Breton embedded V2 and post-syntactic operations

This is a study of how embedded domains are integrated into the syntactic structure. I develop a typology of verb second orders in Breton embedded domains by comparing corpus data and the elicitation data with traditional native speakers from three different dialectal points in Leon and Kerne. The results suggest that some structures undergo a post-syntactic Merge operation at *TRANSFER* before Spell-Out (Wurmbrand 2012). These structures show signs of less integration, among which ban on movement, or resolution of the V2 requirement by last resort strategies against the syntactic rules of head movement constraints or excorporation.

Breton V2. Like Old Romance, Rhaeto-Romance, Karitiana, Germanic Mòcheno or Cimbrian, Breton has a rich left-periphery allowing for V3. Some V3 orders (and more) appear in embedded domains. In (1), the base-generated topic subject of the passive satisfies V2 in the structural region between the ForceP projection and the Fin head, spelled-out as a pre-Tense particle *a* (2). The recursive projection hosting scene-setting adverbials is higher. This rich high periphery hosts elements accidentally satisfying V2 as a by-product of information structure. Breton V2 is typologically peculiar in that it is *linear* (Borsley & Kathol 2000): functional heads count as preverbal constituents, including Q particles (3), preverbal negation (4) or verbal heads (5)-(7). Avoidance of verb-first shows post-syntactic symptoms, as it does not seem to impact semantics or information structure. It allows for syntactic misbehaviours, like *Long Head Movement* (5) or excorporation, leading to analytical tenses (6) or doubling (7) as a last resort. The *linear V2* generalization predicts that embedded domains headed by a complementizer should show no extra V2 effect.

(1) *Bep bloazh neuze e veze dreset, ablamour, pa vez fall an amzer*,
each year then prt was rebuilt because when was bad the weather
a-wechoù ar paper sablet a veze roget gant ar gwallamzer…
sometimes the paper sanded prt was destroyed by the bad weather
'It was rebuilt every year, because, when the weather is bad, the sanded paper was sometimes destroyed by the bad weather.' embedded T4, *Plougerneau*

(2) [Hanging topics [scene setting advs. [$_{\text{TopP}}$ XP subj. [$_{\text{FocP}}$ [$_{\text{FinP}}$ [*a/e*-V] [$_{\text{IP}}$ …

(3) ***Hag*** *eo gwir an dra-se* ?
Q is true the thing-here
'Is that true?'

(4) (Yann ha Lisa) ***ne*** *brenint ket ul levr d'am breur warc'hoazh.*
Yann & Lisa neg will.buy.3PL neg a book to'my brother tomorrow
'Yann and Lisa will not buy a book for my brother tomorrow.'

(5) *Prenet en deus _ Yann ul levr d'am breur.* LHM,
bought has Yann a book to'my brother Stylistic Fronting
'Yann has bought a book for my brother.'

(6) *Prenañ a ra Yann ul levr d'am breur.* analytic tense =
buy prt does Yann a book to'my brother excorporation + *do* support
'Yann buys a book for my brother.'

(7) *Gouzout a ouzon ar wirionez.* excorporation + copy pronunciation
to.know prt know.1SG the truth
'I know the truth.'

Typology. Much like in Germanic, embedded V2 appears in different adjunct clauses denoting cause (1), complements of verbs of saying and thinking (8), including forms of the complementizer 'if'(9) & (10), relatives of temporal nouns (11).

(8) *Me oar a-walh* ***lar*** *eur vuoh wenn he-deus kalz a lêz.* *Uhelgoat*
I know enough that a cow white she-has lot of milk

(9) *N' ouzon ket* ***hag*** *(lennet) e deus (*lennet) an urioù* *Plougerneau*
Neg know neg Q read has read the book
'I don't know if he has read the book.'

(10) *N' ouzon ket* ***ha*** *(lennet /-g-eñ) en deus (lennet) al levr.* *Treger*
neg know neg Q read / expl prt has read the book
'I don't know if he has read the book.' *Lesneven*

(11) *Bevañ a reomp un amzer* ***hag*** *gouzout a ra (ar vugale) diouzh an ordinatourien*
live prt do.we a time that to.know prt does the children from the computers
'We live a time where the children know better of the computers (than their parents).'

Tests. These embedded V2 orders show less integration in the structure: they can not be moved (13), and matrix negation does not have scope over them (14) (they also show a looser pragmatic integration and have restrictions on extraction from them).

(12) (*Peogwir eo lezireg*), *n'eo ket deuet,* (*peogwir eo lezizeg*).
because is lazy NEG'is not come because is lazy embedded C-T…

(13) *(*Peogwir lezizeg eo*), *n'eo ket deuet*, (*peogwir lezireg eo*)
because lazy is NEG'is not come because lazy is embedded C-XP-T…
'He didn't come because he is lazy.'

(14) CONTEXT: 'Don't be nasty! He didn't come with me only because I have a car and he didn't want to walk…'
*N'eo ket deuet peogwir (eo) lezireg (*eo), met evit kaozeal samples.*
NEG'is not come because is lazy is but for discuss together
'He didn't come because he is lazy but for us to have a discussion.'

Analysis. One speaker allows for multiple C heads (15) but accepts very few embedded V2 (16), mild counter evidence for a CP recursion analysis for selected CPs (Vikner 1995, Holmberg & Platzack 1995, Watanabe 1992, Iatridou & Kroch 1992, Heycock 2000). Last-resort verbal head movement in (9)-(11) suggests that the V2 word order rearrangement takes place in a post-syntactic morphological component before spell-out. In these less integrated embedded clauses, the completed FinP phase is first sent to *TRANSFER*. The V2 requirement applies, triggering Merge of an expletive, short-distance constituent inversion (LHM, Stylistic Fronting), or excorporation. Only after do the complementizer and its embedded sentence Merge, making it impossible for the V2 requirement to be saturated by the presence of the complementizer. Embedded V2 is also possible in fully integrated structures like the protasis of conditionals, but it never allows for last-resort expletives/verbal head fronting. Dialectal variation shows a gradation in the richness of their left-periphery (Eastern dialects persistently allow for less options, and Plougerneau in Leon for more (17)).

(15) *N'ouion ket* ***(la) ma*** *teuio.* *Kerne*
NEG'know.I not that if come.FUT
'I don't know if she will come.'

(16) *Me oar mat* ***lar*** *(*'r vuoh wenn) ' ra ('r vuoh wenn) kalz laezh.* *Kerne*
I know well that a cow white does a cow white lot milk
'I know well that a white cow produces a lot of milk.'

(17)a. *Kontant e vichen* ***ma*** *d'ar gouel e teufe Yann.* *Leon (Plougerneau)*
b. * /? *Kontant e vichen* ***ma*** *d'ar gouel e teufe Yann.* *Leon (Lesneven)*
c. * *Kontant ' vefen* ***ma*** *d'ar fest e teufe Yann.* *Kerne*
happy prt be.COND if to'the party prt would.come Yann
'I would be happy for Yann to come to the party.'

Subject behaviour in Old Irish: word order, verbal noun clauses, and relative clauses

In this paper, I will discuss my ongoing research into subjecthood in Old Irish, specifically focussing on potential subject tests for Old Irish. Alignment and argument structure lie at the heart of current theoretical linguistic models. Extensive work has been carried out on argument structure in modern languages, while work on the ancient and medieval Indo-European languages is beginning to take off. Keenan (1976), in his seminal work on the subject concept, developed a set of over 30 "universal" properties, with coding, behavioural, and (some more controversial) semantic properties. These properties do not apply in all languages and it has been argued that subject properties are language specific or perhaps even construction-specific.

Research on argument structure and subjecthood in the medieval Celtic languages is ongoing. Existing work in this area tends to be descriptive, historical, and comparative (Bergin, 1938; Isaac, 1993; Mac Coisdealbha, 1998; Mac Giolla Easpaig, 1980; Watkins, 1963; for an exception, see Lash, 2014a, 2014b). For the modern languages, subject tests have been determined based on the subject properties in these languages. Applying these subject to the ancient and medieval languages is however problematic and Old Irish is no exception. A specific problem for Old Irish is that several common subject tests, such as omission in control constructions and raising, depend on the presence of language-specific structures like the infinitive, which is absent from the language. In the proposed paper, then, I will identify potential subject tests for Old Irish. Specifically, I will discuss position (or word order), the possibility of applying subject tests requiring an infinitive to verbal-noun clauses, and subject behaviour in leniting and nasalising relative clauses.

Position

It is well known that Old Irish has VSO word order in clauses with neutral word order (Thurneysen, 1975, p. 327). Other subject positions are possible (see e.g. Lash, 2014a), although these are restricted by information structure. Old Irish is also a pro-drop language, so a subject pronoun can be left unexpressed. Objects follow the subject, although sentential adverbs may intervene between the two. Both objects and subject-like obliques may also be infixed in the verb (e.g. *ni-s·fil*, 'they are not'), following prepositions, or suffixed to prepositions. Clauses with both nominal subjects and nominal objects are rare, but preliminary research into position as a subject test for Old Irish is promising (Le Mair, Johnson, Frotscher, Eythórsson, & Barðdal, 2017).

Subject tests involving infinitives/verbal nouns

Several common subject tests, such as omission in control constructions and raising, depend on the presence of language-specific structures like the infinitive. Old Irish does not have infinitives, but it does have analogous constructions, using verbal nouns. Verbal noun constructions replace constructions found in other Indo-European languages including participles, gerunds, and genuine infinitives, and in some cases appear where a finite subordinate clause might be expected. An extensive study of verbal nouns in Medieval Irish and Welsh has been undertaken by Ronan (2006).[1] In control infinitives, the subject of the infinitive is left unexpressed under identity with the subject or the object of the preceding clause, or it may even be retrieved from the context. In the case of verbal nouns, the subject can be expressed in the genitive or with a prepositional phrase, or it can be left unexpressed.

Subject-to-object raising involves a main verb and an infinitive, where the subject of the non-finite verb behaves syntactically as the object of the finite verb. The subject of the

[1] See also Jeffers, 1978; Müller, 1999; and Stüber, 2009.

infinitive is in the accusative when embedded under raising-to-object verbs, but the object of the infinitive behaves like it does in ordinary finite clauses. In subject-to-subject raising, the finite verb does not select for a subject of its own; instead the subject of the infinitive assumes the behaviour of the subject of the finite verb. The paper will discuss the behaviour of subjects of verbal nouns in potential control and raising constructions.

Relative clauses

A prospective new subject test for Medieval Irish is behaviour in relative clauses. Of most interest are leniting and nasalising relative clauses. The leniting relative clause is obligatory when the antecedent is felt as the subject of the relative clause. When the antecedent is felt as the object of the relative clause, either a leniting or a nasalising relative clause can be used (Ó hUiginn, 1986; Thurneysen, 1975, pp. 312–325). When the concept expressed in the relative clause is felt as the subject, the relative verb is always in the third person. It may be possible to identify a subject test based on the distribution of leniting and nasalising relative clauses.

Although I cannot hope to expound on all the intricacies of subjecthood and argument structure in Old Irish in a single paper, it is my aim to provide a theoretical analysis of a core aspect of syntax, investigating argument structure and the boundaries of subjecthood in Old Irish.

Bibliography

Bergin, O. (1938). On the syntax of the verb in Old Irish. *Ériu*, *12*, 197–214.

Isaac, G. R. (1993). Issues in the reconstruction and analysis of Insular Celtic syntax and phonology. *Ériu*, *44*, 1–32.

Jeffers, R. J. (1978). Old Irish verbal nouns. *Ériu*, *29*, 1–13.

Keenan, E. (1976). Towards a universal definition of “subject.” In C. N. Lee (Ed.), *Subject and topic* (pp. 303–34). New York: Academic Press.

Lash, E. (2014a). Subject positions in Old and Middle Irish. *Lingua*, *148*, 278–308.

Lash, E. (2014b). The parsed Old and Middle Irish corpus (POMIC).

Le Mair, E., Johnson, C. A., Frotscher, M., Eythórsson, T., & Barðdal, J. (2017). Position as a behavioral property of subjects: the case of Old Irish. *Indogermanische Forschungen*, *122*(1), 111–142.

Mac Coisdealbha, P. (1998). *The syntax of the sentence in Old Irish: selected studies from a descriptive, historical and comparative point of view*. Tübingen: Niemeyer.

Mac Giolla Easpaig, D. (1980). Aspects of variant word order in Early Irish. *Ériu*, *31*, 28–38.

Müller, N. (1999). *Agents in Early Welsh and Early Irish*. Oxford: Oxford University Press.

Ó hUiginn, R. (1986). The Old Irish nasalizing relative clause. *Ériu*, *37*, 33–88.

Ronan, P. (2006). *Aspects of verbal noun constructions in Medieval Irish and Welsh. With reference to similar constructions in Basque*. Maynooth.

Stüber, K. (2009). *Der altirische do-Infinitiv – eine verkannte Kategorie*. Bremen: Hempen Verlag.

