora*. Nijmegen: Radboud University Nijmegen.
- König, Ekkehard. 2001. Intensifiers and reflexive pronouns. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft 20.1), 747–760. Berlin: de Gruyter.
- König, Ekkehard & Peter Siemund. 2000. The development of complex reflexives and intensifiers in English. *Diachronica* 17 (1). 39–84.
- Kortlandt, Frederik. 1979. The Old Irish absolute and conjunct endings and questions of relative chronology. *Ériu* 30. 35–53.
- Kuno, Susumu. 1976. Subject, theme, and the speaker's empathy: A reexamination of relativization phenomena. In Charles N. Li (ed.), *Subject and topic*, 417–444. New York: Academic Press.
- Lambert, Pierre-Yves. 1995. Le complément du comparatif de supériorité en vieil-irlandais. *Études celtiques* 31. 167–177.
- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press.
- Lange, Claudia. 2007. *Reflexivity and intensification in English: A study of texts and contexts*. Frankfurt am Main: Peter Lang Verlag.
- Lash, Elliott. 2014b. Subject positions in Old and Middle Irish. *Lingua* 148. 278–308.
- Lash, Elliott. 2017a. A quantitative analysis of e/i variation in Old Irish *etar* and *ceta*. *Ériu* 67. 139–166.
- Lash, Elliott. 2017b. Evaluating directionality in the internal reconstruction of pre-Old Irish copular clauses. *Indo-European Linguistics* 5. 77–129.
- Lash, Elliott & Aaron Griffith. 2018. Coordinate subjects, expletives, and the EPP in Early Irish. *Journal of Celtic Linguistics* 19 (1). 87–156.
- Le Mair, Esther. 2011. *Secondary verbs in Old Irish: A comparative-historical study of patterns of verbal derivation in the Old Irish glosses*. Galway: National University of Ireland, Galway dissertation. <http://hdl.handle.net/10379/3113>.
- Lewis, Henry. 1946. *Llawlyfr Cernyweg Canol*, 2nd edn. Caerdydd: Gwasg Prifysgol Cymru.
- Lhuyd, Edward. 1707. *Archaeologica Britannica, 1: Glossography*. Oxford.
- Lucht, Ina. 1994. Doppelte Markierung des Akkusativs beim Transitivum im Altirischen. *Zeitschrift für celtische Philologie* 46. 80–118.
- Lyon, Rod & John Pengilly. 1987. *Notes on spoken Cornish*. Redruth: Dyllansow Truran.
- Mac Cana, Proinsias. 1980. *The learned tales of medieval Ireland*. Dublin: Dublin Institute for Advanced Studies.
- Mac Cana, Proinsias. 1984. Varia: V. A note on the use of direct speech in Saltair na Rann. *Ériu* 35. 206–208.
- Mac Cárthaigh, Eoin. 2018. Research case studies: Bringing bardic poetry into the light. *Trinity College Dublin Provost & President's annual review 2017/18*. 28–31. <https://www.tcd.ie/provost/review/2018/annualreview.pdf>.
- Mac Coisdealbha, Pádraig. 1998. *The syntax of the sentence in Old Irish: Selected studies from a descriptive, historical and comparative point of view*, with additional notes and an extended bibliography by Graham R. Isaac. Tübingen: Niemeyer.
- Mac Eoin, Gearóid. 1982. The dating of Middle Irish texts. *Proceedings of the British Academy* 68. 109–137.

- Mac Gearailt, Uáitéar. 1993. The language of some late Middle Irish texts in the Book of Leinster. *Studia Hibernica* 26, 167–216.
- Mac Gearailt, Uáitéar. 1997–1998. Infixes and independent pronouns in the LL text of Táin Bó Cúailgne. *Zeitschrift für celtische Philologie* 49–50. 494–515.
- Mac Gearailt, Uáitéar. 2016. Issues in the transmission of Middle and early Modern Irish translation prose: Togail Troí and Scéla Alaxandair. In Axel Harlos & Neele Harlos (eds.), *Adapting texts and styles in a Celtic context: Interdisciplinary perspectives on processes of literary transfer in the middle ages: Studies in honour of Erich Poppe*, 103–134. Münster: Nodus Publikationen.
- MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. In Lucien M. Le Cam & Jerzy Neyman (eds.), *Berkeley Symposium on Mathematical Statistics and Probability 5, Berkeley, CA, 1965/66*, vol. 1, 281–297. Berkeley: University of California Press.
- Mahon, William J. 2006. Irish literature: 5. 19th century. In John T. Koch (ed.), *Celtic culture: A historical encyclopedia*, vol. 3, 1011–1014. Santa Barbara, CA: ABC-CLIO.
- Maitra, Ranjan & Ivan P. Ramler. 2010. A k-mean-directions algorithm for fast clustering of data on the sphere. *Journal of Computational and Graphical Statistics* 19 (2). 377–396.
- Manning, H. Paul. 2001. *Typology, history and “teratology”: The rise and fall of the “abnormal” main clause with relative structure in P-Celtic*. Chicago: University of Chicago dissertation.
- McCone Kim. 1978. The dative singular of Old Irish consonant stems. *Ériu* 29. 26–38.
- McCone, Kim. 1979. Pretonic preverbs and the absolute verbal endings in Old Irish. *Ériu* 30. 1–34.
- McCone, Kim. 1980. The nasalizing relative clause with object antecedent in the glosses. *Ériu* 31. 10–27.
- McCone, Kim. 1985. The Würzburg and Milan glosses: our earliest sources of ‘Middle Irish’. *Ériu* 36. 84–106.
- McCone, Kim. 1994. An tSean-Ghaeilge agus a réamhstair. In Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach (eds.), *Stair na Gaeilge in ómós do P[h]ádraig Ó Fiannachta*, 61–219. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- McCone, Kim. 1996. *Towards a relative chronology of ancient and medieval Celtic sound change* (Maynooth Studies in Celtic Linguistics 1). Maynooth: Department of Old Irish, National University of Ireland, Maynooth.
- McCone, Kim. 1997 [1987]. *The Early Irish verb* (Maynooth Monographs 1), 2nd edn, revised with index. Maynooth: An Sagart.
- McCone, Kim. 2006. *The origins and development of the Insular Celtic verbal complex* (Maynooth Studies in Celtic Linguistics 5). Maynooth: Department of Old Irish, National University of Ireland, Maynooth.
- McCone, Kim. 2015. Unstressed vowels and consonant quality in Old Irish: *u* or non-*u*? In Liam Breatnach, Ruairi Ó hUiginn, Damian McManus & Katherine Simms (eds.), *International Congress of Celtic Studies 14, Maynooth University, 2–6 August 2011*, 109–135. Dublin: Dublin Institute for Advanced Studies.
- McCone, Kim (author), Elisa Roma (trans.). 2005. *L’irlandese antico e la sua preistoria*, Alessandria: Edizioni dell’Orso.
- McManus, Damian. 1983. A chronology of the Latin loan-words in Early Irish. *Ériu* 34. 21–71.
- McManus, Damian. 1994. An Nua-Ghaeilge chlasaiceach. In Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach (eds.), *Stair na Gaeilge in ómós do P*

- [h]ádraig Ó Fiannachta, 335–445. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- Meelen, Marieke. 2016. *Why Jesus and Job spoke bad Welsh: The origin and distribution of V2 orders in Middle Welsh*. Utrecht: LOT dissertation series.
- Meelen, Marieke & Barend Beekhuizen. 2013. PoS-tagging and chunking historical Welsh. In Chris Yocum (ed.), *Scottish Celtic Colloquium 2012*, Edinburgh. https://www.academia.edu/7336114/PoS-tagging_and_chunking_historical_Welsh.
- Meelen, Marieke & Silva Nurmio. 2020. Adjectival Agreement in Middle and Early Modern Welsh Native and Translated Prose. *Journal of Celtic Linguistics* 21 (1). 1–28.
- Meelen, Marieke & David Willis. forthc. Creating PARSHCWL: Towards a Welsh historical treebank. *Journal of Linguistics*.
- Meelen, Marieke, Paul Russell, Sheila Watts & David Willis. 2017. *Building the Welsh historical treebank: Procedure for the parsed historical corpus of the Welsh Language (PARSHCWL)*. Poster presented at the Cambridge Language Sciences Workshop, 21 November 2017.
- Miller, D. Gary. 2012. *External influences on English: From its beginnings to the Renaissance*. Oxford: Oxford University Press.
- Mitkov, Ruslan (ed.). 2003. *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press.
- Morris Jones, John. 1913. *A Welsh grammar*. Oxford: Clarendon Press.
- Mosteller, Frederick & David L. Wallace. 1963. Inference in an authorship problem. *Journal of the American Statistical Association* 58 (302) 275–309.
- Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty & Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis (eds.), *Language Resources and Evaluation Conference (LREC-2016) 10, Portorož, Slovenia, 23–28 May 2016*. 1659–1666.
- Nyhan, Julianne. 2006. *The application of XML to the historical lexicography of Old, Middle, and early Modern Irish: A lexicon-based analysis*. Cork: University College Cork dissertation. <http://epu.ucc.ie/theses/jnyhan/Nyhanthesisultima.pdf>.
- Nyhan, Julianne. 2008. Developing integrated editions of minority language dictionaries: The Irish example. *Literary and Linguistic Computing* 23 (1). 3–12.
- Ó Cathasaigh, Tomás. 1990. On the Early-Irish prepositional relative without antecedent. *Celtica* 21. 418–426.
- Ó Coileáin, Seán. 1974. The structure of a literary cycle. *Ériu* 25. 88–125.
- Ó Cróinín, Dáibhí. 2001. The earliest Old Irish glosses. In Rolf Bergmann, Elvira Glaser & Claudine Moulin-Fankhänel (eds.), *Mittelalterliche volksprachige Glossen: Internationale Fachkonferenz des Zentrums für Mittelalterstudien der Otto-Friedrich-Universität Bamberg 2. bis 4. August 1999*, 7–31. Heidelberg: Winter.
- Ó Cualaoich, Conchubhar. 1999. *Some irregular syncope patterns in Old Irish*. Maynooth: National University of Ireland, Maynooth dissertation.
- Ó Háinle, Cathal. 2006. Irish literature: 4. Post-classical. In John T. Koch (ed.), *Celtic culture: A historical encyclopedia*, vol. 3, 1005–1111. Santa Barbara, CA: ABC-CLIO.
- Ó hAnluain, Liam A. 1999 [1960]. *Graiméar Gaeilge na mBráithre Críostaí*. Baile Átha Cliath: An Gúm.
- Ó hUiginn, Ruairí. 1983. On the Old Irish *figura etymologica*. *Ériu* 34. 123–133.

