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

https://doi.org/10.1515/9783110680744-001

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 *Ambr. 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 Fransen'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 *Corpws 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 / *ané* '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.