Thurneysen, R. (1975). *A Grammar of Old Irish. Revised and enlarged edition with supplement.* Dublin: Dublin Institute for Advanced Studies.

Watkins, C. (1963). Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb. *Celtica*, *6*, 1–49.

Saothrú Georges Dottin i réimse chanúineolaíocht na Nua-Ghaeilge/ Georges Dottin's contribution to the field of Modern Irish dialectology

French Celticist Georges Dottin (1863-1928) was a pioneer in the field of Irish dialectology – his field-work, undertaken in Galway in the Summer of 1891, and his earliest associated subsequent articles relating to the dialect of Irish he recorded there at that time – Dottin (1893; 1895) – pre-date, for instance, Franz Nikolaus Finck's and Hoger Pedersen's earliest contributions to the field – Finck (1896), Pedersen (1896) – Finck and Pedersen generally being recognised as *the* pioneers of Irish dialectology following their field-work, independent of each other, in the Aran Islands in 1895-6, and following publication of Finck (1896a; 1899) and Pedersen (1897; 1899). Georges Dottin subsequently published two further articles relating to the phonetics of Modern Irish – Dottin (1899; 1900). In the former, Dottin, referring to phonetic examples cited, states (1899: 308): 'Les exemples sans références sont tirés … d'un vocabulaire irlandais d'environ quatre milles mots que j'ai recueilli à Galway en 1891.' Dottin did not publish any further significant contribution drawn from his own 1891 Galway field-work, but at least some of this material has survived to this day in Dottin's personal notebooks preserved in the Library of the University of Rennes.

The present paper, to be delivered in Irish, will contextualise Dottin's contribution to Modern Irish dialectology. It will examine Dottin's early collaboration – both folkloristic and political – with Douglas Hyde – see, for example, Dottin (1893a), Hyde (1899) and de h-Íde (1933) – and it will also provide new information on Dottin's field-work in Galway in 1891, part of which was carried out in the Aran Islands.

References:

de h-Íde, Dubhglas (1933) *An Sgéaluidhe Gaedhealach (Sgéalta as Connachta)*. Institiút Béaloideasa Éireann.

Dottin, Georges (1893) 'Études sur la pronunciation actuelle d'un dialecte irlandais,' *Revue celtique* 14, 97-136

Dottin, Georges (1893a) *Contes irlandais: extraits du* "Leabhar Sgeulaigheachta" *de M. Douglas Hyde*. Rennes: Typographie Oberthur.

Dottin, Georges (1895) 'Contes irlandais,' *Revue celtique* 16, 421-449.

Dottin, Georges (1899) 'Études de phonétique irlandaise,' *Revue celtique* 20, 306-334,

Dottin, Georges (1900) 'Études de phonétique irlandaise (suite),' *Revue celtique* 21, 59-74; 179-199

Finck, Franz Nikolaus (1896) 'Vier neuirische Zaubersprüche,' *Zeitschrift des Vereins für Volkskunde* 6, 88-92.

Finck, Franz Nikolaus (1896a) *Wörterbuch der auf den Araninseln gesprochenen westirischen mundart*. Marburg: Universitäts-Buchdruckerei (R. Friedrich).

Finck, Franz Nikolaus (1899) *Die araner mundart: Ein beitrag zur erforschung des westirischen*. *I*. *Grammatik*. *II*. *Wörterbuch*. Marburg: Elwert.

Hyde, Douglas (1899) 'The National Board and Foreign Opinion on it,' *An Claidheamh Soluis* 7/10/1899, 468-9.

Pedersen, Holger (1896) ‘Zu den neuirischen Zaubersprüchen,‘ *Zeitschrift des Vereins für Volkskunde* 6, 192-196.

Pedersen, Holger (1897) *Aspirationen i Irsk: En Sproghistorisk Undersøgelse*. Leipzig: M. Spirgatis.

Pedersen, Holger (1899) ‘Die aspiration im Irischen. II” *Zeitschrift für vergleichende Sprachforschung* 35, (Neue Folge, Band 15) 315-444.

Two types of vowel reduction in Gaeilge Chorca Dhuibhne

In this talk, I address vowel reduction in Gaeilge Chorca Dhuibhne (GCD), one of the southern dialects of Modern Irish, spoken in Co. Kerry, Ireland. It has been noted in previous literature that the phonological status of the unstressed schwa vowel in GCD, i.e. whether it's a segment present in the lexicon or a surface segment arising as a result of post-lexical reduction process, is not clear (see Iosad 2013), and I propose a possible answer to this question. The data come primarily from my own recordings of GCD speakers gathered in 2015–2016.

Reduction to schwa is applied to unstressed short vowels (long vowels are hardly ever reduced to it). The lexical stress in southern dialects, unlike initial stress in other varieties of Irish, is dependent upon syllable weight (see Ó Sé 1989, Hickey 2014 among others). By default, syllables with phonologically long vowels count as heavy (H) and syllables with short vowels count as light (L). The stress is attracted to the second syllable if it is H, to the third syllable if it is H and the preceding syllables are L, and on the first syllable in all other cases (Ó Sé 2008: 96, see also Ó Sé 2000 or Iosad 2013 for a finer description).

This rule of stress assignment is, however, not exceptionless, and one exception particularly relevant for the discussion in this paper is presented by words with the string /ax/ in the second syllable (and sometimes, probably, tautosyllabic /a.x/ with /a/ in the second syllable). Such words tend to demonstrate peninitial stress in the absence of H's, e.g. *bacach* [bəˈkax] 'beggar'. On the evidence of this pattern, a unique hierarchy of syllable weight has been proposed for southern dialects of Irish: CVV(C) > Cax > CV(C) (Gordon 2006: 27, based on the analysis from Doherty 1991). I try to show that rather than resulting from the properties of the consonant /x/ in combination with the vowel /a/ (this position is taken not only by Doherty 1991, but by various scholars including Blankenhorn 1981, Green 1996, and Bennett 2017), the aforementioned pattern is due to the presence of an underlying reduced vowel /ə/ in the first (extra-light) syllable, at least synchronically (this idea is previously mentioned in Ó Sé 2000: 38, 53 and Iosad 2013: 100, yet without much detail; see O'Rahilly 1932 for the history of /ax/).

It can be gleaned from the process of stress retraction that there exist two types of vowel reduction. Roughly, stress retraction is a process in GCD which replaces the stress towards the beginning of a given word in situations, where a clash with the stress in the following word could eventually arise (Ó Sé 2000). In word combinations leading to potential stress-clash, like *bacach óg* [bəˌkaˈxoːg] 'young beggar' (also in some non-initially stressed words without /ax/), retraction is blocked, whereas in examples like *buidéal mór* [ˌbɪdʲeːlˈmuːəɾ] 'big bottle' retraction succeeds and

restores the quality of the vowel in the first syllable, cf. *buidéal* [bəˈdʲeːl] ‘bottle’ in the citation form. Thus, one (“phonological”) type of reduction applies on the word level and the other (“phonetic”) follows a phrase-level process of stress retraction, cf. a similar situation in Modern Standard Russian (Iosad 2012).

Another piece of evidence comes from the acoustic analysis of GCD vowels coming primarily from the data of four GCD speakers collected in 2016. Vowels resulting from the two types of reduction are overtly distinct, /ə/ surfacing with formants characteristic of a mid vowel, quite unlike “phonetically” reduced vowels, which demonstrate a more diverse distribution.

Finally, whether this situation, when both a word-level reduced vowel and phrase-level reduction coexist in one variety, can be regarded as a case of rule scattering (Bermúdez-Otero 2010) will also be discussed.

References

Bennett, Ryan. 2017. Output optimization in the Irish plural system. *Journal of Linguistics* 53(2). 229–277.

Bermúdez-Otero, Ricardo. 2010. Morphologically conditioned phonetics? Not proven. Paper given at On Linguistic Interfaces II, Belfast, 2 December 2010.

Blankenhorn, Virginia S. 1981. Pitch, quantity and stress in Munster Irish. *Éigse* 18. 225–250.

Doherty, Cathal. 1991. Munster Irish stress. *Phonology at Santa Cruz* 2. 19–32.

Gordon, Matthew. 2006. *Syllable weight: Phonetics, phonology, typology*. New York: Routledge.

Green, Antony D. 1996. Stress placement in Munster Irish. In Lise M. Dobrin, Kora Singer & Lisa McNair (eds.), *Chicago Linguistic Society 32: Papers from the Main Session*. 77–92.

Hickey, Raymond. 2014. *The sound structure of Modern Irish*. Berlin, Mouton de Gruyter.

Iosad, Pavel. 2012. Vowel reduction in Russian: No phonetics in phonology. *Journal of Linguistics* 48. 521–571.

Iosad, Pavel. 2013. Head-dependent asymmetries in Munster Irish prosody. *Nordlyd* 40(1). 66–107.

O’Rahilly, Thomas F. 1932. *Irish Dialects: Past and present, with chapters on Scottish and Manx*. Dublin: Browne & Nolan.

Ó Sé, Diarmuid. 1989. Contributions to the study of word stress in Irish. *Ériu* 40. 147–178.

Ó Sé, Diarmuid. 2000. *Gaeilge Chorca Dhuibhne*. Baile Átha Cliath: Institiúid Teangeolaíochta Éireann.

Ó Sé, Diarmuid. 2008. Word stress in Munster Irish. *Éigse* 36. 87–112.

Relative clauses in the *Book of the Anchorite of Llanddewibrefi*

My paper will present the results of a corpus-based analysis of the relative clauses in *Llyfr Ancr Llanddewibrefi (*Oxford, Jesus College MS. 119*,* 1346). In his edition of the manuscript containing religious texts that are all with one exception translations, John Morris-Jonescomments repeatedly on the construction of relative clauses in this text as one of the best examples of 'the effect upon literary Welsh of translation from Latin'(Morris-Jones 1894: xxvi-xxvii). The construction he criticises mostly (cf. also Morris-Jones 1913: 288; Morris-Jones 1931: 104 and Sims-Williams 2016: 151-2) is the one described in *The Syntax of Welsh* in the following way (for Modern Welsh):

> 'In archaic literary style, an overt demonstrative pronoun, such as *yr hwn* 'that one (masc.)', may be used as a relative pronoun:
>
> | y | dyn | yr | hwn | a | gafodd | y | wobr |
> |---|---|---|---|---|---|---|---|
> | the | man | the | DEM.MS | PRT | get.PAST.3S | the | prize |
>
> 'the man who got the prize'
>
> This usage was largely modelled on foreign languages, and has mostly fallen out of use (for details, see Richards 1938: 75)' (Borsley, Tallerman, Willis 2007: 119).

The same construction is characterised as 'imitation of the syntax of Latin or of the dominant neighbouring languages, rather than natural developments in speech' (Borsley, Tallerman, Willis 2007: 335)

On the whole these constructions have been often regarded as one of the 'traces of translation' (see Evans 1964: 66, 69; Luft 2016: 171-2, 176 with a collection of examples and discussions from earlier scholarship).

It is therefore worthwhile to investigate the evidence of the Book of the Anchorite in its entirety and not just selected examples.

First the relative frequency of the so called 'proper relative clauses' vs sentences with overt demonstratives in different texts of the manuscript will be presented. Attention is also being paid to the semantics of the clauses, since according to John Morris-Jones the overt demonstratives 'are also employed even when the antecedent is expressed, if the relative clause is coordinate, that is, introduces a new idea instead of merely qualifying a noun in the principal sentence' (Morris-Jones 1931: 98), that is there is a possible distinction being made between restrictive and non-restrictive clauses. Secondly, the variation of different types of overt antecedents *(*e.g. *yr hwnn, y gwr, ar yr hwnn)* will be analysed. Thirdly, the data of these religious texts will be compared to the corpus of canonical Middle Welsh prose, including the *Four Branches of the Mabinogi* and *Culhwch ac Olwen*.

This quantitative approach allows on the one side to refine our understanding of different forms and semantics of relative clauses in Middle Welsh and on the other hand to map individual texts of the manuscript in a continuum from the language of the native tales to a specific register of translated texts and thus elaborate the conception of stylistic choices within the language of Middle Welsh prose.

References:

Borsley, Robert D.; Tallerman, Maggie; Willis, David (2007): The Syntax of Welsh. Cambridge: Cambridge University Press.

Evans, D. Simon (1964): A grammar of Middle Welsh. [Dublin]: Dublin Institute for Advanced Studies (Mediaeval and modern Welsh series. Supplementary volume).

Luft, Diana (2016): Tracking ôl cyfieithu. Medieval Welsh translation in criticism and scholarship. In *Translation Studies* 9 (2), pp. 168–182. DOI: 10.1080/14781700.2015.1118404.

Morris-Jones, John (1913): A Welsh Grammar, Historical and Comparative: phonology and accidence. Oxford: At the Clarendon Press.