- Ó hUiginn, Ruairí. 1986. The Old Irish nasalizing relative clause. *Ériu* 37. 33–87.
- Ó hUiginn, Ruairí. 1998. Complementation in Early Irish: The *uerba dicendi*. *Ériu* 49. 121–148.
- Ó hUiginn, Ruairí. 2013. Transmitting the text: some linguistic issues in the work of the Franciscans. In Raymond Gillespie & Ruairí Ó hUiginn (eds.), *Irish Europe: 1600–1650*, 85–104. Dublin: Four Courts Press.
- Ó Lochlainn, Colm. 1941–1942. Poets on the Battle of Clontarf, part I. *Éigse* 3. 208–218.
- Ó Lochlainn, Colm. 1943–1944. Poets on the Battle of Clontarf, part II. *Éigse* 4. 33–45.
- Ó Maolalaigh, Roibeard. 1995–1996. The development of eclipsis in Gaelic. *Scottish Language* 14–15: *International Conference on the Languages of Scotland and Ulster* 4. 158–173.
- Ó Maolalaigh, Roibeard. 2003. Processes in nasalisation and related issues. *Ériu* 53. 109–132.
- Ó Maolalaigh, Roibeard. 2008. The Scotticisation of Gaelic: A reassessment of the language and orthography of the Gaelic notes in the Book of Deer. In Katherine Forsyth (ed.), *Studies in the Book of Deer*, 179–274. Dublin: Four Courts Press.
- Ó Maolalaigh, Roibeard. 2016. The mutational effects of the preposition *ós*: *Bile ós chrannaibh* and related matters. *Scottish Gaelic Studies* 30. 73–111.
- Ó Muircheartaigh, Peadar. 2015. *Gaelic dialects present and past: a study of modern and medieval dialect relationships in the Gaelic languages*. Edinburgh: University of Edinburgh dissertation.
- Ó Nolan, Gerald. 1920. *Studies in Modern Irish*, part I. Dublin: The Educational Company of Ireland.
- O’Leary, Aileen M. 1999. The identities of the poet(s) Mac Cisi: A reinvestigation. *Cambrian Medieval Celtic Studies* 38. 53–71.
- Parina, Elena. 2007. Reflexivpronomina im Mittelkymrischen. In Helmut Birkhan (ed.), *Kelten-Einfälle an der Donau: Akten des Vierten Symposiums Deutschsprachiger Keltologinnen und Keltologen. Linz/Donau, 17.–21. Juli 2005*, 389–396. Vienna: Verlag der Österreichischen Akademie der Wissenschaft.
- Parina, Elena & Erich Poppe. forthcoming. *Adjectival agreement in the Llyfr Ancr*.
- Park, Hae-Sang & Chi-Hyuck Jun. 2009. A simple and fast algorithm for K-medoids clustering. *Expert Systems with Applications* 36 (2). 3336–3341.
- Passarotti, Marco. 2007. Verso il Lessico Tomistico Biculturale. La treebank dell’Index Thomisticus. In Raffaella Petrilli & Diego Femia (eds.), *Il filo del discorso: Intrecci testuali, articolazioni linguistiche, composizioni logiche: Società di Filosofia del Linguaggio* 13, Viterbo, 14–16 Settembre 2006, 187–205. Rome: Aracne Editrice.
- Pedersen, Holger. 1899. Die aspiration im Irischen, zweiter Theil. *Zeitschrift für vergleichende Sprachforschung* 35. 315–444.
- Pedersen, Holger. 1909–1913. *Vergleichende Grammatik der keltischen Sprachen*, 2 vols. Göttingen: Vandenhoeck & Ruprecht.
- Peitsara, Kirsti. 1997. The development of reflexive strategies in English. In Matti Rissanen, Merja Kytö & Kirsi Heikkonen (eds.), *Grammaticalization at work: Studies of long-term developments in English*, 277–370. Berlin/New York: Mouton de Gruyter.
- Pesetsky, David. 1987. *Wh*-in-situ: Movement and unselective binding. In Eric J. Reuland & Alice G. B. ter Meulen (eds.), *The representation of (in)definiteness*, 98–129. Cambridge, MA: MIT Press.
- Piotrowski, Michael. 2012. *Natural language processing for historical texts* (Synthesis Lectures on Human Language Technologies 5.2). San Rafael: Morgan & Claypool Publishers.

- Poletto, Cecilia. 2002. The left periphery of V2-Rhaetoromance dialects: A new view on V2 and V3. In Sjeff Barbiers, Leonie Cornips & Suzanne van der Kleij (eds.), *Syntactic microvariation*, 214–242. Amsterdam: Meertens Instituut.
- Poppe, Erich. 1991. Word order in Cyfranc Lludd a Llefelys: Notes on the pragmatics of constituent-ordering in MW narrative prose. In James Fife & Erich Poppe (eds.), *Studies in Brythonic word order*, 155–204. Amsterdam: John Benjamins.
- Poppe, Erich. 1999. Reconstructing medieval Irish literary theory: The lesson of Airec Menman Uraird maic Coise. *Cambrian Medieval Celtic Studies* 37. 33–54.
- Poppe, Erich. 2008. *Of cycles and other critical matters: Some issues in medieval Irish literary history and criticism* (E. C. Quiggin Memorial Lectures 9). Cambridge: Department of Anglo-Saxon, Norse, and Celtic.
- Poppe, Erich. 2009. Standard Average European and the Celticity of English intensifiers and reflexives: Some considerations and implications. *English Language and Linguistics* 13 (2). 251–266.
- Preusler, Walther. 1938. Keltischer Einfluss im Englischen. *Indogermanische Forschungen* 56. 178–191.
- Puddu, Nicoletta. 2005. *Riflessivi e intensificatori: Greco, latino e le altre lingue indoeuropee*. Pisa: Ed. ETS.
- Qiu, Fangzhe. 2017. The Ulster Cycle in the law tracts. In Mícheál Ó Mainnín & Gregory Toner (eds.), *Ulidia* 4, 9–22. Dublin: Four Courts Press.
- Qiu, Fangzhe. 2019. Old Irish *aue* ‘descendant’ and its descendants. *Indogermanische Forschungen* 124. 343–374.
- Qiu, Fangzhe, David Stifter, Bernhard Bauer, Elliott Lash & Tianbo Ji. 2018. Chronologicon Hibernicum: A probabilistic chronological framework for dating Early Irish language developments and literature. In Marinos Ioannides, Eleanor Fink, Raffaella Brumana, Petros Patias, Anastasios Doulamis, João Martins & Manolis Wallace (eds.), *Digital heritage, progress in cultural heritage: Documentation, preservation, and protection* (LNCS 11196), 731–740. Springer: Cham, Switzerland.
- Quin, Ernest Gordon. 1960. Old-Irish *ol* ‘inquit’. *Celtica* 5. 95–102.
- Quin, Ernest Gordon. 1979. Nasalization of *g- d- b-* in Early Irish. *Studia Celtica* 14–15. 255–259.
- Randall, Beth, Ann Taylor & Anthony Kroch. 2005. *CorpusSearch 2*. Philadelphia: University of Pennsylvania.
- Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27. 53–93.
- Ringe, Don & Joseph Eska. 2012. *Historical linguistics: Toward a twenty-first century reintegration*. Cambridge: Cambridge University Press.
- Rissanen, Matti. 1989. Three problems connected with the use of diachronic corpora. *ICAME Journal* 13. 16–19.
- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: Handbook in generative syntax*, 281–337. Dordrecht: Kluwer.
- Rizzi, Luigi. 2004. On the cartography of syntactic structures. In Luigi Rizzi (ed.), *The structure of CP and IP* (Cartography of Syntactic Structures 2), 3–15. Oxford: Oxford University Press.
- Rizzi, Luigi. 2013. Notes on cartography and further explanation. *Probus* 25 (1). 197–226.
- Rodway, Simon. 2013. *Dating medieval Welsh literature: Evidence from the verbal system*. Aberystwyth: Cambrian Medieval Celtic Studies (CMCS) Publications.

- Roma, Elisa, 2000a. *Da dove viene e dove va la morfologia: Forme sintetiche e forme analitiche nella storia del verbo irlandese*. Milan: Franco Angeli.
- Roma, Elisa. 2000b. How subject pronouns spread in Irish: A diachronic study and synchronic account of the third person + pronoun pattern. *Ériu* 51. 107–157.
- Roma, Elisa. 2014. Old Irish noun phrases: Data from the Milan Glosses and a hypothesis for the origin of the single article constraint. In Elisa Roma & David Stifter (eds.), *Linguistic and philological studies in Early Irish*, 131–176. Lewiston, NY: Edwin Mellen Press.
- Roma, Elisa. 2018a. Nasalization after inflected nominals in the Old Irish glosses: A reassessment. *Journal of Celtic Linguistics* 19. 1–30.
- Roma, Elisa. 2018b. Old Irish pronominal objects and their use in verbal pro-forms. In Raimund Karl & Katharina Möller (eds.), *European symposium in Celtic studies 2, Bangor, Wales, 31 August – 3 July 2017*, 7–20. Hagen: Curach Bhán.
- Roma, Elisa & David Stifter (eds.). 2014. *Linguistic and philological studies in Early Irish*. Lewiston, NY: Edwin Mellen Press.
- Rossiter, Trudy. 2004. *Verbal composition in Old Irish with special reference to multi-preverb compounds*. Maynooth: National University of Ireland, Maynooth dissertation.
- Rousseeuw, Peter J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20. 53–65.
- Russell, Paul. 1988. The Celtic preverb *uss and related matters. *Ériu* 39. 95–126.
- Russell, Paul. 2005. “What was best of every language”: The early history of the Irish language. In Dáibhí Ó Cróinín (ed.), *A new history of Ireland, vol. 1: Prehistoric and early Ireland*, 405–450. Oxford: Oxford University Press.
- Scannell, Kevin. 2009. Standardization of corpus texts for the new English-Irish dictionary. Paper presented at the 15th annual NAACL conference, New York, NY, 22 May 2009. <https://cs.slu.edu/~scannell/pub/naactl09.pdf>.
- Scannell, Kevin. 2017. *Caighdeánú na Gaeilge*. Paper presented at the American Conference for Irish Studies (ACIS-2017), Kansas City, MO, 29 March – 1 April 2017. <https://cs.slu.edu/~scannell/pub/acis17-ga.pdf>.
- Schafer, Robin. 1995. Negation and verb second in Breton. *Natural Language and Linguistic Theory* 13. 135–172.
- Schlüter, Dagmar. 2010. *History or fable? The Book of Leinster as a document of cultural memory in twelfth-century Ireland*. Münster: Nodus Publikationen.
- Schrijver, Peter. 1994. The Celtic adverbs for “against” and “with” and the early apocope of *-i. *Ériu* 45. 151–189.
- Schrijver, Peter. 1997a. On the nature and origin of word-initial h- in the Würzburg Glosses. *Ériu* 48. 205–227.
- Schrijver, Peter. 1997b. *Studies in the history of the Celtic pronouns and particles* (Maynooth Studies in Celtic Linguistics 2). Maynooth: Department of Old Irish, National University of Ireland, Maynooth.
- Schrijver, Peter. 2007. Some common developments of Continental and Insular Celtic. In Pierre-Yves Lambert & Georges-Jean Pinault (eds.), *Gaulois et Celtique continentale*, 354–371. Genève: Droz.
- Schrijver, Peter. 2011. Old British. In Elmar Ternes (ed.), *Brythonic Celtic-Britannisches Keltisch: From medieval British to modern Breton*, 1–84. Bremen: Hemen.
- Schumacher, Stefan (author), Britta Schulze-Thulin & Caroline aan de Weil (eds.). 2004. *Die keltischen Primärverben: Ein vergleichendes etymologisches und morphologisches*

- Lexikon* (Innsbrucker Beiträge zur Sprachwissenschaft 110). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.
- Schumacher, Stefan. 2012. Mittelmährisch *tra*, *tros*, *traws*, *traw*, altirisch *trá* und verwandtes. In Velizar Sadovski & David Stifter (eds.), *Iranistische und indogermanistische Beiträge in memoriam Jochem Schindler (1944–1994)*, 361–375. Vienna: Österreichischen Akademie der Wissenschaften.
- Scowcroft, R. Mark. 1987. Leabhar Gabhála, part I: The growth of the text. *Ériu* 38. 81–142.
- Siemund, Peter. 2010. Grammaticalization, lexicalization and intensification: English “itself” as a marker of middle situation types. *Linguistics* 48 (4). 797–836.
- Siemund, Peter. 2014. The emergence of English reflexive verbs: An analysis based on the Oxford English Dictionary. *English Language and Linguistics* 18. 49–73.
- Sims-Williams, Patrick. 1984. The double system of verbal inflexion in Old Irish. *Transactions of the Philological Society* 82. 138–201.
- Singhal, Amit. 2001. Modern information retrieval: A brief overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24 (4). 35–43.
- Skjærholt, Arne. 2011. More, faster: Accelerated corpus annotation with statistical taggers. *Journal for Language Technology and Computational Linguistics* 26 (2). 153–165.
- Solka, Jeffrey L. 2008. Text data mining: Theory and methods. *Statistics Surveys* 2. 94–112.
- Sommer, Ferdinand. 1897. Das Pronomen Personale Infixum im Altirischen. *Zeitschrift für celtische Philologie* 1. 177–231.
- Spencer, Andrew. 1998. Morphophonological operations. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 123–143. Oxford: Blackwell.
- Stamatatos, Efstathios. 2009. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology* 60 (3). 538–556.
- Stéfanini, Jean. 1962. *La voix pronominale en ancien et en moyen français*. Gap: Ophrys.
- Stifter, David. 2006. *Sengoidélc: Old Irish for beginners*. Syracuse: Syracuse University Press.
- Stifter, David. 2009. Early Irish. In Martin J. Ball & Nicole Müller (eds.), *The Celtic languages* (Routledge Language Family Descriptions), 2nd edn, 55–116. London: Routledge.
- Stifter, David. 2013. The history of the Old Irish preverb *to-*. In Elisa Roma & David Stifter (eds.), *Linguistic and philological studies in Early Irish*, 203–246. Lewiston, NY: Edwin Mellen Press.
- Stifter, David. 2015. The language of the poems of Blathmac. In Pádraig Ó Riain (ed.), *The poems of Blathmac son of Cú Brettan: Reassessments* (Irish Texts Society Subsidiary Series 27), 47–103. London: Irish Texts Society.
- Strachan, John. 1899. The substantive verb in the Old Irish glosses. *Transactions of the Philological Society* 24 (1). 1–82.
- Strachan, John. 1903a. On the language of the Milan Glosses. *Zeitschrift für celtische Philologie* 4. 48–71.
- Strachan, John. 1903b. On the language of the St. Gall Glosses. *Zeitschrift für celtische Philologie* 4. 470–492.
- Strachan, John. 1904. The infixed pronoun in Middle Irish. *Ériu* 1. 153–179.
- Strachan, John. 1929. *Old-Irish paradigms and selections from the Old-Irish glosses*, 3th edn, revised by Osborn Bergin. Dublin: Royal Irish Academy.
- Strawson, Peter. F. 1964. Identifying reference and truth values. *Theoria* 30. 96–118.
- Stüber, Karin. 2010–2012. Die Figura etymologica des Altirischen im sprachgeschichtlichen Vergleich. *Keltische Forschungen* 5. 229–260.

- Sukhareva, Maria, Francesco Fuscagni, Johannes Daxenberger, Susanne Görke, Doris Prechel & Iryna Gurevych. 2017. Distantly supervised POS tagging of low-resource languages under extreme data sparsity: The case of Hittite. In Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Feldman, Anna Kazantseva, Nils Reiter & Stan Szpakowicz (eds.), *Association for Computational Linguistics: Joint SIGHUM Workshop on computational linguistics for cultural heritage, social sciences, humanities and literature (LaTeCH-CLFL-2017) 1, Vancouver, BC, 4 August, 2017*. 95–104.
- Svenonius, Peter. 2007. Interpreting uninterpretable features. *Linguistic Analysis* 33. 375–413.
- Tan, Pang-Ning, Michael Steinbach & Vipin Kumar. 2005. *Introduction to data mining*. Boston: Addison-Wesley.
- TEI Consortium. 2009. *TEI P5: Guidelines for electronic text encoding and interchange, version 1.3*. <http://www.tei-c.org/Guidelines/P5/>.
- Theuerkauf, Marie-Luise. 2017. The death of Boand and the recensions of Dindsénchas Érenn. *Ériu* 67. 49–97.
- Thomas, Peter W. 1996. *Gramadeg y Gymraeg*. Caerdydd: Gwasg Prifysgol Cymru.
- Thomson, Derick S. 1970. The poetry of Niall MacMhuirich. *Transactions of the Gaelic Society of Inverness* 46. 281–307.
- Thurneysen, Rudolf. 1905. Miscellen zur altirischen Grammatik: I. Die Nasalisierung des Anlauts nach deklinierten Wörtern im Altirischen. *Zeitschrift für celtische Philologie* 5. 1–19.
- Thurneysen, Rudolf. 1909. *Handbuch des Alt-Irischen, I. Teil: Grammatik*. Heidelberg: Winter Verlag.
- Thurneysen, Rudolf. 1912. *Zu irischen handschriften und litteraturdenkmälern* (Abhandlungen der Akademie der Wissenschaften in Göttingen, Philologisch-Historische Klasse). Berlin: Weidmannsche buchhandlung.
- Thurneysen, Rudolf. 1940. Irisches, 3. *frisdo-gair*. *Zeitschrift für celtische Philologie* 22. 27–29.
- Thurneysen, Rudolf (author), Daniel A. Binchy & Osborn Bergin (trans.). 1946. *A Grammar of Old Irish*, revised and enlarged edition, reprinted 1975, 1993, with supplement. Dublin: Dublin Institute for Advanced Studies.
- Tristram, Hildegard L. C. 1999. *How Celtic is standard English?* St Petersburg: Nauka.
- Uhlich, Jürgen. 1993. *Die Morphologie der komponierten Personennamen des Altirischen*. Beiträge zu Sprachwissenschaften 1. Witterschlick/ Bonn: M. Wehle.
- Uhlich, Jürgen. 1995. On the fate of intervocalic *-w- in Old Irish, especially between neutral vowels. *Ériu* 46. 11–48.
- Uhlich, Jürgen. 2009–2010. Altirisch *arae* “Wagenlenker”, *aithesc* “Antwort”, keltische Präverbien auf *-i und die frühe Apokope von *-i. *Zeitschrift für celtische Philologie* 57. 141–160.
- Uhlich, Jürgen. 2013. Zum Artikelgebrauch beim Bezugswort eines Relativsatzes im frühen Irischen. In Dónall Ó Baoill, Donncha Ó hAodha & Nollaig Ó Muraíle (eds.), *Saltair saíochta, sanasaíochta agus seanchais: A Festschrift for Gearóid Mac Eoin*, 429–462. Dublin: Four Courts Press.
- Uhlich, Jürgen. 2018. The poems of Blathmac. *Cambrian Medieval Celtic Studies* 75. 53–77.
- Uí Dhonnchadha, Elaine & Joseph van Genabith. 2006. A part-of-speech tagger for Irish using finite state morphology and constraint grammar disambiguation. *Language Resources*