Morris-Jones, John (1931): Welsh syntax. An unfinished draft. Cardiff: The Univ.of Wales Press.

Morris-Jones, John; Rhŷs, John (Hg.) (1894): The Elucidarium and other tracts in Welsh from Llyvyr agkyr Llandewivrevi ("The book of the anchorite of Llandewivrevi") A.D. 1346 (Jesus college ms. 119). Oxford: Clarendon Press (Anecdota Oxoniensia, 1,6).

Sims-Williams, Patrick (2016): John Morris-Jones and his *Welsh Grammar*. In *Transactions of the Honourable Society of Cymmrodorion* 22, pp. 134–153.

Developing An Auto-Glosser for Scottish Gaelic Using a Corpus of Interlinear Glossed Text

1 Introduction

Interlinear Glossed Text (IGT) is widely used in linguistic studies. (1) is an example of Scottish Gaelic IGT.

(1) Tha a athair nas sine na a mhàthair.
be.pres 3sm.poss father comp old.cmpr comp 3sm.poss mother
'His father is older than his mother.'

IGT is essential in linguistic research and analysis of a language's grammar. However, building large IGT corpora of under-resourced languages is expensive and time consuming. Endeavors to develop resources and tools for under-resourced languages are ongoing (Raghallaigh and Měchura, 2014; Lamb and Sinclair, 2016; Lamb and Danso, 2014).

In this paper, we present the first auto-glosser for Scottish Gaelic, which automatically generates the gloss line. We also introduce a corpus of Scottish Gaelic IGT from which the auto-glosser is trained. Additionally, we show that this glossing data significantly improves the performance of the machine translation systems that we are currently developing.

2 The Auto-Glosser

The Hidden Markov Model (HMM) is used widely and successfully in part of speech tagging tasks (Kupiec, 1992). We treat the glosses as special part of speech tags, and build the auto-glosser using HMM. Specifically, the gloss of a target Gaelic word is inferred by considering the relationship of the target Gaelic word to the predicted glosses of the two preceding Gaelic words. Consider the following example:

(2) ... word_1 word_2 word_3 ...
... gloss_1 gloss_2 ??? ...

To determine the gloss of word_3, the glosser selects the gloss that is most likely in the given context of word_3, gloss_1 and gloss_2. The accuracy rate of this auto-glosser is 65.8%, with 3986 possible glosses.

The primary goal of the auto-glosser is to facilitate the glossing process. We continue to expand the IGT corpus by incorporating data collected during a language documentation project. The auto-glosser aids and expedites the glossing task by generating a draft of Gaelic IGT, which a team member then corrects and verifies.

To understand the utility of the auto-glosser, we consider the IGT in (1) again. Specifically, only the first line in (1) is independent; the second line is determined by the first line. If the line one in (1) is our newly collected data, it will be glossed by the auto-glosser first, and the output is (3). Our team member then reviews and corrects (3) into (4), and then provides a free-translation of the Gaelic sentence into English.

(3) Tha **a** athair nas sine na a mhàthair.
be.pres **det** father comp old.cmpr comp 3sm.poss mother

(4) Tha **a** athair nas sine na a mhàthair.
be.pres **3sm.poss** father comp old.cmpr comp 3sm.poss mother
'His father is older than his mother.'

In this manner, the task is machine-aided, and we do not need to gloss from scratch.

3 Description of the Corpus

The essential component for any machine learning systems is a sizable and accurate training data. The key of the auto-glosser is our corpus of Scottish Gaelic IGT. The corpus has 8,367 Gaelic sentences, and in term of words, it has 52,778 Gaelic words/glosses. The data of the corpus is from two different sources: fieldwork and data elicitation. IGT in our corpus is treated as parallel text, a format is commonly used in machine translation. Specifically, IGT is stored in three related plain text files: 1) language text in orthography, 2) sequences of glosses, and 3) free English translations.

4 Other Application of the IGT Corpus: Machine Translation System

Given that the IGT data can be viewed as parallel texts, they can be the used directly as training data for machine translation systems. Using our data and OpenNMT (Klein et al., 2017), we built neural network machine translation systems and ran several experiments comparing two types of systems. Both types of systems have the Gaelic transcription and English translation data; The critical difference is whether the gloss data is incorporated. To the best of our knowledge, no published machine translation system exists that incorporates gloss data. The performances of the systems with gloss data incorporated are a lot better than those built only with Gaelic and English data. By including the gloss data, the BLEU score is more than doubled (without gloss: 0.165; with gloss: 0.357). Moreover, the gloss-incorporated systems also outperform Google translation (a BLEU of 0.248).

The important implication of this application is that linguistics can be really relevant and beneficial for NLP.

5 Conclusion

We introduce a IGT corpus and its applications. The corpus and the source codes of the applications will be made freely available. These endeavors are meant to provide open, useful and usable tools and resources for Scottish Gaelic. Moreover, the applications of IGT demonstrate that natural language processing techniques and linguistics research can be mutually beneficial and informative.

References

G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. OpenNMT: Open-Source Toolkit for Neural Machine Translation. *ArXiv e-prints*, 2017.

J. Kupiec. Robust part-of-speech tagging using a hidden markov model. *Computer Speech & Language*, 6(3):225–242, 1992.

W. Lamb and S. Danso. Developing an automatic part-of-speech tagger for scottish gaelic. In *Proceedings of the First Celtic Language Technology Workshop*, pages 1–5, 2014.

W. Lamb and M. Sinclair. Developing word embedding models for scottish gaelic. *PARIS Inalco du 4 au 8 juillet 2016*, page 31, 2016.

B. O. Raghallaigh and M. B. Měchura. Developing high-end reusable tools and resources for irish-language terminology, lexicography, onomastics (toponymy), folkloristics, and more, using modern web and database technologies. *CLTW 2014*, page 66, 2014.

The interaction of gender and mutation in Breton:
a cross-generational perspective

This paper uses fieldwork data from a range of speakers to examine grammatical gender and initial consonant mutation (ICM) in Breton. Breton nouns may have masculine or feminine gender, which is marked by agreement with pronouns and the cardinal numerals 2, 3 and 4, and by patterns of initial consonant mutation. For example, feminine singular nouns undergo lenition following the definite article, but masculine nouns do not: feminine *taol* 'table' becomes *an daol* 'the table', but masculine *ti* 'house' is simply *an ti* 'the house'. The numerals also trigger mutations of their own on the following noun: *daou di* 'two houses'; *div daol* 'two tables'.

The gender of Breton nouns is not generally reflected in their phonology or morphology, as it is in some languages (Corbett, 1991), although some suffixes tend to result in a noun being masculine or feminine – for example, the singulative *-enn* always makes the noun feminine (Irslinger, 2014). Like in Welsh, the Breton gender system can be considered **opaque**. This is first because not all nouns mutate: there are a number of exceptions to the mutation rules, and in addition to this, not all initial consonants undergo mutation. Secondly, there is no one-to-one relationship between mutation and gender: feminine singular nouns undergo lenition following the article, but so do masculine plural nouns referring to humans, such as *pesketerien* 'fishermen' > *ar besketerien* 'the fishermen'. Lenition is thus marking both masculine and feminine, and singular and plural nouns. Studies of Welsh have shown that this system is difficult for children to acquire, and they have not reached adult proficiency even by the age of 11 (Gathercole et al., 2001; Thomas & Gathercole, 2007; Thomas & Mayr, 2010).

The system of gender and mutation in Breton is particularly interesting in light of language revitalisation movements and the growing number of younger speakers whose parents may not speak Breton, and whose only input is Breton-medium education. The Breton context therefore differs somewhat from that in Wales, where a greater number of young Welsh speakers come from Welsh-speaking homes. Various claims have been made about the 'Neo-Breton' variety used by younger Breton speakers, including that they omit or are confused about the system of initial consonant mutation (Hornsby, 2005). In light of the fact that many younger speakers come from French-speaking homes, it would not be surprising to find some degree of influence from French. Equally, older speakers who now have few opportunities to speak Breton may show attrition effects (Dressler, 1991).

To investigate these issues, three groups of Breton speakers were interviewed: (i) 9 older adult speakers aged 58-83 who grew up speaking Breton (mean age = 72.3); (ii) 7 younger adult speakers aged 24-52 who learnt Breton largely through education (mean age = 38); (iii) 12 teenagers aged 15-17 (mean age = 16.1) who are currently educated through the medium of Breton.

Speakers were asked to produce Breton words in response to pictures, and three contexts were examined: first, a singular noun preceded by the definite article; secondly, the noun preceded by the numeral 2 (*daou/div*); and thirdly, the plural. For example, for 'boat', speakers were asked to produce 'the boat', 'two boats' and '(many) boats'. A separate picture was provided for each context, and 18 distinct nouns were elicited. The goal was to examine whether speakers used the expected patterns of gender and mutation: agreement in the numeral, lenition of feminine nouns following the definite article, and lenition of all nouns following the numeral. It also permitted a

comparison of distinctive mutation, where the mutation is subject to grammatical constraints and conveys information, and contact lenition, which affects all items regardless of their grammatical features (Press, 1986). Finally, including the plural allowed the elicitation of the unmutated form as a control.

The findings from the fieldwork indicate that most speakers have a strong grasp of the mutation system, and this is particularly true of both groups of adult speakers, who use the expected mutation in at least 85 per cent of contexts. This applies to both lenition of feminine nouns and lenition following *daou/div*. The teenage speakers lag a little further behind, and are more likely to omit the mutation than the adults. There is also a lot of interspeaker variation: while some teenagers use mutation and gender with adult-like proficiency, others seem to struggle: a few speakers, for example, use only one form of the numeral, regardless of the gender of the noun. This suggests that some younger speakers may be confused about gender in Breton, and this consequently impacts upon their use of initial consonant mutation – speakers are unlikely to use the expected mutations if they are unsure of the gender of the nouns in question.

It is interesting to note speakers' own feelings about gender and mutation in Breton. Several speakers reported that they found gender 'difficult', and were unsure of the gender of certain nouns, particularly less frequently-used nouns, such as *gozh* 'mole'. Even older speakers made contradictory claims about the gender of nouns – stating for example, that a noun was masculine, but then using the feminine patterns of agreement. This is unsurprising, given the high degree of dialectal variation in traditional Breton, and the lack of opportunities speakers may have to use Breton frequently. It also makes the system more challenging for younger speakers to acquire.

In sum, speakers of all ages are more proficient at using the Breton gender and mutation system than some accounts would suggest. Some teenage speakers struggle with the gender of certain nouns, or with the system in general, which in turn affects their use of mutation. However, there is little to suggest that younger speakers as a whole are confused about initial consonant mutation, or that they omit with any frequency.

Corbett, Greville. 1991. *Gender* (Cambridge: Cambridge University Press). **Dressler, Wolfgang U. 1991.** 'The sociolinguistic and patholinguistic attrition of Breton phonology, morphology and morphonology' in Herbert W. Seliger and Robert Michael Vago (eds.) *First Language Attrition* (Cambridge: Cambridge University Press) 99–112. **Gathercole, Virginia C. Mueller, Enlli M. N. Thomas & N. I. Laporte. 2001.** 'The acquisition of grammatical gender in Welsh' *Journal of Celtic Language Learning* 6: 53-87. **Hornsby, Michael. 2005.** 'Néo breton and questions of authenticity' *Estudios de Sociolingüística* 6(2): 191-218. **Irslinger, Britta. 2014.** 'The gender of abstract noun suffixes in the Brittonic languages' in Sergio Neri and Roland Schuhmann (eds.) *Studies on the collective and feminine in Indo-European from a diachronic and typological perspective* (Leiden; Boston: Brill) 57-113. **Press, J. Ian. 1986.** *A Grammar of Modern Breton* (Berlin: Mouton de Gruyter). **Thomas, Enlli Môn & Virginia C. Mueller Gathercole. 2007.** 'Children's productive command of grammatical gender and mutation in Welsh: an alternative to rule-based learning' *First Language* 27(3): 251–278. **Thomas, Enlli Môn & Robert Mayr. 2010.** 'Children's Acquisition of Welsh in a Bilingual Setting: A Psycholinguistic Perspective' in Delyth Morris (ed.) *Welsh in the Twenty-First Century* (Cardiff: University of Wales Press) 99–117.

Abstract consonant representations in Irish

This paper makes the case for a phonological analysis of Irish as containing abstract consonants, defined only for secondary localisation. Positing these abstract consonants can simplify the description of a number of phenomena in both Modern and Old Irish, and brings greater generalisations.

One of the most heated debates in early generativist linguistics revolved around how abstract phonological representations should be. Some scholars sought to restrain the excessive generative power of transformational grammars (e.g. Chomsky and Halle 1968) by placing limits on the degree of abstractness permitted (i.a. Postal 1968; Kiparsky 1968; Hooper 1976 etc.). Throughout the 1970s and into the 1980s, debate raged over 'abstract' versus 'concrete' analyses of given languages, such as French (Schane 1968), Nupe (Hyman 1970, 1973; Harms 1973), and Seri (Marlett 1981). Nowadays, abstract representations are commonplace, both in descriptive work written in a broadly structuralist framework (Drude 2014 for Awetí), and in phonological approaches that are primarily representational rather than computational (e.g. Ségéral and Scheer 2001).