- and Evaluation Conference (LREC-2006) 5, Genoa, 22–28 May 2006. 2241–2244.
http://www.lrec-conf.org/proceedings/lrec2006/pdf/193_pdf.pdf.
- Uí Dhoonchadha, Elaine, Kevin Scannell, Ruairí Ó hUiginn, Eilís Ní Mhearraí, Máire Nic Mhaoláin, Brian Ó Raghallaigh, Gregory Toner, Séamus Mac Mathúna, Déirdre D’Auria, Eithne Ní Ghallchobhair & Niall O’Leary. 2014. *Corpas na Gaeilge (1882–1926): Integrating historical and Modern Irish texts. Language Resources and Evaluation Conference (LREC-2014) 9: Language resources and technologies for processing and linking historical documents and archives Workshop (LRT4HDA): Deploying linked open data in cultural heritage, Reykjavik, 26 May 2014.* 12–18.
<http://www.lrec-conf.org/proceedings/lrec2014/workshops/LREC2014Workshop-LRT4HDA%20Proceedings.pdf>.
- Väänänen, Veikko. 1981. *Introduction au latin vulgaire*, 3rd edn. Paris: Klincksieck.
- Vendryes, Joseph. 1927. Les verbes composés avec *ym-* dans les Mabinogion. *Annales de Bretagne: Mélanges bretons et celtiques offerts à M. J. Loth, Volume hors-série.* 49–62.
- Vendryes, Joseph. 1928. Sur les adverbes de manière du type *v.-irl.* in *biucc*, gall. *yn fychan*. *Zeitschrift für celtische Philologie* 17. 73–78.
- Vennemann, Theo. 2013. Concerning *myself*. In Robert Mailhammer (ed.), *Lexical and structural etymology: Beyond word histories*, 121–146. Boston/Berlin: Mouton De Gruyter.
- Vezzosi, Letizia. 2005. The development of *himself* in Middle English: A “Celtic” hypothesis. In Nikolaus Ritt & Herbert Schendl (eds.), *Rethinking Middle English: Linguistic and literary approaches*, 228–243. Frankfurt am Main: Peter Lang Verlag.
- Walkden, George. 2015. Verb-third in early West Germanic: A comparative perspective. In Theresa Biberauer & George Walkden (eds.), *Syntax over time: Lexical, morphological, and information-structural interactions*, 236–248. Oxford: Oxford University Press.
- Watkins, Calvert. 1962. *Indo-European Origins of the Celtic verb: 1. The sigmatic aorist*. Dublin: Dublin Institute for Advanced Studies.
- Watkins, Calvert. 1963. Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb. *Celtica* 6. 1–49.
- Watkins, Calvert. 1995. *How to kill a dragon: Aspects of Indo-European poetics*. New York: Oxford University Press.
- Watkins, Thomas Arwyn. 1987. Constituent order in the Old Welsh verbal sentence. *Bulletin of the Board of Celtic Studies* 34. 51–60.
- Williams, Nicholas. 1994. Na canúintí a theacht chun solais. In Kim McCone, Damian McManus, Cathal Ó Háinle, Nicholas Williams & Liam Breatnach (eds.), *Stair na Gaeilge in ómós do P[h]ádraig Ó Fiannachta*, 447–478. Maigh Nuad: Roinn na Sean-Ghaeilge, Coláiste Phádraig.
- Williams, Nicholas. 2006. *Cornish today: An examination of the revived language*, 3rd edn. Cathair na Mart: Evertype.
- Williams, Nicholas. 2011. Middle and Late Cornish. In Elmar Ternes (ed.), *Brythonic Celtic-Britannisches Keltisch: From medieval British to modern Breton*, 237–357. Bremen: Hemen.
- Williams, Nicholas. 2016. *Studies in traditional Cornish*. Portlaoise: Evertype.
- Willis, David. 1998. *Syntactic change in Welsh: A study of the loss of the verb-second*. Oxford: Clarendon Press.
- Willis, David. 2009. Old and Middle Welsh. In Martin J. Ball & Nicole Müller (eds.), *The Celtic languages* (Routledge Language Family Descriptions), 2nd edn, 117–160. London: Routledge.

- Winford, Donald. 2003. *An introduction to contact linguistics*. Oxford: Blackwell Publishing.
- Wmffre, Iwan. 1998. *Late Cornish*. München: LINCOM-EUROPA.
- Wodtko, Dagmar S., Britta Irslinger & Carolin Schneider. 2008. *Nomina im indogermanischen Lexikon*. Heidelberg: Universitätsverlag Winter.
- Wolfe, Sam. 2016. On the left periphery of V2 languages: Evidence from Romance Fin and Force V2 system. *Rivista di grammatica generativa* 38. 287–310.
- Wolfe, Sam. 2018. Syntactic variation in two sister languages: A study of word order in Old French and Old Occitan. In Gabriela Pană Dindelegan, Adina Dragomirescu, Irina Nicula & Alexandru Nicolae (eds.), *Comparative and diachronic perspectives on Romance syntax*, 53–83. Cambridge: Cambridge Scholars Publishing.
- Young, Robert W. & William Morgan. 1980. *The Navajo language: A grammar and colloquial dictionary*. Albuquerque, NM: University of New Mexico Press.
- Zair, Nicholas. 2012. *The reflexes of the Proto-Indo-European laryngeals in Celtic*. Leiden: Brill.
- Zavrel, Jakob & Walter Daelemans. 1997. Memory-based learning: Using similarity for smoothing. In Philip R. Cohen & Wolfgang Wahlster (eds.), *Association for Computational Linguistics (ACL/EACL-'97) 35, Madrid, 7–12 July 1997*, 436–443. Stroudsburg, PA: Association for Computational Linguistics.
- Zeman, Daniel, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinkova, Jan Hajič jr., Jaroslava Hlavacova, Václava Kettnerová, Zdenka Uresova, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droганova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonça, Tatiana Lando, Rattima Nitisaroj & Josie Li. 2017. CoNLL 2017 shared task: Multilingual Parsing from raw text to universal dependencies. In Jan Hajič and Dan Zeman (eds.), *Conference on Natural Language Learning (CoNLL 2017) shared task: Multilingual parsing from raw text to universal dependencies*, 1–19.
- Zhong, Shi & Joydeep Ghosh. 2005. Generative model-based document clustering: a comparative study. *Knowledge and Information Systems* 8 (3). 374–384.

Index

Language Index

- Armenian 17
- Baltic 273
- Breton
- Middle 315–8
 - Modern 308, 325, 328–9
 - Old 313
- British Celtic, Brittonic 240, 313
- Old Southwest Brittonic 314–5
- Cornish 313–37
- Late 327–32, 335–7
 - Middle 315–8, 320–9, 331–4
 - Old 313
- English
- Middle 15, 33, 275
 - Modern 15, 22, 29, 33–4, 63, 90, 269–71, 273–6, 310, 312, 327–32, 335–6
 - Old 6, 17–8, 33, 270–1, 278, 283–4, 295, 321
- French
- Modern 269, 274–5, 292, 328
 - Old (incl. Old Occitan) 17–8, 289, 302, 306, 308, 312, 319–20
- Gaulish 240
- German (incl. Old High German) 269–70, 273, 275–7, 292, 321
- Germanic (incl. Gothic, Norse, Saxon) 17, 270, 273–4
- Greek 15, 17, 25, 291, 303
- Hittite 26
- Icelandic 15, 33
- Insular Celtic 163, 239, 242, 265, 269
- Irish
- Classical Modern 49–51, 61–3, 115
 - Middle 50–1, 54, 60–2, 74, 76, 78, 81–3, 85–104, 115, 117–130, 132–141, 258
 - Modern 50–1, 61–7, 80–3, 189–93, 224, 269
 - Old 49–83, 101, 115–32, 136, 138–9, 141, 143–177, 179–193, 195–237, 239–69, 322
- Italian 269, 274
- Latin 15, 17, 19, 25, 37, 51, 173–4, 187, 190–1, 199–200, 202–4, 215, 230, 250, 256, 259, 274, 302–5, 307–8, 312
- Navajo 40
- Portuguese 15, 17–8
- Proto-Indo-European 70, 74, 162–4, 273
- Romance 273–4, 319
- Russian 17–8, 22
- Sanskrit 26
- Scottish Gaelic 189–93
- Slavic (incl. Old Church Slavonic) 17–8, 25, 273
- Spanish 17–8
- Surselvan 274
- Swedish 25
- Welsh
- Middle 27–47, 240–1, 269–312, 320–1
 - Modern 30, 42, 206, 269–71, 274–5, 295–6, 315–7, 321, 326, 335
 - Old 270–1, 313–4

Index of Examples

This index contains the references to all primary quotations, except those listed in the Appendix to Aaron Griffith's article (266–8)

Irish

Adomn. 2nd Vision §21 (Stokes 1891) 119

Airec Menman Uraird Meic Coise (Byrne 1908: 70.14) 136

Aislinge Meic Conglinne (Meyer 1892: 129.1) 140

Annals of Four Masters 899 (O'Donovan 1856) 138

Bechbretha §55 (Charles-Edwards & Kelly 1983) 129

Blathm. 20 126

Blathm. 41 126

Blathm. 52 118

Blathm. 108 126

Blathm. 117 119–20

Blathm. 122 121

Blathm. 127 125

Blathm. 128 120

Blathm. 140 126

Blathm. 155 136

Blathm. 159 125

Blathm. 187 125

Blathm. 190 130

Blathm. 208 126

Blathm. 225 127

Blathm. 233 125

Blathm. 237 125

Blathm. 245 118

Book of Armagh, Add. (Thes. 2: 240.1) 188

Book of Armagh, Add. (Thes. 2: 240.20) 187

Book of Armagh, Add. (Thes. 2: 242.20–1) 187

Book of Deer 189

Bretha Déin Chécht §37 (Binchy 1966) 129

Cambrai Homily (Thes. 2: 244.27–8) 168

Cambrai Homily (Thes. 2: 244.31) 168

Cambrai Homily (Thes. 2: 244.33) 168

Cambrai Homily (Thes. 2: 245.12) 168

Cambrai Homily (Thes. 2: 245.14) 168

Cambrai Homily (Thes. 2: 245.33) 168

Cambrai Homily (Thes. 2: 246.5–6) 168

Cambrai Homily (Thes. 2: 246.14) 167

Cambrai Homily (Thes. 2: 247.3) 168

Cambrai Homily (Thes. 2: 247.8) 210

Cambrai Homily (Thes. 2: 247.18, 39) 237

Carlsruhe Augustine 10^a2 213

Carlsruhe Augustine 12^d1 210

Carlsruhe Bede 18^b10 228

Carlsruhe Bede 32^a8 118

Carlsruhe Bede 32^c8 125

Carlsruhe Bede 33^d11 241–2

Carlsruhe Priscian 2^a3 252

Christian Brothers' Grammar (Ó hAnluain 1960 §11.36) 224

CIH 43.10 132

CIH 538.19 (Senchas Már) 131–2

CIH 885.5 128

CIH 920.32 123

CIH 1794.15 135

CIH 1833.30 135

Coibnes Uisci Thairidni §15 (Binchy 1955) 129

Córus Bésgnai §30 (Breatnach 2017a) 129

Críth Gablach 248 (Binchy 1941) 131

Críth Gablach 280 (Binchy 1941) 131

Críth Gablach 350 (Binchy 1941) 131

Críth Gablach 448 (Binchy 1941) 131

Críth Gablach 459 (Binchy 1941) 131

Críth Gablach 593 (Binchy 1941) 131

Fél. Prol. 25–8 116

Fél. June 20 116

Fél. Aug 2 116

Fél. Oct 12 116

Fél. Oct 16 116

Fél. Oct 20 116

Fél. Epil. 313 117

- Fél. Epil. 411–2 116
Fo réir Choluimb céin ad-fias (Kelly 1973 §12) 322
- Gein Branduib maic Echach* §9 (Meyer 1899) 132
- Irish Gospel of Thomas* 18 127
Irish Gospel of Thomas 22 131
Irish Gospel of Thomas 33 126
Irish Gospel of Thomas 39 136
Irish Gospel of Thomas 44 127
Irish Penitential (Gwynn 1914: 154–5 §1e) 126
Irish Penitential (Gwynn 1914: 160–1 §23) 127
Irish Penitential (Gwynn 1914: 166.7) 126
Irish Penitential (Gwynn 1914: 166.13) 118
- LL 3565–8 (Met. Dinds.) 139
 LL 3573–6 (Met. Dinds.) 139
 LL 4816–9 128
 LL 8367 (Táin Bó Cúailnge) 134
 LL 8796 (Táin Bó Cúailnge) 134
 LL 9081 (Táin Bó Cúailnge) 139
 LL 12066 (Táin Bó Cúailnge) 123
 LL 12726 130
 LL 19708 (Met. Dinds.) 139
 LL 20690–1 (Met. Dinds.) 133
 LL 20219 (Met. Dinds.) 120
 LL 21057–8 (Prose Dinds.) 122
 LL 21149 (Prose Dinds.) 120
 LL 21473 (Met. Dinds.) 124
 LL 22899 (Cath Ruis na Ríg) 123
 LL 25163 (Met. Dinds.) 135
 LL 25209 (Met. Dinds.) 135
 LL 25233 (Met. Dinds.) 120
 LL 25401–2 (Met. Dinds.) 134
 LL 25708 (Met. Dinds.) 140
 LL 25712 (Met. Dinds.) 122
 LL 26022 (Met. Dinds.) 133
 LL 26094 (Met. Dinds.) 133
 LL 26633 (Met. Dinds.) 129
 LL 26846 (Met. Dinds.) 122
 LL 26995 (Met. Dinds.) 122
 LL 27259–62 (Met. Dinds.) 123–4
- LL 27297 (Met. Dinds.) 130
 LL 27637 (Met. Dinds.) 130
 LL 27775 (Met. Dinds.) 122
 LL 27837 (Met. Dinds.) 122
 LL 28898 132
 LL 29020 (Met. Dinds.) 132
 LL 29245 122
 LL 29807 (Met. Dinds.) 134
 LL 29927 (Met. Dinds.) 124
 LL 30203 (Met. Dinds.) 140
 LL 30502 (Met. Dinds.) 133
 LL 37651 (Bórama) 135
 LU 1632 (Táin Bó Flidais) 141
 LU 1714 (Máel Dúin) 125
 LU 1731 (Máel Dúin) 128
 LU 1757 (Máel Dúin) 126
 LU 1846 (Máel Dúin) 118
 LU 1930 (Máel Dúin) 127
 LU 2562 (Scéla na Esérgi) 133
- Meyer 1903 §6 140
 Meyer 1912 v. 23 134
 Ml. 16^{a7} 258–9, 262
 Ml. 17^{b2} 221
 Ml. 17^{b26} 229
 Ml. 17^{c3} 260–2
 Ml. 18^{a7} 156
 Ml. 19^{b11} 227, 231–2
 Ml. 20^{b4} 122
 Ml. 20^{d11} 158
 Ml. 21^{b2} 256
 Ml. 21^{b10} 156
 Ml. 21^{d4} 235
 Ml. 22^{c3} 159
 Ml. 22^{c14} 206, 227, 232
 Ml. 22^{d7} 247
 Ml. 23^{a5} 187
 Ml. 23^{a12} 257, 262
 Ml. 23^{d21} 156
 Ml. 24^{a15} 156
 Ml. 24^{a18} 262
 Ml. 24^{c4} 125
 Ml. 24^{d17} 262
 Ml. 25^{c16} 229
 Ml. 25^{d18} 198
 Ml. 26^{b8} 258

- Ml. 26^c4 234–5
 Ml. 27^d22 258
 Ml. 28^a10 205
 Ml. 29^c13 247
 Ml. 29^d3 156
 Ml. 30^a3 202, 224
 Ml. 30^b2 220
 Ml. 30^b10 187
 Ml. 31^b16 262
 Ml. 31^b22 221
 Ml. 31^b25 258
 Ml. 31^c14 158, 173, 229
 Ml. 32^a5 169
 Ml. 32^d1 203
 Ml. 32^d5 196
 Ml. 32^d27 158
 Ml. 33^a1 158
 Ml. 33^a19 219
 Ml. 34^b6 258
 Ml. 35^b5 215
 Ml. 36^b1 247
 Ml. 36^c21 119
 Ml. 36^d2 221
 Ml. 37^b23 208
 Ml. 37^a12 258
 Ml. 37^b24 216
 Ml. 38^a5 262
 Ml. 38^a11 156
 Ml. 38^c3 242
 Ml. 38^c26–7 155
 Ml. 39^a10 246
 Ml. 39^c15 242, 244, 264–5
 Ml. 39^c27 158
 Ml. 39^c28 155, 229
 Ml. 39^d11 156, 221
 Ml. 39^d13 157
 Ml. 39^d14 229
 Ml. 40^a15 196
 Ml. 40^a20 184, 190
 Ml. 40^c22 252
 Ml. 41^d9 202, 225
 Ml. 42^b18 208
 Ml. 42^c8 252
 Ml. 43^b11 159
 Ml. 43^d14 244
 Ml. 44^a19 258
 Ml. 44^b4 242, 244
 Ml. 44^b22 247
 Ml. 44^b26 158
 Ml. 44^c15 247
 Ml. 44^c17 158
 Ml. 44^c31 159
 Ml. 44^c32 159
 Ml. 45^a8 249, 251, 261–2
 Ml. 45^a16 157
 Ml. 45^c9 257, 262
 Ml. 46^a1 187
 Ml. 46^a7 156
 Ml. 46^c20 182
 Ml. 46^d6 212
 Ml. 48^d6 229
 Ml. 48^d8 228
 Ml. 49^a3 196
 Ml. 49^a11 258
 Ml. 50^b8 257, 262
 Ml. 51^a5 187
 Ml. 51^d2 147
 Ml. 51^d12 147
 Ml. 51^d28 258
 Ml. 53^b15 157
 Ml. 53^b17 246
 Ml. 53^b18 226
 Ml. 53^c13 208
 Ml. 53^c14 159
 Ml. 54^a7 212, 233
 Ml. 54^a12 246
 Ml. 54^d6 147
 Ml. 54^d14 159
 Ml. 54^d23 256
 Ml. 55^c1 216
 Ml. 55^d11 169
 Ml. 56^a20 229
 Ml. 56^b15 251
 Ml. 56^c11 252
 Ml. 56^c22 205
 Ml. 57^a10 230
 Ml. 57^c12 119, 218, 230–1, 235
 Ml. 58^d14 156
 Ml. 59^a12 260, 262
 Ml. 59^a21 156
 Ml. 59^b9 230–1
 Ml. 59^c12 157
 Ml. 61^a21 202–3
 Ml. 61^a22 202–3