Evidence for abstract consonants in Modern Irish comes from two main sources: the behaviour of prothetic consonants before vowel-initial words, and past tense allomorphy.

The secondary localisation of a prothetic consonant before a vowel-initial word cannot be predicted from the quality of the initial vowel (the nasal of the definite article is taken here as prothetic, with examples from Ní Chiosáin 1991: 80ff.):

iontas	iːntəs	'wonder'	(m. sg.)
an iontas	ə**nʲ** iːntisʲ	'the wonder'	(m. sg. gen. def.)
aois	iːsʲ	'age'	(f. sg.)
an aois	ə**n** iːsʲ	'the age'	(f. sg. nom. def.)

This phenomenon, consistently represented in Modern Irish orthography, was already noted by Gussmann (1986), who proposed that the unexpected vowel stems, e.g. *aois* above, have an empty C-slot on the CV-tier to which an autosegmental specification for consonant quality is associated. Ní Chiosáin (1991: 84) extends this to all vowel stems, arguing that they all begin with an onset specified only for secondary localisation, here transcribed as broad /ø/ and slender /øʲ/. The only alternative would be to argue that the selection of either a broad or a slender allomorph of prothetic /t/ or /n/ is assigned simply arbitrarily to each stem.

The past tense in Modern Irish is formed by leniting an initial concrete consonant, e.g. *cuir* → *chuir* 'put' and prefixing /d/ to an initial abstract consonant, e.g. *ith* → *d'ith* 'ate'. Words beginning in /f/ combine both rules, e.g. *freagair* → *d'fhreagair* 'answered'. Assuming that /fʲ/ disappears under lenition makes it difficult to explain the prefixed /dʲ/, as words beginning in a sonorant lack this, e.g. *rith* → *rith* 'ran' not ***drith* (Armstrong 1975). However, if /f~fʲ/ is rather considered lenited to the abstract consonant /ø~øʲ/, then the formation of the past tense can be subsumed under a single rule: lenite an initial concrete consonant and prefix /d/ to an initial abstract consonant.

Anderson (2016) adopts an abstract consonant analysis from Modern Irish into Old Irish, and extends it by considering long vowels to be sequences of short vowel plus abstract consonant, i.e. /Vø/, contrasting with disyllabic sequences /VøV/ for vowels

in hiatus (this remains perfectly compatible with the zero consonant in Ó Cuív 1966). Indeed, many person markers in the verbal system can thus be considered to consist solely of abstract consonants that assimilate a preceding consonant, so that the conjugation of consonant-final strong verb forms and vowel-final hiatus verbs can be unified as follows:

bʹər-ø°ə ø°	→	bʹər°ə ø°	*biru*	'I carry' (absolute)
bʹə-ø°ə ø°	→	bʹə ø°ə ø°	*biuu*	'I do be' (absolute)
·bʹər-ø°	→	·bʹə ø°	*·biur*	'I carry' (conjunct)
·bʹə-ø°	→	·bʹə ø°	*·bíu*	'I do be' (conjunct)

Similar generalisations can be made for other aspects of the verbal morphology, such as the reduplication of vowel-initial verbal stems.

It is preferable, a priori, to assume that underlying phonological representations are as similar to the surface reality as possible, a principle known as *Lexicon Optimisation* in Optimality Theory (Prince and Smolensky 1993). On the other hand, it is also desirable to maximise the level of generalisation of our linguistic analyses. In a language such as Irish, with its complex system of consonant alternations, the abstract consonant solution seems able to simplify considerably the description of several morphological phenomena, and is worth the cost of a little abstraction.

Anderson, C. 2016. *Consonant colour and vocalism in the history of Irish*. PhD disseration, University of Poznań.

Armstrong, J. 1975. 'A note on initial mutation in Modern Irish' in *Linguistic Inquiry* 6.2, 317-23.

Chomsky, N. and Halle, M. 1968. *The sound pattern of English*. New York: Harper and Row.

Drude, S. 2014. 'Reduplication as a tool for morphological and phonological analysis in Awetí' in: Gomez, G. G. and Van der Hoort (eds.) *Reduplication in indigeneous languages of South America*, 185-215.

Gussmann, E. 1986. 'Autosegments, linked matrices and Irish lenition' in: Kastovsky, D. and A. Szwedek (eds.), *Linguistics across Historical and Geographic Boundaries*. Amsterdam: Mouton de Gruyter, 891-907.

Harms, R. T. 1973. 'How abstract is Nupe?' in *Language* 49.2: 439-46.

Hooper, J. B. 1976. *An introduction to natural generative phonology*. New York: Academic Press.

Hyman, L. M. 1970. 'How concrete is phonology' in *Language* 46: 58-76.

Hyman, L. M. 1973. 'Nupe three years later' in *Language* 46: 447-52.

Kiparsky, P. 1968. *How abstract is phonology?* Bloomington: Indiana University.

Marlett, S. A. 1981. 'The abstract consonant in Seri' in *Proceedings of the Seventh Annual Meeting of the Berkeley Linguistics Society*, 154-65.

Ní Chiosáin, M. 1991. *Topics in the phonology of Irish*. PhD dissertation, University of Massachusetts.

Ó Cuív, B. 1966. 'The phonetic basis of Classical Modern Irish rhyme' in *Ériu* 20: 94-103

Postal, P. M. 1968. *Aspects of phonological theory*. New York: Harper and Row.

Prince, Alan and Smolensky, Paul. 1993. *Constraint interaction in generative grammar*.

Schane, S. 1968. *French phonology and morphology*. Cambridge, Mass: MIT Press.

Ségéral, P. and Scheer, T. 2001. ‘Abstractness in phonology: the case of virtual geminates’ in: Dziubalska-Kołaczyk, Katarzyna (ed.), *Constraints and Preferences*. Berlin and New York: Mouton de Gruyter, 311–337.

Vowel nasalisation in Scottish Gaelic:
The search for paradigm uniformity effects in fine-grained phonetic detail

According to the modular feedforward architecture of grammar, the phonetics is sensitive only to the output of the phonology and is thus blind to morphological or lexical conditioning (Pierrehumbert 2002). However, this prediction is challenged by claims that fine-grained phonetic detail may display e.g. *paradigm uniformity* (PU) effects (Steriade 2000) or effects of usage factors such as lexical frequency (Bybee 2001). In the present study I search for potential phonetic PU effects in vowel nasalisation in Scottish Gaelic by investigating alternating items in which a nasalising environment is removed by a morpho(phono)logical process. A clear distinction is found between categorical phonological nasalisation, which displays overapplication in derived forms, and gradient phonetic nasalisation, which disappears completely when the triggering environment is removed. I present this as evidence for the modularity of the phonetics-phonology interface and the non-existence of phonetic PU effects.

Some putative instances of phonetic PU effects can be accounted for in a modular architecture by allowing prosodic structure to grant the phonetics indirect access to morphological structure. Thus, the subtly differing degrees of /l/-darkening and GOOSE-fronting found by Strycharczuk & Scobbie (2016) between simplex *hula* and complex *fool-ing* are compatible with an analysis in which *-ing* is adjoined directly to the prosodic word (*cf.* Bermúdez-Otero 2011: 2028). I argue that the search for phonetic PU effects must therefore be restricted to cases where a prosodic explanation is unavailable, and that Scottish Gaelic provides an ideal testing ground since it is rich in morpho(phono)logical processes that do not involve overt affixation.

In Scottish Gaelic, vowels normally display strong nasalisation after initial [m], e.g. *madainn* [mãtəɲ̟ʲ] 'morning', but some exceptional lexical items display far less nasalisation, e.g. *marag* [maɾak] 'pudding'. Under the *lenition* mutation, *radical* initial [m] alternates with *lenited* initial [v] under certain morphosyntactic conditions. In a nasal airflow study of one 62-year-old speaker from Ness, Isle of Lewis, I show that both radical *madainn* [mãtəɲ̟ʲ] and lenited *mhadainn* [vãtəɲ̟ʲ] display high levels of nasalisation throughout the vowel; radical *marag* [maɾak] displays some nasalisation early in the vowel; and lenited *mharag* [vaɾak] displays no nasalisation at all (*see* Figs. 1 and 2). Bonferroni-corrected unpaired *t*-tests among these and other items, comparing nasal airflow both early and late in the vowel, paint a clear picture of two distinct types of nasalisation: a categorical phonological process of progressive [nasal] spreading that overapplies in lenited forms and may be subject to lexically conditioned blocking, and a gradient phonetic process of coarticulatory nasalisation that is bled by lenition and always applies after [m]. Crucially, *mharag* [vaɾak] displays no trace of nasalisation in spite of the presence of gradient phonetic nasalisation in paradigmatically related *marag* [maɾak]. This is in line with the predictions of modular architectures, in which categorical phonology has direct access to morphological information but gradient phonetics does not.

Non-modular theories designed to account for phonetic PU effects include phonetic output-output (OO-)correspondence (Steriade 2000), in which OO-constraints penalise differences between paradigmatically related forms at the phonetic level, and Exemplar Theory, in which phonetic detail is stored in the lexicon and production of one form is influenced by simultaneous activation of paradigmatically related forms. While a negative result can in principle be handled by phonetic OO-correspondence using appropriate constraint rankings, it is more problematic for Exemplar Theoretic approaches in which phonetic PU effects emerge mechanically from exemplar dynamics. I conclude that my results are more compatible with a modular architecture in which phonetics has no direct access to morphological (or lexical) information and that wholesale dismissal of this empirically more restrictive framework is premature.

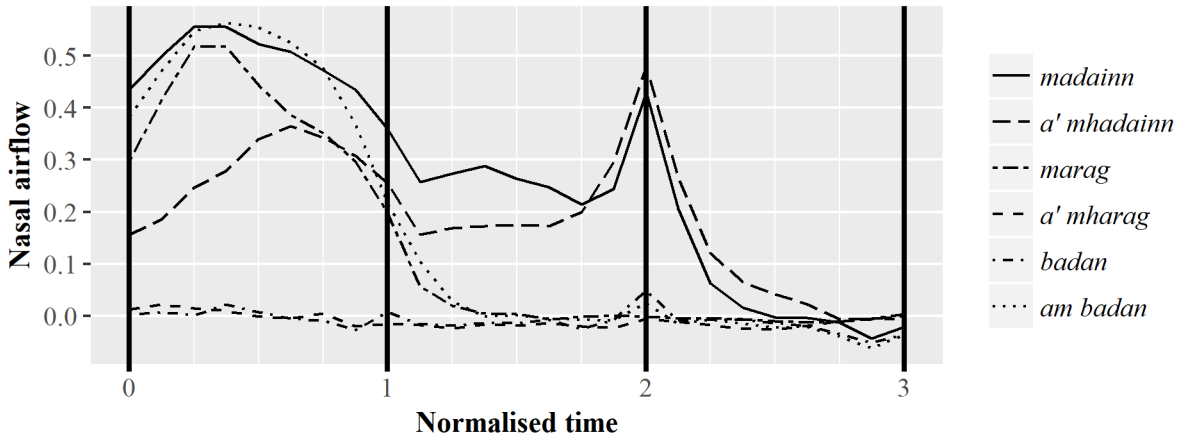


Fig. 1: Dynamic nasal airflow profiles for the underlined portions of madainn [mãtəɲʲ] *'morning'*, a' mhadainn [ə vãtəɲʲ] *'the morning'*, marag [marak] *'pudding'*, a' mharag [ə varak] *'the pudding'*, badan [patan] *'thicket' and* am badan [ə matan] (< /əm patan/) *'the thicket'. The* x*-axis represents normalised time, where 0-1 is the duration of the initial consonant, 1-2 is the duration of the vowel, and 2-3 is the duration of the following consonant.*