- Ml. 61^{b7} 125
 Ml. 61^{b17} 244
 Ml. 62^{a2} 252
 Ml. 62^{c7} 257
 Ml. 62^{c16} 158
 Ml. 62^{c21} 158
 Ml. 63^{b1} 159
 Ml. 63^{b9} 213
 Ml. 64^{a12} 259
 Ml. 64^{a13} 248, 259, 262
 Ml. 65^{a12} 122
 Ml. 66^{d14} 205
 Ml. 68^{d15} 247
 Ml. 69^{a17} 196
 Ml. 69^{a18} 244
 Ml. 69^{a23} 188
 Ml. 70^{a12} 229
 Ml. 70^{c6} 125
 Ml. 71^{c9} 252
 Ml. 72^{c4} 122
 Ml. 72^{c9} 187
 Ml. 73^{c2} 231
 Ml. 74^{b22} 148
 Ml. 74^{d3} 258
 Ml. 74^{d7} 198
 Ml. 75^{b7} 119
 Ml. 75^{b7} 209
 Ml. 75^{c8} 122
 Ml. 76^{a12} 246
 Ml. 79^{b5} 211
 Ml. 84^{c9} 258
 Ml. 85^{b11} 219
 Ml. 85^{d10} 230
 Ml. 86^{c3} 125
 Ml. 86^{c10} 157
 Ml. 86^{d6} 209
 Ml. 86^{d19a} 242, 245, 263, 265
 Ml. 87^{a8} 208, 233
 Ml. 90^{b10} 119
 Ml. 92^{b9} 209, 232
 Ml. 93^{a13} 169
 Ml. 93^{a15} 159
 Ml. 93^{d12} 157
 Ml. 93^{d14} 221
 Ml. 95^{c2} 213–4
 Ml. 96^{a13} 187
 Ml. 96^{b13} 122
 Ml. 96^{c11} 258
 Ml. 97^{a7} 243
 Ml. 98^{d8} 246
 Ml. 100^{a10} 200
 Ml. 104^{a4} 125
 Ml. 104^{b5} 212
 Ml. 106^{b8} 174
 Ml. 108^{a9} 229
 Ml. 109^{b2} 171
 Ml. 110^{a10} 205
 Ml. 110^{d10} 187
 Ml. 111^{b15} 252
 Ml. 111^{b17} 213
 Ml. 111^{d4} 209
 Ml. 112^{b12} 210
 Ml. 114^{a7} 156, 221
 Ml. 114^{a9} 257
 Ml. 114^{c3} 256
 Ml. 114^{d10} 158
 Ml. 115^{b15} 251–2
 Ml. 115^{c1} 125
 Ml. 115^{d9} 214
 Ml. 117^{b9} 256
 Ml. 120^{a3} 159
 Ml. 120^{a6} 256
 Ml. 120^{d2} 258
 Ml. 121^{d22} 158
 Ml. 122^{c9} 125
 Ml. 123^{b7} 199
 Ml. 124^{b3} 147, 242, 244
 Ml. 124^{d7} 221
 Ml. 125^{a9} 156
 Ml. 125^{d4} 242
 Ml. 126^{b12} 246
 Ml. 127^{c8} 157
 Ml. 127^{d8} 184
 Ml. 129^{b1} 213, 230
 Ml. 129^{b2} 227–8, 230
 Ml. 130^{a16} 125
 Ml. 131^{b8} 229
 Ml. 131^{c10} 154, 172
 Ml. 131^{d12} 119, 154
 Ml. 134^{a1} 156
 Ml. 135^{a13} 209
 Ml. 135^{d4} 159
 Ml. 136^{c11} 201
 Murphy 1956: 26–7 128

- O'Brien 1952, verse 12c 132
 Ó Cuív 1966: 174 140
Olr. Homily (Strachan 1907: 3) 222
Olr. Table of Commutations §12 (Binchy 1962) 118
- Passions and Homilies from Leabhar Breac*
 (Atkinson 1887: 643) 133
- RIA MS 23 P 2, f. 222^r10 137
- Saltair na Rann* 1306 137
Saltair na Rann 1942 137
Saltair na Rann 5367 124
Scél Mongáin (White 2006: 76) 118
- Sg. 2^b2 190
 Sg. 3^a1 247
 Sg. 4^b12 125
 Sg. 6^b11 188
 Sg. 8^b2 256
 Sg. 9^a16 256
 Sg. 17^a7 156
 Sg. 26^a5 215
 Sg. 26^a8 215
 Sg. 26^b12 183
 Sg. 28^a4 256
 Sg. 29^a3 255
 Sg. 33^a17 256
 Sg. 33^a19a 191
 Sg. 33^a23 256
 Sg. 36^b1 256
 Sg. 39^b1 150, 256
 Sg. 39^a25 215
 Sg. 40^a15 229
 Sg. 45^a15 214
 Sg. 45^b19 190, 251–2
 Sg. 50^a1 256
 Sg. 51^a4 256
 Sg. 53^a11 256
 Sg. 56^b10 256
 Sg. 59^b12 256
 Sg. 61^a1 256
 Sg. 62^b2 204
 Sg. 63^a16 184
 Sg. 70^a6 215
 Sg. 73^b1 191
- Sg. 77^b5 252
 Sg. 90^b12 191
 Sg. 104^b1 125
 Sg. 137^b2 208
 Sg. 139^b2 159
 Sg. 145^a4 213
 Sg. 146^b16 255
 Sg. 148^b12 125
 Sg. 150^b1 121
 Sg. 150^b5 121
 Sg. 161^a1 252
 Sg. 180^b2 255
 Sg. 183^a3 159
 Sg. 183^b3 220
 Sg. 188^a7 256
 Sg. 188^a8 256
 Sg. 188^a12 256
 Sg. 188^a13 256
 Sg. 188^a16 256
 Sg. 188^a19 256
 Sg. 193^a1 256
 Sg. 197^a11 210–11
 Sg. 197^b4 255
 Sg. 198^b3 255
 Sg. 200^b6 187
 Sg. 201^b6 190
 Sg. 202^b3 235
 Sg. 203^a16 125
 Sg. 208^b15 224
 Sg. 209^b10 187
 Sg. 209^b13 255
 Sg. 210^a10 255
 Sg. 215^a2 183
 Sg. 218^a9 256
 Sg. 220^a6 215
 Sg. 221^b7 229
- Southampt. Psalter* (Ó Néill 2012: LXIII no. 12) 117
- Táin Bó Fraích* (Meid 1974: 130–1) 260
- Thes. 2: 241.15 131
 Thes. 2: 241.16 242
 Thes. 2: 241.8–9 228, 235–6
 Thes. 2: 252.14 117
 Thes. 2: 253.16 117
 Thes. 2: 255.7 117

- Thes. 2: 290.14 119
 Thes. 2: 292 v.8 119
 Thes. 2: 294.13 119
 Thes. 2: 312.4 217–8, 222
 Thes. 2: 317.6 205–6
Tochmarc Étaíne (Bergin & Best 1934–8
 §23) 221
 Tur. 58^a 234
 Tur. 60 242
 Tur. 108 183
- Vienna Bede* 23 183
Vienna Bede 31 183
Vita Tripartita (Mulchrone 1939: l. 116;
 Stokes 1887: 11) 138
Vita Tripartita (Mulchrone 1939: l. 2417;
 Stokes 1887: 207) 138–9
- Wb. 1^a2 153
 Wb. 1^b14 198
 Wb. 1^c20 203
 Wb. 2^a4 203
 Wb. 2^a11 226
 Wb. 2^a12 156
 Wb. 3^a15 262
 Wb. 3^b1 251
 Wb. 3^b10 236
 Wb. 3^b23 159
 Wb. 3^c21 251
 Wb. 3^c25 226
 Wb. 3^c33 169
 Wb. 3^c34 169
 Wb. 3^d10 196
 Wb. 3^d27 228
 Wb. 4^b19 226
 Wb. 4^c13 226
 Wb. 4^c27 216–8, 222, 227
 Wb. 4^c32 207, 211, 232–3
 Wb. 4^c33 195, 207, 211, 232–3
 Wb. 4^d32 244
 Wb. 5^b12 251
 Wb. 5^b28 184, 219, 222
 Wb. 5^b29 254
 Wb. 5^b30 157
 Wb. 5^b32 217–8, 222, 227
 Wb. 5^b35 163
 Wb. 5^c6 153, 232
- Wb. 5^c16 181
 Wb. 5^c23 226
 Wb. 5^d5 169
 Wb. 6^a13 229
 Wb. 6^c8 201, 227, 250
 Wb. 6^b13 157
 Wb. 6^b22 226
 Wb. 6^d11 154
 Wb. 7^a2 252
 Wb. 7^a4 158
 Wb. 7^a5 158
 Wb. 7^a12 159, 181
 Wb. 7^b11 240, 242
 Wb. 7^b15 232
 Wb. 7^d3 181
 Wb. 8^c16 157
 Wb. 9^a16 159
 Wb. 9^a23 156
 Wb. 9^b3 226
 Wb. 9^b13 226
 Wb. 9^b19 156
 Wb. 9^c10 226, 228
 Wb. 9^c20 157
 Wb. 9^d6 204
 Wb. 9^d9 198
 Wb. 10^b3 182, 226
 Wb. 10^b27 122
 Wb. 10^c1 182
 Wb. 10^c22 237
 Wb. 10^d26 157, 220
 Wb. 11^a19 226
 Wb. 11^d12 181
 Wb. 12^c5 200
 Wb. 12^c8 247
 Wb. 13^a29 217–8, 227, 231–2
 Wb. 13^b7 215
 Wb. 13^c24 226
 Wb. 13^d20 211
 Wb. 14^a29 242
 Wb. 14^b13 226
 Wb. 14^c18 58, 154
 Wb. 14^c20 157
 Wb. 14^c42 169
 Wb. 14^d37 157
 Wb. 15^d12 181
 Wb. 16^a2 226
 Wb. 16^a11 211