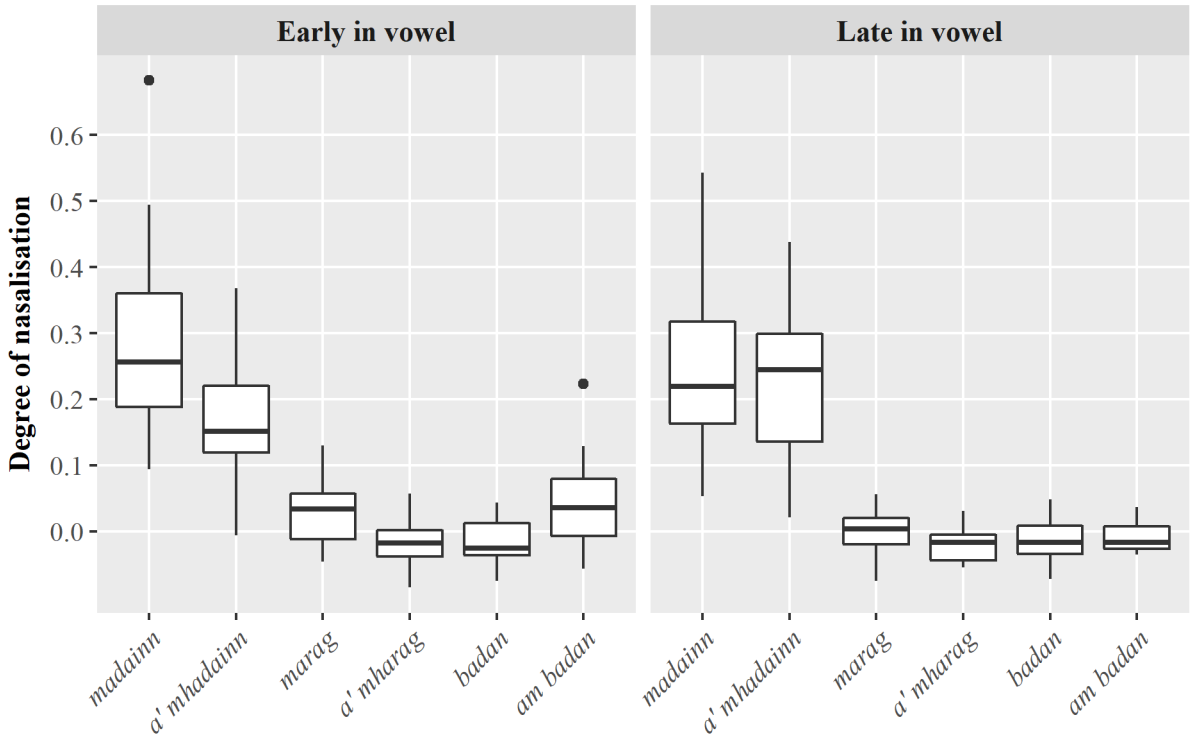


Fig 2: Degree of nasalisation early in the vowel and late in the vowel for the items in Fig. 1.

References: Bermúdez-Otero, R. 2011. Cyclicity. In M. van Oostendorp, C. J. Ewen, E. Hume & K. Rice (eds.), *The Blackwell Companion to Phonology*, 2019-48. Oxford: Blackwell. · Bybee, J. 2001. *Phonology and Language Use*. Cambridge: CUP. · Pierrehumbert, J. B. 2002. Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology* 7, 101-39. Berlin: M de G. · Steriade, D. 2000. Paradigm uniformity and the phonetics-phonology boundary. In M. B. Broe & J. B. Pierrehumbert (eds.), *Papers In Laboratory Phonology* V, 313-34. Cambridge: CUP. · Strycharczuk, P. & J. M. Scobbie. 2016. Gradual or abrupt? The phonetic path to morphologisation. *Journal of Phonetics* 59, 76-91.

Towards a computational lexical resource for the diachronic study of Irish verbs

In this paper, we propose a computational framework for a lexical resource that will better facilitate diachronic study of Irish verbs. The verbal system is subject to major morphological changes between Early Irish (c. 7th-12th centuries A.D.) and Modern Irish varieties (post-12th centuries) (McCone 1997). Moreover, whereas the literary output in the Old Irish period (c. 8th-9th centuries A.D.) points to a standardised language (Stifter 2009), all post-Old Irish historical varieties, except for bardic poetry (Early Modern Irish period, c. 13th-17th centuries A.D.), show a substantial degree of grammatical, orthographical and – particularly evident in the case of Early Modern Irish prose (Ó hUiginn 2013) – stylistic variation (cf. contributions in McCone 1994). The available digital support is insufficient to systematically trace the linguistic change and variation.

The research described here aims to mitigate the lack of digital support by creating and linking verb forms in morphologically annotated corpora by using a morphological analyser for contemporary, standardised Irish – already in the process of being adapted for successively earlier Modern Irish texts (Uí Dhonnchadha *et al.* 2014) – and by developing new tagging tools for Old Irish, to project forward to later forms.

This paper will focus on the creation of a morphological analyser for Old Irish using finite-state morphology (Beesley and Karttunen 2003). Recognition rates for an Early Irish sample text and associated findings and challenges will be reported on. The paper concludes with an outlook on the implementation stage of the lexical resource, its benefits and potential further research. We will (a) discuss challenges in morphologically tagging and accurately linking verbal cognates across historical corpora, (b) explore the ways in which this resource can serve and advance (digital) scholarship in historical Irish philology and linguistics, and (c) address more general questions relating to the balance between computational methods and manual work in successfully linking cognate verb forms.

References

Beesley, Kenneth R. & Lauri Karttunen, *Finite State Morphology*. Stanford, CA: CSLI Publications, 2003.

McCone, Kim, *The early Irish verb*. Maynooth monographs, 1 (1987). Revised second edition with index verborum. Maynooth: An Sagart, 1997.

McCone, K.; D. McManus; C. Ó Háinle; N. Williams; L. Breatnach. (eds.), *Stair na Gaeilge: in ómós do Pádraig Ó Fiannachta.* Maynooth: Department of Old Irish, St. Patrick's College, 1994.

Ó hUiginn, Ruairí, "Transmitting the text: some linguistic issues in the work of the Franciscans." In: Gillespie, R. and R. Ó hUiginn (eds.), *Irish Europe, 1600–1650: Writing and Learning,* Dublin: Four Courts Press, 2013. 85–104.

Stifter, David, "Early Irish." In: Ball, Martin J., and Nicole Müller (eds.), *The Celtic languages, Routledge Language Family Descriptions,* 2nd ed. (1993), London, New York: Routledge, 2009. 55–116.

Uí Dhonnchadha, E.; K. Scannell, K.; R. Ó hUiginn; E. Ní Mhearraí; M. Nic Mhaoláin; B. Ó Raghallaigh; G. Toner; S. Mac Mathúna; D. D'Auria; E. Ní Ghallchobhair; N. O'Leary, "Corpas na Gaeilge 1882-1926: Integrating Historical and Modern Irish Texts", *LREC 2014 Workshop LRT4HDA: Language Resources and Technologies for Processing and Linking Historical Documents and Archives - Deploying Linked Open Data in Cultural Heritage*, Reykjavik, Iceland, May 2014. 12-18.

"It's lazy but a part of everyday talk" – the acceptability of Welsh phrasal verbs among professional speakers of Welsh

Phrasal verbs have been largely absent from the linguistic description of Welsh or treated as a marginal phenomenon. Although Welsh possesses a number of native verb-particle constructions (Rottet 2005) and a substantial number of phrasal verbs have emerged as a result of natural processes of metaphorical extension of particle meaning, contact with English, in particular direct calquing, has been found to significantly reinforce the usage and productivity of verb-particle construction. For that reason, some Welsh scholars (e.g. Jones 1979; Thomas 1996) have described phrasal verbs as redundant borrowings from English or substandard elements of the colloquial. Numerous teaching materials aimed at advanced users of Welsh strive to sensitise Welsh native speakers and learners to the phenomenon of calquing, presenting PVs as incorrect, careless translations from English. Prescriptive rules against some of these constructions are also visible in a number of Welsh dictionaries. These norms are in obvious disagreement with the wide-spread usage of phrasal verbs not only in spoken, but also written semi-formal varieties of Welsh, which have been first observed by Rottet (2005) and confirmed by my own study on a corpus of Welsh fiction and press.

Bearing in mind that direct translations of PVs occur at a deeper level of interference than obvious lexical borrowings, it was considered worth investigating whether language-aware users of Welsh perceive PVs as tokens of interference from English and how deeply the monolingual-orientated standardisation processes permeate. The paper will present results of a field study conducted in Wales in the years 2015-2016, which comprised interviews with 55 professional speakers of Welsh accompanied with questionnaires on the acceptability of phrasal verbs. The term "professional speaker", adopted after McEvan-Fujita (2008) denoted speakers of Welsh who used the language in a professional environment. The study investigated the level of accommodation of phrasal verbs across four registers of the language, focusing on the written semi-formal variety. Answers to the questionnaire have been set against data retrieved from the interviews regarding the speakers' linguistic awareness, perceived competence, use of borrowings and declared attitudes towards loanwords and calques. In this way the study touched on the broader question of linguistic ideologies manifested by speakers.

The study has demonstrated that the selected PVs were considerably well-accommodated in informal spoken Welsh, while in the written registers the acceptability was much lower

depending on a number of intra- and extralinguistic factors. In general, the phenomenon of calquing phrasal verbs was not perceived as unacceptable or threatening for Welsh and a number of these constructions have been shown to be well-accommodated in the standard language.

References:

Jones, Morris. 1979. "The present condition of the Welsh language," in: Meic Stephens (ed.), *The Welsh Language Today*. Llandysul: Gomer, 112–130.

McEwan-Fujita, Emily. 2008. "Working at '9 to 5' Gaelic: Speakers, Contexts, and Ideologies of an Emerging Minority Language Register," in: Kendall A. King (ed.), *Sustaining linguistic diversity: endangered and minority languages and language varieties*. Washington, D.C.: Georgetown University Press, 81–94.

Rottet, Kevin. 2005. "Phrasal verbs and English influence in Welsh," *Word* 56, 1: 39–70.

Thomas, Peter Wynn. 1996. *Gramadeg y Gymraeg* [The Grammar of Welsh]. Caerdydd: Gwasg Prifysgol Cymru.

Lexical variation in 'néo'-Breton: A corpus-based approach

The field of minority language sociolinguistics has recently seen much work on 'new speakers', particularly in the last five years as a result of the work of the COST New Speakers Network (www.nspk.org.uk). Definitions of the new speaker have varied depending on the context of study, although they are typically defined with reference to "transmission, attitude and origin" (Hornsby, 2015:108). New speakers thus do not acquire the language as a result of uninterrupted intergenerational transmission, instead often doing so through education; they have a positive attitude towards the language; and they need not originate from the traditional speaker community, i.e. the community of those who acquired the language intergenerationally.

While in general terms the 'new speaker' category has emerged only recently, as a development of the 'néo-speaker' theorised by Grinevald and Bert (2011:51), language-specific new speaker categories are more established in academic discourse. One example is the category of the 'néo-bretonnant', or new speaker of Breton. This first came to prominence in the work of Maryon McDonald (1989), who identified a group of "militant" speakers, and was further examined by Mari Jones (1995; 1998). Since then, the body of work dealing with the 'néo-bretonnant' has grown. With this longer-established position in academic discourse, 'néo-bretonnants' have become associated with various attributes: among these, it is claimed that they are typically young, well-educated, mobile, literate in Breton, and highly motivated by the desire to express Breton identity and even political activism, and that they wish to use Breton in as many domains as possible (Jones, 1995:428; Hornsby and Quentel, 2013:75; Rottet, 2014:213). These definitions put them at odds with traditional speakers, who are characterised as older, rurally located, and unable to read and write Breton, and are said to restrict their use of the language to the home and local community (Adkins, 2013:59; ibid:63; Hornsby and Quentel, 2013:75). The language of new speakers has similarly been assigned certain attributes in the literature: it is portrayed as highly standardised and non-dialectal, and while said to be similar to French in terms of its deep linguistic structure, i.e. phonology and morphosyntax, it is also characterised as having an artificially 'purified' lexicon, which makes use of Celtic-based neologisms rather than French borrowings, despite the latter being prevalent in the language of traditional speakers (Hewitt, 2016). These characterisations contribute to the perception of a large linguistic and ideological distance between new and traditional speakers of Breton.

Work on new speakers has generally tended to focus on attitudes and perceptions, often based on interviews and ethnographic techniques. This has been no less the case for Breton, the empirical research carried out by Holly Kennard (2014; 2018) being a rare exception. This work, however, found that new and traditional speaker syntax do not notably differ from each other. Similarly, Ó hIfearnáin (2013) noted that some traditional speakers are well-educated and literate in Breton, and can take up influential positions in Breton language planning, while Adkins (2014) found that adult education classes seek to teach partially dialectal varieties of Breton appropriate to their geographical location, rather than a supradialectal standard. All these findings point to a situation where the gap between new and traditional speakers, and their respective language varieties, may not be as large as is commonly reported.

My work aims to contribute more empirical research to the field in order to investigate this further. This paper presents results of a quantitative analysis of data gathered from Breton-language media, a context associated with new speakers, given their supposed literacy in Breton and greater motivation to speak the language in non-traditional domains, contrasting with traditional speakers. Based on data gathered in 2016–2017 from radio programmes, magazines, and posts on Facebook, this paper focuses on the lexicon, investigating quantities of French borrowings and neo-Celticisms and the presence of dialectal or other non-standard vocabulary, taking into account factors such as

stage of entry into the language and morphological integration, and highlighting some ways in which the lexicon displays variation across the media contexts examined. Such variation attests the heterogeneous nature of Breton in the media, undermining some of the stereotypes about new speakers and their language, particularly the claim that new speakers use only a 'pure', non-dialectal variety.