Wb. 16^b8 244
 Wb. 16^b18 212
 Wb. 16^c9 207, 233
 Wb. 16^d4 215
 Wb. 16^d9 226
 Wb. 17^b10 156, 226
 Wb. 17^c5 172
 Wb. 18^b16 228
 Wb. 18^c6 216
 Wb. 19^a20 155, 218–9
 Wb. 19^b5 158, 237
 Wb. 19^b10 215
 Wb. 19^d10a 157
 Wb. 19^d16 215
 Wb. 19^d27 159
 Wb. 20^d15 216
 Wb. 21^a2 206, 223, 225–6
 Wb. 21^a3 249–50
 Wb. 21^c5 242
 Wb. 21^c12 227
 Wb. 21^d9 207, 233
 Wb. 22^d18 226
 Wb. 23^a3 210
 Wb. 23^a9 184
 Wb. 23^b1 207
 Wb. 23^b12 181
 Wb. 23^b27 226
 Wb. 23^c17 250
 Wb. 24^a17 228
 Wb. 24^a32 226
 Wb. 24^b13 228
 Wb. 24^b26 252
 Wb. 24^c17 229
 Wb. 25^a26 157
 Wb. 25^a38 212, 233
 Wb. 25^d1 181
 Wb. 26^c4 242, 244
 Wb. 27^b1 230–1
 Wb. 27^d19 195, 208, 226
 Wb. 28^a3 251–2
 Wb. 28^a17 212
 Wb. 28^b17 197
 Wb. 28^b3 243
 Wb. 28^b32 195, 212
 Wb. 28^c11 226

Wb. 28^c14 199
 Wb. 29^a25 163
 Wb. 29^c4 209, 224
 Wb. 29^d29 254
 Wb. 30^a9 156
 Wb. 30^d13 226
 Wb. 31^a6 223
 Wb. 31^c11 228
 Wb. 31^c16 159, 250
 Wb. 31^c17 226
 Wb. 32^a25 201–2
 Wb. 32^c10 230
 Wb. 33^a12 158
 Wb. 33^b1 228, 231, 237
 Wb. 33^b13 169
 Wb. 33^c12 169

Latin

2 Cor. 7:13 212
Historia Regum Britanniae, Geoffrey of
 Monmouth (Reeve & Wright 2007: 45
 §32) 303
Imago Mundi (Flint 1983; 48–9) 307
 Mt. 27:42 299
Philemon 21 201
Pilgrim Girardus, BL Arundel MS 346
 (Neuhaus 1886: 38) 305
Pilgrim Girardus, BL Cotton Cleopatra MS C X
 (Neuhaus 1886: 38) 305–6
Pilgrim Girardus, BL Cotton Cleopatra MS C X
 (Neuhaus 1886: 39) 306
 Rom. 3:2 203
 1 Tim. 3:16–4:1 199–200

English

Ælfric, *Genesis* 3.8 271
 Ælfric, *Grammar* 96.11–2 271
Genesis 2629 284
Orosius IV.11 271
 Wycliffe, *Old Testament*, *Genesis* 3.8 272

French

Boeve de Haumtone (Stimming 1899: 68,
 l. 1797) 307

Queste del Saint Graal (Pauphilet 1923
l. 129) 319

Vie de sainte Doucline (Albanès 1879
§41) 320

Southwest British (Cornish and Breton)

Angers MS 477 fol. 62b (Fleuriot 1964b:
413) 314

Angers MS 477 fol. 65a (Fleuriot 1964b:
262) 314

Angers MS 477 fol. 79a (Fleuriot 1964b:
412) 314

Beunans Meriasek l. 594 317

Beunans Meriasek l. 941 318

Beunans Meriasek l. 970–77 324

Beunans Meriasek l. 978 323

Beunans Meriasek l. 978–82 323

Beunans Meriasek l. 1801–8 334

Beunans Meriasek l. 1807–8 332–4

Beunans Meriasek l. 1888 332–4

Beunans Meriasek l. 3224–5 332–4

Beunans Meriasek l. 4340–2 332–4

Beunans Meriasek l. 4515–6 332–4

Bewnans Ke (Thomas & Williams 2007)
ll. 208–12) 321

Bewnans Ke (Thomas & Williams 2007)
ll. 213) 321

Bible translation of Wella Rowe (Loth 1902:
180) 330

Boson family, writings (Padel 1975 §3) 330

Boson family, writings (Padel 1975 §46)
330

Middle Cornish Charter (Toorians 1991
l. 15–22) 322

Middle Cornish Charter (Toorians 1991
l. 20) 321

Norris 1859, 1: l. 1259–64 323

Norris 1859, 1: l. 1263 321

Norris 1859, 1: l. 2423–6 325

Norris 1859, 2: l. 909 317

Passion (Stokes 1860–1: st. 9^a) 317

Passion (Stokes 1860–1: st. 42^a) 325

Passion (Stokes 1860–1: st. 54^a) 325

Tregear Homilies 1^r, ll. 1–5 335–6

Tregear Homilies 1^r, l. 3 328

Tregear Homilies 1^r, l. 5 328

Tregear Homilies 2^r, ll. 18–9 328

Tregear Homilies 4^r, ll. 2–3 328–9

Tregear Homilies 4^r, ll. 2–6 336–7

Tregear Homilies 8^v, l. 14 328

Vie de sainte Catherine (Ernault 1887a
§8) 316

Vie de sainte Catherine (Ernault 1887a
§12) 316

Vie de sainte Catherine (Ernault 1887a
§13) 316

Vie de sainte Nonne (Ernault 1887b
l. 240) 318

William Bodinar's letter (Pool & Padel
1976–6: 234.7) 331

Welsh

Breuddwyd Maxen (Williams 1908 ll. 18–9)
44

Brut y Brenhinedd (Roberts 1971 ll. 795–7)
320

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 2^v: 7–9 282–3

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 5^v: 27 293

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 6^r: 10–11 293–4

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 11^r: 9 286

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 11^r: 21 285

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 11^v: 3 283

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 19^v: 1 285–6

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 20^r: 14–6 303–4

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 24^v: 15 283

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 34^v: 23 283

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 55^v: 19 293

Brut y Brenhinoedd, BL Cotton Cleopatra MS
B V i, 96^v: 22–3 295

Brut y Brenhinoedd, Cardiff MS 1.362 (Hafod 1),
63^v: 19 284

Brut y Brenhinoedd, NLW MS 3035, 61^v:
13 281

- Brut y Brenhinoedd*, NLW MS Peniarth 46, 254: 16 281
- Brut y Brenhinoedd*, Oxford Jesus Coll. MS 111, 39^f (154): 27 302
- Credo Athanasius* (Intr.), NLW MS Peniarth 5, 14g. B v. 196 301
- Culhwch ac Olwen* (Bromwich & Evans 1992 l. 595) 38
- Cynghorau Catwn* (Lewis 1935: 23.28) 277
- Cynghorau Catwn* (Lewis 1935: 24.25) 277
- Cynghorau Catwn* (Lewis 1935: 29.37) 277
- Cynghorau Catwn*, NLW MS Llanstephan 27, 31: 3 309
- Cynghorau Catwn*, NLW MS Llanstephan 27, 32: 20 309
- Cynghorau Catwn*, NLW MS Llanstephan 27, 165^v: 15 310
- Cynghorau Catwn*, NLW MS Llanstephan 27, 168^f: 16 290, 308
- Cynghorau Catwn*, NLW MS Peniarth 3 ii, 37: 3 308–9
- Cynghorau Catwn*, NLW MS Peniarth 3 ii, 38: 11 290
- Delw'r Byd*, Oxford Jesus Coll. 111, 243^f (976): 14 307–8
- De Mensuris et Ponderibus* 85 (Lambert 2003: 113) 314
- Efengyl Nicodemus*, NLW MS Peniarth 5, 32^f: 14 300
- Gwyrtheu Mair*, NLW MS Llanstephan 27, 176^f: 10 301
- Gwyrtheu Mair*, NLW MS Peniarth 14 (Jones 1939: 148) 301
- Gwyrtheu Mair*, NLW MS Peniarth 14 (Jones 1941: 24) 300, 305–6
- Gwyrtheu Mair*, NLW MS Peniarth 14 (Jones 1941: 25) 300, 306
- Gwyrtheu Mair* (Jones 1939: 336.33) 276
- Juvenicus Englynion* 5^{a-b} 314
- NLW MS Peniarth 9, 1^v: 4 302
- Oxford Jesus College MS 111, 95^v (400): 9 302
- Pedeir Keinc y Mabinogi* (Williams 1951: 1. 3–4) 38
- Pedeir Keinc y Mabinogi* (Williams 1951: 2. 12–13) 277
- Pedeir Keinc y Mabinogi* (Williams 1951: 16. 18–9) 316
- Pedeir Keinc y Mabinogi* (Williams 1951: 20.29) 276
- Pedeir Keinc y Mabinogi* (Williams 1951: 44.11) 316
- Pedeir Keinc y Mabinogi* (Williams 1951: 45. 2–3) 316
- Pedeir Keinc y Mabinogi* (Williams 1951: 46.27) 276
- Pedeir Keinc y Mabinogi* (Williams 1951: 68. 15–6) 317
- Poppe 2009: 254 296
- Saith Doethion Rhufain*, Oxford Jesus Coll. MS 20, 56^{av}: 1 295
- Saith Doethion Rhufain*, Oxford Jesus Coll. MS 111, 131^f (541): 34 294
- 'Surexit' Memorandum 313
- Thomas 1996 (Grammar), 269 295
- Y Groglieth*, NLW MS Llanstephan 27, 105^v: 17 299
- Y Groglieth*, NLW MS Peniarth 5, 7^v: 18 309
- Y Groglieth*, NLW MS Peniarth 5, 7^v: 20 300
- Y Groglieth*, NLW MS Peniarth 7, 58^v (215): 2 299
- Y Groglieth*, Shrewsbury MS 11, 113: 16 299
- Y Groglieth*, Shrewsbury MS 11, 114: 1 309
- Ystoria Bilatus*, NLW MS Peniarth 5, 11^f: 14 303–4
- Ystoria Bilatus*, NLW MS Peniarth 7, 63^v (236): 20 303–4
- Ystoria Bown de Hamtwn*, NLW MS Peniarth 5, 134^v (301): 22 307
- Ystoria Bown de Hamtwn*, NLW MS Peniarth 5, 148^v 292
- Ystoria Carolo Magno: Chronicl Turpin*, NLW MS Peniarth 5, 74^f (63): 34 302
- Ystoria Titus*, NLW MS Peniarth 5, 37^f: 48 303–4
- Ystoriau Saint Greal*, NLW MS Peniarth 11, 63^v: 24 294
- Ystoriau Saint Greal*, NLW MS Peniarth 11, 239^v: 9 287