This paper concludes with an assessment of what such variation could mean for portrayals of Breton in academic discourse and in language policy, stressing the importance of accounting for the multiple motivations and ideologies that operate in contexts where new speakers are present, in order to ensure greater success and sustainability for language maintenance in the future.

References

Adkins, M. (2013). Will the real Breton please stand up? Language revitalisation and the problem of authentic language. *International Journal of the Sociology of Language* 223, 55–70.

Adkins, M. (2014). An American abroad: My adult language immersion studies in Brittany. *Journal of Celtic Language Learning* 19, 109–128.

Grinevald, C. and Bert, M. (2011). Speakers and communities. In P. K. Austin and J. Sallabank (Eds.), *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press, pp. 45–65.

Hewitt, S. (2016). The problem of neo-speakers in language revitalisation: The example of Breton. Paper presented at International Conference on Endangered Languages, Tbilisi State University, Georgia, 20–24 October. Available at https://www.academia.edu/29610759/The_problem_of_neo-speakers_in_language_revitalization_The_example_of_Breton (accessed 9 Nov 2016).

Hornsby, M. (2015). The 'new' and 'traditional' speaker dichotomy: Bridging the gap. *International Journal of the Sociology of Language* 231, 107–125.

Hornsby, M. and Quentel, G. (2013). Contested varieties and competing authenticities: Neologisms in revitalised Breton. *International Journal of the Sociology of Language* 223, 71–86.

Jones, M. C. (1995). At what price language maintenance? Standardisation in modern Breton. *French Studies* 49(4), 424–438.

Jones, M. C. (1998). Death of a language, birth of an identity: Brittany and the Bretons. *Language Problems and Language Planning* 22(2), 129–142.

Kennard, H. J. (2014). The persistence of verb second in negative utterances in Breton. *Journal of Historical Linguistics* 4, 1–39.

Kennard, H. J. (2018). Verbal lenition among young speakers of Breton: Acquisition and maintenance. In C. Smith-Christmas, N. P. Ó Murchadha, M. Hornsby and M. Moriarty (Eds.), *New speakers of minority languages: Linguistic ideologies and practices*. London: Palgrave Macmillan, pp. 231–252.

McDonald, M. (1989). *"We are not French!": Language, culture and identity in Brittany*. London: Routledge.

Ó hIfearnáin, T. (2013). Institutional Breton language policy after language shift. *International Journal of the Sociology of Language* 223, 117–135.

Rottet, K. J. (2014). Neology, competing authenticities, and the lexicography of regional languages: The case of Breton. *Journal of the Dictionary Society of North America* 35, 208–247.

THE POSITION OF MIDDLE IRISH: HISTORICAL LINGUISTICS AND GLOTTOCHRONOLOGY

Traditional historical linguists "don't do dates" (McMahon and McMahon 2006), yet "the issue of dating lies at the core of historical linguistics as the very fact of language change implies 'before' and 'after', 'earlier' and 'later' phases" (Pereltsvaig and Lewis 2015: 92). While 'not doing dates', it only applies to language reconstructions. Dealing with well-attested languages, they usually try not only to determine specific stages or periods of a language's evolution, but also to correlate them with the known crucial historical events.

A new method of 'dating' language changes was proposed in 1950s by M. Swadech (1952, 1955), who examined the changes in the basic vocabulary of a certain language and posited an assumption: the 1000-year retention rate represents 86 %, in other words, 14 words (of the 100-word list!) must be replaced throughout this period. Later this method of dating was modified (Starostin 1999; Blažek 2007) and compromised in Gray and Atkinson (2003; but supported in Renfrew 2013). J. Mallory characterizes this technique as 'often cited but usually rejected' (Mallory 2013: 258). "We cannot help concluding that the *lapis philosophorum* of glottochronology is unable to transmute the baser metals of lexicostatistical numbers into pure gold of reliable dating" (Dolgopolsky 2000: 404).

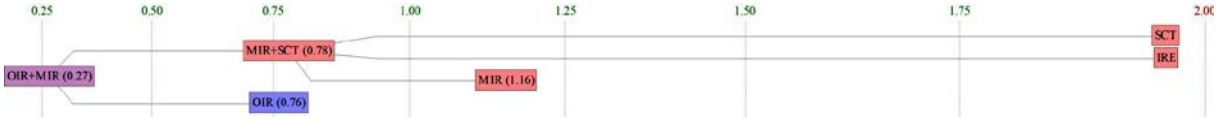
An attempt to pinpoint the time of the split between Goidelic and Brittonic was made in Greene (1964), and later in (Elsie 1979) containing, unfortunately, some inaccuracies (e.c. 'to lie' - in Irish *luím*, in Gaelic - *innis breug*). See also (Fowkes 1971). In Blažek and Novotná (2006) the split between Goidelic and Brittonic is dated c. 1200 BC. The authors use a new calibration, shifting the constant of disintegration λ from 0,14 to 0,05 per millennium, eliminating loanwords and adding synonyms to the list. The use of synonyms compromises Swadech's idea of the **basic** vocabulary and automatically results in its artificial archaisation (thus, the change of λ). Good for 'exotic' languages whose genetic kinship is not clear, this method is hardly fit for 'historical' languages from well-known language families. In this case, a linguist is only able to pinpoint the historical nodes of language divergence more accurately. So, glottochronology, rather than being a supplement to lexicostatistics, could be seen as a self-valuable field that not only does "do dates" but can describe mechanisms of semantic changes.

Middle Irish, traditionally associated with the Viking invasion, is described in detail in regard to its phonology, morphology, verbal system and syntax (copula etc). See (Breatnach 1994: McCone 2005, including references). Semantic development of the Irish lexica, changes in its vocabulary received little attention from linguists (with the exception of Latin and Norse loanwords). The analysis of changes in the Irish **basic** word-list could be a step towards filling this gap. The arithmetic average of all shared words between OI (VII-VIII c.) and late MI (XII c.?) is 0,90, so we have 10 (problematic and supposed) changes: bark: *rúsk* > *coirt*; to die: *at-baill* > *téit* (*do*) *éc*; ear: *ó* > *cluas*; egg: *óg* > *ubh*; meat: *carna* > *feóil*; red: *ruad* > *derg*; star: *rind* > *retla*; tooth: *dét* > *fiacail*; tree: *bile* > *crann*; to walk: *téit* > *siblaid*. It doesn't mean of course that all these lexemes were not in use in the MI period (and later), but they lost their basic status. Each change deserves discussion and explanation based on cultural changes.

The 90 % retention rate corresponds to the period of 720 years (according to the Swadech λ-constant of disintegration) or of 2000 years (according to the Blažek-Starostin recalibration!). A simple idea - that the glottochronology technique automatically extends MI period to the

XVI c. - is demonstrably false, not because it conflicts with the traditional periodization of the history of Irish, but because of later new changes in its basic vocabulary (*madra*, *gruaig*, *gealach*, *bothar* etc.).

As a preliminary conclusion: Middle Irish is not a 'stage' or a 'period' in the history of the Irish, but represents a specific variety of the written language preserved in manuscripts. Both Modern Irish and Modern Gaelic do not descend from MI but go back to oral dialectal (?) forms of the language continuum. MI, as well as OI, are rather "dead-end branches": see the table -



For a new dating of the beginning of MI - cf. McCone 1985!

References

Blažek, V. (2007) 'From August Schleicher to Sergei Starostin: On the development of the tree-diagram models of the Indo-European languages' *JIES*, 35, P. 82-110.
Blažek, V. and Novotná, P. (2006) 'On application of Glottochronology for Celtic Languages' *Linguistica Brunensia*, A 54, P. 71-100.
Breatnach, L. (1994) 'An Mheán-Ghaeilge' I K.McCone, D. McManus agus eile (eag.) *Stair na Gaeilge*. Maigh Nuad, L.61-220.
Dolgopolsky, A.B. (2000) 'Sources of linguistic chronology'. In: C. Renfrew, A. McMahon and L. Trask (eds.) Time Depth in Historical Linguistics. Cambridge, UK, P. 401–409.
Gray, R.D. & Atkinson Q. (2003) 'Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin' *Nature*, 426, P. 435-439.
Greene, D. (1964) 'The Celtic Languages' In J.Raftery (ed.) *The Celts*. Cork: Mercier, P. 9-21.
Fowkes, R.A. (1971) 'Glottochronology and Brythonic?' *Studia Celtica*, VI, P. 189-194.
Mallory, J. (2013) *The Origins of the Irish*. Thames & Hudson.
McCone, K. (1985) 'The Würzburg and Milan Glosses: Our earliest sources for MI' *Ériu*, 36, P. 85-106.
McCone, K. (2005) *A First Old Irish Grammar and Reader, including an Introduction to Middle Irish*. Maynooth.
McMahon, A. and McMahon, R. (2006) 'Why linguists don't do dates: evidence from Indo-European and Australian languages'. In: P. Forster and C. Renfrew (eds.) *Phylogenetic Methods and the Prehistory of Languages*. Cambridge, UK, P. 153-160.
Pereltsvaig, A. and Lewis, M.W. (2015) *The Indo-European Controversy. Facts and Fallacies in Historical Linguistics.* Cambridge University Press.
Renfrew, C. (2013) 'Early Celtic in the West: The Indo-European Context' In J.T. Koch and B. Cunliffe (eds.) *Celtic from the West 2*. Oxford and Oakville: Oxbow Books, P. 207-218.
Starostin, S. (1999) 'Comparative-historical linguistics and lexicostatistics'. In V. Shevoroshkin & P.Sidwell (eds.) *Historical Linguistics & Lexicostatistics*. Melbourne, P. 3-50.
Swadech, M. (1952) 'Lexico-statistic dating of prehistoric ethnic contacts' *PAPS*, 96, P. 452-463.
Swadech, M. (1955) 'Towards greater accuracy in lexicostatistic dating' *IJAL*, vol. 21, P. 121-137.

The interplay of syntax and stress in Old Irish: Evidence for indirect reference

The complex interplay of syntax and stress in the Old Irish (ca. 750–900 CE), in particular in the verbal complex, has long been a crux in both traditional historical studies on Celtic and Indo-European languages (e.g. Watkins 1963, Meid 1963, Cowgill 1975, McCone 1996, 2006) and in more theoretically informed linguistics (e.g. Carnie, Pyatt and Harley 1994, 2000, Adger 2001, Newton 2006). Although the evidence is primarily historical, it is descriptively quite clear and, moreover, supported by comparison with Modern Irish dialects (Thurneysen 1980). In this paper I advance an analysis of the complex situation in the Old Irish verbal complex which is in line with theories proposing that the phonology–syntax interface is mediated through phrasal prosodic constituents (e.g. Chen and Downing 2016) rather than phonological information being directly encoded in the syntactic structure.

In Old Irish stress is regularly assigned to the initial syllable of a word (Thurneysen 1980). The main exception to this general rule involves finite compound verbs (CV), containing one or more prefixes (P), when occurring in initial position in main clauses. In this case the leftmost P is unstressed but the stress falls on the verb (V), and the CV assumes a "deuterotonic" form, e.g. *do-beir* 'brings' (P-'V). In other contexts, however, the stress in CVs is on P, giving a "prototonic" form in accordance with the usual pattern, e.g. *tabair* 'brings' ('P-V). Note that Old Irish is a VSO language, finite verbs normally occurring initially in both main and embedded clauses.

It has been claimed that the exceptional deuterotonic stress of CVs in Old Irish is due to a "prosodic juncture" after an element (a complementizer, P or V) in the highest functional head (C) in the clause (Eska 1996, Carnie et al. 1994, 2000, Adger 2001, Newton 2006). However, this analysis is problematic in view of the existence of CVs exhibiting the regular prototonic stress, e.g. 3 person imperative *taibred* 'let him bring' ('P-V).

I propose that the deuterotonic stress in CVs results from the effects of enclitic Wackernagel elements (pronouns/agreement markers, particles) between P and V; crucially, this includes morphologised historical reflexes of such clitics, synchronically manifested as covert function particles (proposed in the "Particle Theory" of Cowgill 1975; cf. also Sims-Williams 1984, Schrijver 1997). Expanding on earlier suggestions (Watkins 1963, Isaac 1993), I argue that these clitics block the stress assignment on P in absolute clause-initial position.

The placement of Wackernagel clitics between P and V ("infixation", cf. e.g. Griffith 2015) would seem to violate the lexical integrity of the word (Di Sciullo and Williams 1987, Harris 2000). However, I adopt the view that a CV arises by incorporation of P into V. I assume that P is a stress neutral function word, normally incorporating as a proclitic into the host prosodic word (pword, ω) (Selkirk 1995, Green 2001). When CVs are not in absolute initial position, notably in embedded clauses, their structure is [ω('P-V)], the stress in the complex head being assigned initially according to the general rule. But when P combines with a Wackernagel element (E) in main clauses, the resulting complex P-E constitutes an affixal proclitic outside the host pword, and the entire structure of the CV is to be analyzed as a recursive pword [ω(P-E ω('V))]. Only the right-hand member of this recursive pword receives the stress. The exceptions involving prototonic imperative forms to CVs are due to the fact that they do not contain a Wackernagel element.

In conclusion, CVs in Old Irish exhibit both [ω(Clitic+Host)] and [ω(Clitic ω(Host))] structures, the former appearing as prototonic forms and the latter as deuterotonic forms. The conditions on the formation of the different structures and their respective stress assignment depend on the syntactic position of the CV and the placement of Wackernagel clitics. The analysis sustains

the predictions of indirect reference theories such as Chen and Downing (2016) that the phonology–syntax interface is mediated through constituents in the Prosodic Hierarchy rather than being directly conditioned by syntactic structure.

References

Adger, D. 2001. 'V1 phenomena and the syntax prosody interface.' In R. Billeray and B. D. Lillehaugen (eds) *Proceedings of WCCFL* 19. Somerville: Cascadilla Press. 1-14.

Carnie, A., H. Harley and E. Pyatt. 1994. 'The Resurrection: Raising to Comp, Evidence from Old Irish.' *Studies in the Linguistic Sciences* 24,1/2:85-100.

Carnie, Andrew, H. Harley and E. Pyatt. 2000. 'VSO order as Raising to out of IP.' In A. Carnie and E. Guilfoyle (eds) *The Syntax of Verb Initial Languages*. Oxford University Press. 39-60.

Chen, Lisa Lai-Shen and Laura J. Downing. 2016. 'Phasal Syntax = Cyclic Phonology?' *Syntax* 19,2:156–191.

Cowgill, W. 1975. 'The origin of the Insular Celtic conjunct and absolute verbal inflections.' In H. Rix (ed.) *Flexion und Wortbildung: Akten der V. Fachtagung der indogermanischen Gesellschaft.* Wiesbaden: Harrassowitz. 40-70.

Di Sciullo, A.-M., and E. Williams. 1987. *On the definition of word*. Cambridge MA: MIT Press.

Eska, J. 1996. 'On the syntax and phonology within the early Irish verbal complex.' *Diachronica* 13:225-257.

Green, A. D. 2001. 'The prosodic representation of clitics in Irish.' In B. Gerlach and J. Grijzenhout (eds) *Clitics in phonology, morphology and syntax*, Amsterdam: John Benjamins. 181-218.

Griffith, A. 2011. 'Old Irish pronouns: agreement affixes vs. clitic arguments.' In A. Carnie (ed.) *Formal approaches to Celtic linguistics*. Newcastle upon Tyne: Cambridge Scholars Publishing. 65-94.

Harris, A. 2000. 'Where in the word is the Udi clitic?' *Language* 76,3:593-616.

Isaac, G. 1993. 'Issues in the reconstruction and analysis of Insular Celtic syntax and phonology.' *Ériu* 44:1-22.

McCone, K. 1996. *The early Irish verb*. 2nd ed. Maynooth: An Sagart.

McCone, K. 2006. *The origins and development of the insular Celtic verbal complex*. Maynooth: Department of Old Irish, National University of Ireland.

Meid, W. 1963. *Die indogermanischen Grundlagen der altirischen absoluten und konjunkten Verbalflexion*. Wiesbaden: Harrassowitz.

Newton, Glenda. 2006. The development and loss of the Old Irish double system of verbal inflection. PhD dissertation, University of Cambridge.

Schrijver, P. 1997. *Studies in the history of Celtic pronouns and particles*. Maynooth: The Department of Old Irish, National University of Ireland.

Selkirk, E. 1995. The prosodic structure of function words. In J. Beckman, L. Walsh Dickey, and S. Urbanczyk *Papers in Optimality Theory*. Amherst, MA: GLSA Publications. 439-70.

Sims-Williams, P. 1984. 'The double system of verbal inflexion in Old Irish.' *Transactions of the Philological Society* 82:138-201.

Thurneysen, R. 1980. *A grammar of Old Irish: an enlarged edition with supplement.* 2nd ed., translated and revised by D. A. Binchy and O. Bergin. Dublin: Dublin Institute for Advanced Studies.

Watkins, C. 1963. 'Preliminaries to a historical and comparative analysis of the syntax of the Old Irish verb.' *Celtica* 6:1-49.

On the use of the Old Irish indefinite pronoun *nech ní*

In this paper, I would like to address a number of issues related to the use of the Old Irish indefinite pronoun *nech ní* in the corpus provided by the main collections of glosses attested in the contemporaneous manuscripts. Due to the quality and quantity of the involved texts, Wb (c. 160 cases) and Ml (c. 250 cases) are especially valuable in this regard; less so the Sg (c. 27 cases) evidence. The different value of these collections is much the same as in the case of the use of the neuter singular demonstrative pronouns *aní* and *a*^{N}, which are characterized by the fact that they constitute a NP together with the relative clause which regularly follows. One of the questions which I would like to deal with in this paper is precisely the difference between the use of these (definite) demonstratives and that of the mentioned indefinite pronoun, one of whose main uses is precisely as the head of a relative clause.

The use of this indefinite pronoun has been analyzed from the point of view of its semantic uses (the functions considered by M. Haspelmath 1997, *Indefinite Pronouns*), gender, pragmatic use, syntactic context (namely, the combination with negation or an associated element, as well as the inclusion in a main or subordinate clause), syntactic function (A, S, O, oblique or genitive), as well as the presence of a relative clause depending on this indefinite pronoun (which has a NP function in this relative clause). A preliminary glance at the results of this analysis permits to advance a number of worth discussing observations:

1. The Old Irish *nech ní* apparently has a wide range of semantic uses, from the "specific known to the speaker" to the "free choice" uses, to quote the two extreme poles of the semantic map proposed by Haspelmath, so that one must say that this Old Irish indefinite pronoun possesses a remarkable quantity of functions, even for an cross-linguistically multifunctional element like this.
2. The Old Irish indefinite pronoun can be combined with both negation and a relative clause (and can also be used without any of these elements), but a certain complementary distribution of negation and relative clause can be observed.
3. A very remarkable fact which clearly calls for an explanation is the fact that the indefinite pronouns which have the function of A (agent of a transitive predicate) virtually have no relative clause (neither in Wb nor in Ml), whereas the indefinite pronouns with other functions (namely, S, O and oblique) have a relative clause in more or less the same proportion (i.e. in the half of the cases).
4. As is also the case of the previous observation, the use of the aforementioned neuter singular demonstrative pronouns *aní* and *a*^{N} contrasts with the corresponding indefinite pronoun *ní*: whereas *aní* and *a*^{N} most often express the combination of NP functions S – O, this combination is precisely the less frequent in the equivalent use of *ní* in both Wb and Ml.

The adduced list does not exhaust the number of observations and questions which can be put forward on the use of the Old Irish indefinite pronoun *nech ní*. However, it shows the potential of the language of the glosses for a productive investigation on the syntactic use of relevant elements and perhaps even the potential of the Old Irish linguistic evidence for general linguistic issues.

Ten New Indo-European Etymologies for Old Irish

ABSTRACT: The Proto-Indo-European etymology of the cognates, i.e. providing correspondences between two or more Indo-European morphemes belonging to different subgroups and thus enabling the reconstruction of their original prototype, remains a primary task of Indo-European linguistics.

The paper at hands contributes to the solution of the comparative etymology of Old Irish by means of presenting twelve new Indo-European etymologies from the rest of the group. The etymologies discussed are:

1. OIr. áith 'pinna: wing': RV. átia- 'eilend, rennend'
2. OIr. nēire 'seer, diviner': OInd. nayana- 'Auge'
3. OIr. tlí 'protection, force, récomfort': RV. trā́- 'beschützen, behüten, retten'
4. OIr. ate 'en verité, certes': RV. addhā́ 'fürwahr'
5. OIr. cēcht 'Macht': Gr. κῖκυς 'Kraft, Stärke'
6. OIr. all 'bride, rêne': Thess. ἄλλιξ, ἄλλικα 'χλαμύς'
7. OIr. ness- 'weasel': AV. nak·ulá- 'Ichneumon, Viverra ichneumon'
8. OIr. dīn- 'covering, thatch, sparing': Lith. dẽ̃ni- '(Ver)Deck, Schiffsdeck'
9. OIr. nāt 'posterior, arse': Goth. notin- 'Schliffshinterteil, Achtedeck' (πρύμνη)
10. OIr. aⁱnmīan 'passion, lust, concupiscence': TochB. añme 'wish, desire'

References (incomplete sample)

BARTHOLOMAE, Christian. 1904. Altiranisches Wörterbuch. Straßburg: Trübner.

FRAENKEL, Ernst. 1962-65. Litauisches etymologisches Wörterbuch I-II. Heidelberg & Göttingen: Winter, Vandenhoeck & Ruprecht

FRISK, Hjalmar. 1960-722. Griechisches etymologisches Wörterbuch I-III. Heidelberg: Winter.

GRASSMANN, Hermann. 19966. Wörterbuch zum Rig-Veda, 6. überarbeitete und ergänzte Auflage von Maria Kozianka. Wiesbaden: Harrassowitz.

LEHMANN, Winfred, P. 1986. A Gothic Etymological Dictionary Based on the Third Edition of Vergleichendes Wörterbuch der gothischen Sprache by Sigmund Feist. Leiden: Brill.

MARSTRANDER, Carl J. S. et al. 1913ff. Dictionary of the Irish language. Dublin: Royal Irish Academy.

MAYRHOFER, Manfred. 1956-1980. Kurzgefaßtes etymologisches Wörterbuch des Altindischen I-IV. Heidelberg: Winter.

----- 1986-2000. Etymologisches Wörterbuch des Altindoarischen. Heidelberg: Winter.
POKORNY, Julius. 1959. Indogermanisches etymologisches Wörterbuch I. Tübingen: Francke
TISCHLER, Johann. 1977-2016. Hethitisches etymologisches Glossar. Mit Beiträgen von Günter Neumann und Erich Neu. (IBS 20). Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
UNTERMANN, Jürgen. 2000. Wörterbuch des Oskisch-Umbrischen. (Handbuch der Italischen Dialekte, III.) Heidelberg: Winter.
VENDRYES, Joseph & BACHELLERY, Edouard & LAMBERT, Pierre Y. 1959ff. Lexique étymologique de l'irlandais ancien. Dublin: Dublin Institute for Advanced Studies.
WALDE, A. & HOFMANN, J.B. 1938. Lateinisches etymologisches Wörterbuch. 3., neuarbeitete Auflage von J. B. Hofmann I-II. Heidelberg: Winter.

IRISH CONSTRUCTIONS OF METAPHORIC GIVING

The paper presents an analysis of a type of constructions with the light verb *tabhair* 'give' and two prepositions – *do* 'to' and *ar* 'on'. A fully lexical *tabhair* 'give' combines with the preposition *do* 'to' introducing the beneficiary – the DATIVE CONSTRUCTION:

tabhair bronntanas do Mháire
give present to Máire
'give a present to Máire'.

As a light verb, however, it combines not only with *do* but also with *ar* 'on' which is not predicted by the meaning of *tabhair* and makes these combinations immediately idiomatic. The following table presents some predicate nouns that combine with *tabhair do* and *tabhair ar*:

Examples of *tabhair* with *do* 'to'	**Examples of *tabhair* with *ar* 'on'**
aire 'care', *cúnamh* 'help', *tacaíocht* 'support', *suntas* 'notice', *cead* 'permission'	*aird* 'attention', *aghaidh* 'face', *cuairt* 'visit'
thug Seán tacaíocht do Mháire give.PST Seán support to Máire 'Seán gave support to Máire'	*thug Seán aird ar Mháire* give.PST Seán attention on Máire 'Seán paid (gave) attention to Máire'

The idiomatic nature of both types of constructions can be seen in the following. Assigning semantic roles to the participants can be done on two levels – that of the metaphor's source domain and that of the target domain: at the level of the source domain *Máire* can be said to be an adressee but in the target domain – a beneficial. This means that general Irish syntax only predicts the semantic role metaphorically, whereas the real semantic role can only be assigned if one knows the exact meaning of the *do* + object in this exact construction. This is even more so in the case of the predicate. In the source domain 'care' can be understood to be an object of 'giving'. However, ascribing it a true semantic role in this sentence is problematic as the predicate noun does not refer to any real object but to the situation itself. The predicate noun in the target domain should therefore be analysed as a part of the syntactic predicate, forming a complex predicate *thabhair* 'give' + *tacaíocht* 'support'.

source domain		AGENT	THEME		RECIPIENT
		↓	↓		↓
target domain	PREDI-	AGENT	-CATE		BENEFICIARY
	thug	*Seán*	*tacaíocht*	*do*	*Mháire*
	give.PST	Seán	support	to	Máire
	'Seán gave support to Máire'				

This shows that even though formally the construction in question resembles the dative construction, to adequately understand the sentence above one has to have the construction with a predicate noun stored in one's lexicon as a separate unit, if not for decoding the meaning of the sentence (which might be assumed to be decodable with little effort), but at least to know how such meanings are encoded in the language. This is what was called 'encoding idioms' in Fillmore et al. 1988: they seem to be compositional, e.g. *answer the door*, but are arbitrary or conventional for a particular meaning (Fillmore et al. 1988: 504-5).