Thematic Index

- absolute / conjunct distinction 239–40, 265
- adjectives 36
- comparison 36, 119, 121, 206–11, 223–4, 284
 - positive 211–6
- adverbs, adverbials 36, 38, 152, 172, 179–80, 195–204, 206, 214–5, 223, 225, 228–9, 248–9, 257–8, 263, 269, 272–3, 284, 315–8, 325
- antecedent 250
- adjectival 195–237
 - adverbial 197, 202, 225, 228–9, 249
 - mixed 204, 224–5, 233
 - prepositional 239, 249–50, 253–63
 - subject/object 196, 200–2, 218, 220, 222–3, 227, 233
- authorship attribution 85–87, 89–90, 97, 99, 101–4
- clause types 143–8, 151–4, 163–73, 175–7, 313
- cleft sentence 44, 152–4, 195–206, 215–9, 222, 224–5, 228–36, 239, 245, 248–60, 263–6, 315, 317
- complement clause 36, 172, 198
- complexity of language 24–5, 49–50, 59, 65–66, 69, 72
- configuration/constituent order 17, 19, 44–5, 47, 152, 313–32
- contracted verbs 251–2
- convergence 270, 302, 308, 312
- corpora, treebanks and texts 15–26, 33, 45–7
- dictionaries 19, 22, 49, 61–5, 70, 79–80, 83
 - Irish 1–3, 63–4, 67, 78, 85–9, 94–5, 101–2
 - Late Cornish 327–32
 - Middle Cornish 325–7, 332
 - Welsh 27–9, 46–7, 272, 278–9, 281–2, 286–90, 296–7, 311
- deictic particle 130, 136–7
- dialects 188–92
- figura etymologica 153, 196–7, 220, 248
- finite-state morphology/transducer 21, 23–4, 66–73, 75–6, 78–83
- information structure 17–8, 20, 29–30, 33, 36, 41, 44–5, 47
- intensifier 269–78, 280–5, 287, 293–7, 299, 303–5, 307, 309–11
- left periphery 314–5, 317–9, 321–2, 324, 334
- machine learning 24, 66, 79, 85, 87, 89, 94, 103–4
- morphological analyser 23, 49–50, 65–8, 82–3
- morphosyntax, morphosyntactic annotation 15, 17, 19, 22–3, 28, 31–3, 37, 40–1, 47, 54, 65, 154
- mutations 28–9, 32–3, 40, 56, 90, 92, 145–6, 189, 239, 250
- lenition 252
 - nasalisation 56, 58, 120, 145–6, 152–3, 165–7, 179–93, 195–8, 202–3, 205–25, 227–33, 237, 250, 252, 256–8, 261–3
- natural language processing (NLP), toolkit 22, 28, 31, 33, 41, 50, 61, 64, 66
- NLP pipeline 29–45, 64
- annotation 15–47, 63–4
 - lemmatisation/lemmatiser 16, 19, 25, 64–7, 71, 79–82
 - POS tagging/tagger 15–6, 21–40, 44–7, 64–8, 75, 81–3, 89–90, 101, 104
 - pre-processing 29–32
 - syntactic parsing/parser 15–6, 21–2, 25–31, 35–7, 41, 43–7, 52, 62, 64, 67–8, 70, 75–6, 78–9, 332
 - tokenisation 26, 31–2, 47, 64
- nota augens* 56–60, 115–6
- orthography 25, 28, 31–3, 40, 64, 71, 82, 92, 115, 250, 329

- standardisation, normalisation 19, 24–5, 28–30, 33, 47, 50–1, 64–7, 71, 79–82, 90, 92–3, 104, 189
- paradigmatic split 173–5
- particle theory/'Cowgill particle' 239–40, 246–8, 263–5
- phonological changes
 - depalatalisation 166–9
 - i-apocope 239–40
 - loss of nasals 189, 192, 196, 229
 - pretonic e > a 242
- preverbs 37–8, 52–8, 60, 70–2, 76–8, 80–1, 143–5, 147–51, 155–71, 174–6, 214, 220–1, 225–6, 231, 239–8, 261, 263–5, 308, 310
- pronouns 34–36
 - demonstrative, Old Irish 115–141, 153
 - infixes (general) 56, 86, 90, 119, 246, 261, 271, 288, 294–5, 309–11
 - infixes, Class A 56, 143–55, 162–7, 169, 172, 175–6, 216–9, 221, 226–7, 229–31, 233, 244, 250
 - infixes, Class B 56, 143–67, 169–70, 172–7, 216–8, 220–2, 229, 233, 244, 250
 - infixes, Class C 56, 143–7, 149, 151–67, 169–76, 218–22, 227, 230–1, 233, 250
 - prepositions with pronominal objects 34–35, 38, 90, 121, 127–8, 134, 166–8, 179–180, 182–3, 187, 191, 243, 281, 283–5
- reflexive pronouns/markers,
 - reflexivity 269–78, 281–2, 286–312
- relative clauses 58, 143–7, 151–3, 162–3, 165–7, 169–72, 175, 195–237, 248–262
 - leniting relative clauses 166, 169–70, 196, 201, 218–23, 225, 232–3, 237, 250, 252, 254, 256, 258–61, 263
 - nasalising relative clauses 58, 152–3, 195–233, 237, 249, 252, 256–8, 262–3
- relative marker 76, 78, 174, 230
- statistical methods 22, 24, 65–6, 85–6, 89–93, 99–100
 - Fisher's exact test 241
 - F-scores 65
 - k-medoids 91–104
 - Principal Component Analysis 93–4
 - ten-fold cross-validation 39
- syntactic raising 206, 223–4, 228, 231, 233, 237
- syntax, syntactic annotation 16–21, 23–4, 29–31, 36, 41–3, 45–7, 64, 143, 146, 152–6, 160–1, 171–2, 176, 179–80, 183–6, 293–7, 308, 313–32
- verbal complex 52–60, 70–1, 75, 143, 145–55, 166, 168–9, 171, 176, 223, 243–5, 248
- XML, TEI XML 20, 30, 43, 46–7, 63, 86, 88–9

Word Index

Old and Middle Irish

a^N ‘when’ 228–9
ad- 147, 151, 160, 220
aisndís, aisndisse 186
aith- 163–4
améin/amné 258, 263
ar-, ara- 168, 226, 233, 244
arindí 257
as-, asa- 147, 244, 255
as-beir, ·epir 73–4
as-oilgi 70
at-tá 234–7
beirid 75
bete 237
ceta-, cita-, ciata- ‘first’ 239–48, 263–5
ceta- ‘along’ 241, 246–8, 264–5
cían 205–6, 225
co^H 215
con- 151
conid- 174
cruth, in chruth 228–9, 249
-d^L- 152
de/i- 150
di-sruthaigedar, dísruthigthe 255–6
do- 150
dó 252
do-esta 169
do-léici, ·teilci, ·tarlaic 71–3, 76–7
dommae 236
etar, itar 242–4
fo-fera 169–71, 173, 176
forngaire 186
frecndairc, frecndarcus 186
fris- 160, 174–5, 221
frit-racatar 172
í 130, 136–7
imm-, imma- 244
immaircide 258–9
in(d)-, inda- (preverb) 244
in(d) (adverbial particle) 214–5
léicid 68, 73–4, 76–7
ocu- 246–8, 264
ol sé, ol sí 137

-s (infix pronoun as relative marker) 230
samlaid 258, 263
sé, sí (personal pronoun) 137–41
sin, sen, sein, sain 115, 117–24, 131–4, 136–7, 141
-sin 115–6
siú 128, 130, 136
só, sé, séo (demonstrative pronoun) 115, 117, 124–31, 134–5, 138–40
-so, -se 115–7
sommae 236–7
sund 129
túailnge, túailngigidir 186

Welsh

efo 280–1
hun(an) 269–70, 276–312
sef 30
ym- 270–1, 274–7, 286–93, 296–7, 300–2, 307, 310–1
ynteu 280

Breton

em- 308
en, em 308

Latin

deus 173–4
dīuus 173–4
ipse 305–7
se 274, 291, 303–4, 306–7
te 304–5

Proto-Celtic

**ambi* 276
**enter* 242–3
**eti* 240, 246, 263
**kanta* 241, 247, 264
**kentu-* 240, 242–8, 263
**onku-* 246, 248, 264

Proto-Indo-European

**s(ǵ)e-* 269–70, 274

Germanic

hine, hem, him (Old and Middle English) 271–2, 284

selbst (German) 269, 273, 276–7, 310–1

-self (English) 269–74, 276, 284, 310–1

ser, sik (Old Norse) 270

sich (German) 269, 273, 275, 277, 292, 311

sih (Old High German) 270

sik (Old Saxon) 270

sik, sis (Gothic) 270

Romance

me (Old French) 307

se (French) 269, 274–5, 308

se- (Surselvan) 274

si (Italian) 269, 274