The degree of encoding idiosyncrasy is even higher in the case of the construction *tabhair* + *ar* 'on':

source domain		AGENT ↓	THEME ↓		?
target domain	PREDI-	AGENT	-CATE		GOAL?/BENEFICIARY?
	thug	*Seán*	*aird*	*ar*	*Mháire*
	give.PST	Seán	attention	on	Máire

'Seán paid (gave) attention to Máire'

The knowledge of this construction involves not only the form itself of how this meaning is expressed in Irish, but also a semantically unpredictable preposition governing the indirect object of *tabhair* 'give' so that it is impossible to assign the object a semantic role at the source domain level. However, it is quite difficult to assertively assign a semantic role to the indirect object at the level of the target domain either, which questions the plausibility of a semantic-role analysis of such constructions given the abstract nature of the situations they describe. The metaphor itself should be then seen as a tool to grasp these abstract notions in terms of more concrete situation models along the lines of Lakoff 1993. In this case the target domain level of semantic roles could be dispensed with. The semantics of the constructions would then include the semantic roles of the participants in the source domain situation and then metaphorically mapped onto the target domain situation.

Another question concerns the use of the preposition *ar* 'on' in combination with the verb *tabhair* 'give' which does not occur with non-metaphoric use of *tabhair*. I suggest that the the combination of *tabhair* and *ar* is an instantiation of a more abstract construction [ATTENTION PREDICATE *ar* OBJECT OF ATTENTION]. Some other instantiations of this construction include, for example:

- *breathnaigh/féach/amharc* 'look' *ar X* 'look at X';
- *cuir* 'put' *aithne, eolas* 'knowledge' *ar X* 'learn to know X'.

It can therefore be argued that the *TABHAIR AR* CONSTRUCTION inherits its form from a more abstract CONSTRUCTION OF ATTENTION [ATTENTION PREDICATE *ar* OBJECT OF ATTENTION], which allows for the use of the verb *tabhair* 'give' with the preposition *ar* 'on', a combination that would otherwise be difficult to account for.

The two constructions of metaphoric giving both show traits of idiomaticity. However, they differ in that the *TABHAIR DO* CONSTRUCTION allows for a more straightforward mapping of source domain semantic roles onto the participants of the target domain situation. This is more complicated in the case of the *TABHAIR AR* CONSTRUCTION, in which the verb *tabhair* 'give' unusually for the basic, literal meaning of the verb combines with the preposition *ar* 'on'. This construction is argued to be a result of the use of a light verb *tabhair* with predicate nouns in a more abstract CONSTRUCTION OF ATTENTION, where the predicate slot can be occupied by single-unit verbs or by light-verb constructions.

References

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. Language, 501-538.
Lakoff, G. (1993). The syntax of metaphorical semantic roles. In Semantics and the Lexicon (pp. 27-36). Springer Netherlands.

What is a revived language?
Linguistic features of Revived Manx in comparison with the traditional language

As minority language communities worldwide increasingly face the cessation of intergenerational transmission and the process of language shift and death, the practice of 'language revival', meaning the revernacularization of 'dead', 'dormant' or 'sleeping' languages attested in written texts, recordings and linguistic documentation, is becoming more prominent, and the study of it more urgent. In addition to the most well-known and, in numerical and political terms, most successful example of Hebrew (Zuckermann & Walsh 2011), a number of examples of revival of indigenous languages in North America (e.g. Miami and Wampanoag), Australia (e.g. Kaurna) and East Asia (e.g. Siraya) have been recently documented, with efforts often led or guided by trained linguists, sometimes from within the indigenous ethnic group itself. Older movements, such as those to revive the Celtic languages Cornish and Manx, which both go back over a century, have mostly been shaped by gifted amateurs.

Within the Celtic languages, the topic of language revival, as opposed to revitalization of still-existing speech communities, is becoming more significant as traditional speech communities are rapidly becoming moribund. Networks of second-language speakers, for example in the cities and towns of Ireland and Brittany, can often be seen as cases of language revival, since their links to traditional native speakers are often tenuous (cf. the ground-breaking study of a Belfast community in Maguire [1991]). It has been argued that such 'new speakers' increasingly represent the future of these languages (O'Rourke & Walsh 2015; Hornsby & Quentel 2013), and that new varieties or dialects may be developing in these networks (Ó Broin 2014; Snesareva 2017).

In this context, the study of Revived Manx has a good deal to offer both to Celtic linguistics and to the wider fields of language revitalization and 'revival linguistics' (Zuckermann & Walsh 2011: 122). Although a small number of studies have examined certain linguistic features of the revived variety of Manx (e.g. Clague 2004–5; Kewley Draskau 2005; Broderick 2013), hitherto there has been no general overview or description of the linguistic features of the revived language and how it has developed, and diverged, from the natively-spoken variety, which is usually considered to have become extinct in 1974.

Other studies have examined the language movement from an ethnographic, anthropological, sociological or historical perspective, including speakers' language ideologies (e.g. Ó hIfearnain 2015). However, a lack of engagement with formal linguistic data from Revived Manx speakers, and the evidence of the Traditional Manx corpus, means that scholars have not fully examined the complex interplay between language ideology, historical circumstances, the second language acquisition process and the linguistic and orthographic features of Manx, as well as the nature of the long-standing contact relationship between Manx and English. In the present paper, a variety of phonological, morphosyntactic and lexical features will be presented, showing how they illustrate various factors which can be seen to mould speakers' (and writers') usage. These include:

- Substratal influence from English, the first language of almost all Revived Manx speakers, including features consciously avoided but frequently occurring, as well as features of which most speakers seem to be unaware.
- Purism and 'hyper-Gaelicism', whereby attested structures are replaced with those perceived, sometimes spuriously, to be more native or more Gaelic.
- 'Hyper-archaisms', including restoration of forms to those considered etymologically 'correct', even when these may in fact have been ungrammatical in attested periods of the traditional language.

- Spelling pronunciations, based on misinterpretation of the complex and inconsistent Manx orthography.
- Internal analogy, overgeneralization and simplification.
- Erroneous, ambiguous or incomplete information in dictionaries, grammars, language courses, and the usage of prominent individuals or organizations.
- Widespread use of neologisms which may be seen as 'out of character' from the perspective of the traditional language, especially those based on Irish.
- Mixing of dialects, registers and historical periods of the language.

Evidence from these features is taken primarily from a corpus of recent literary and pedagogical publications, as well as videos of interviews with Revived Manx speakers available on Youtube.

Manx, like most revived languages, but unlike the usually-cited textbook example of Hebrew, remains largely a second language, moulded by conscious learning by relatively small numbers of adults in each generation, rather than being subject to the usual unconscious processes of language change during intergenerational transmission. In this respect it is perhaps a more typical example of what to expect when a second language is maintained over the long term in small networks of enthusiasts and activists. This has lessons for those considering the future prospects of the other Celtic languages, as well as endangered languages elsewhere, in terms of corpus and acquisition planning, planning language ideology (Armstrong 2012), documentation of surviving native varieties, and preparation of pedagogical resources.

References:

Armstrong, Timothy Currie. 2012. 'Establishing new norms of language use: The circulation of linguistic ideology in three new Irish-language communities', *Language Policy*, 11: 145–68

Broderick, George. 2013. 'The revival of Manx Gaelic in the Isle of Man', *Scottish Gaelic Studies*, 29: 132–71

Clague, Marie. 2004–5. 'Cross-Linguistic Discourse Markers in Manx Gaelic and English', *Proceedings of the Harvard Celtic Colloquium*, 24/25: 195–205

Hornsby, Michael and Gilles Quentel. 2013. 'Contested varieties and competing authenticities: neologisms in revitalized Breton', *International Journal of the Sociology of Language*, 223: 71–86

Kewley Draskau, Jennifer. 2005. 'Language Death and Resurrection in the Isle of Man: the continuity of Manx Gaelic Exemplified by the Use of Inflected Verb Tenses', *Proceedings of the Harvard Celtic Colloquium*, 25: 229–46

Maguire, Gabrielle. 1991. *Our Own Language: An Irish Initiative* (Clevedon: Multilingual Matters)

Ó Broin, Brian. 2014. 'New Urban Irish: Pidgin, Creole, or Bona Fide Dialect? The Phonetics and Morphology of City and Gaeltacht Speakers Systematically Compared', *Journal of Celtic Linguistics*, 15: 69–91

Ó hIfearnáin, Tadhg. 2015. 'Sociolinguistic vitality of Manx after extreme language shift: authenticity without traditional native speakers', *International Journal of the Sociology of Language*, 231: 45–62

O'Rourke, Bernadette and John Walsh. 2015. 'New speakers of Irish: shifting boundaries across time and space', *International Journal of the Sociology of Language*, 231: 63–83

Snesareva, Marina. 2017. 'Palatalisation in Dublin Irish, or how to speak Irish with a Dublin accent', *Studia Celtica Posnaniensia*, 2: 63–80

Zuckermann, Ghil'ad and Michael Walsh. 2011. 'Stop, Revive, Survive: Lessons from the Hebrew Revival Applicable to the Reclamation, Maintenance and Empowerment of Aboriginal Languages and Cultures', *Australian Journal of Linguistics*, 31: 111–27

Three birds with one stone: A methodological triangulation to gain insight into Scottish Gaelic morphosyntax

Abstract

Linguistic variation in Scottish Gaelic is a vibrant field of study, and impressive and comprehensive projects have sought to understand linguistic and sociolinguistic phenomena found in Gaelic, most notably the Linguistic Survey of Scotland (Gaelic) (LSS(G)) between 1951 and 1963, and Dorian's work in East Sutherland since the 1970s. Much of the interest in regional variation in Gaelic has been concerned with phonetic description or phonological analysis, e.g. the published volumes of the Survey of the Gaelic Dialects of Scotland (SGDS) (Ó Dochartaigh 1997). However, much of the work has overlooked morphosyntax (see Iosad and Lamb 2016). As well as morphosyntax proving to be elusive in the literature on variation in Gaelic, I identified two other immediate concerns for Gaelic linguistics that could be incorporated into my methodology: (1) Bell et al. (2014) stipulated that the description of the traditional vernacular Gaelic of older speakers was an immediate priority for corpus development, and (2) documentation methods have developed, and another generation has grown up and reached senior years since the last meaningful dialectology project, LSS(G). Therefore, I sought to address these gaps in Gaelic scholarship by undertaking fieldwork in today's traditional speaker heartland (the Hebrides), which made use of a triangulation of linguistic research methods: documentation, dialectology, and variationist sociolinguistics. In this paper, I will describe how the methods I used in my doctoral research capture data on morphosyntactic processes and how they mark morphosyntactic function in Gaelic, while simultaneously documenting natural speech and providing data that can be compared directly with the data in LSS(G). I will focus on nominal case morphology (including attributive adjectives), providing a preliminary analysis of the data. I will discuss whether the data suggests that morphosyntactic variation emits a geographic signal and/or indicates changes, and where this will take my doctoral research.

Keywords

Scottish Gaelic, morphology, syntax, sociolinguistics, variation, dialectology, documentation

References

Bell, Susan, Mark McConville, Wilson McLeod, and Roibeard Ó Maolalaigh. 2014. *Dlùth is Inneach – Linguistic and Institutional Foundations for Gaelic Corpus Planning*. University of Glasgow, U.K.: Soillse.

Dorian, Nancy C. 1973. "Grammatical change in a dying dialect." *Language* 49 (2): 413-438.

Dorian, Nancy C. 1978. *East Sutherland Gaelic: the dialect of the Brora, Golspie, and Embo fishing communities*. Dublin: Dublin Institute for Advanced Studies.

Dorian, Nancy C. 1981. *Language death: The life cycle of a Scottish Gaelic dialect*. Philadelphia: University of Pennsylvania Press.

Dorian, Nancy C. 2010. *Investigating variation: the effects of social organization and social setting*. Oxford: Oxford University Press.

Iosad, Pavel, and William Lamb. 2016. "Morphology and dialectology in the Linguistic Survey of Scotland: A quantitative approach." Rannsachadh na Gàidhlig, Sabhal Mòr Ostaig, Isle of Skye, 24 June 2016.

Ó Dochartaigh, Cathair, ed. 1997. *Survey of the Gaelic dialects of Scotland: questionnaire materials collected for the Linguistic Survey of Scotland*. Dublin: School of Celtic Studies, Dublin Institute for Advanced Studies.