



PROJECT MUSE®

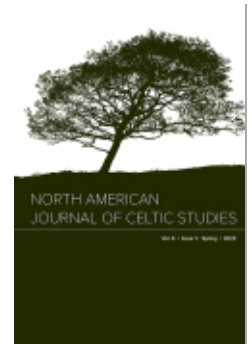
Examining linguistic layers in the St. Gall glosses, again

Aaron Griffith

North American journal of Celtic studies, Volume 6, Issue 1, Spring
2022, pp. 143-160 (Article)

Published by The Ohio State University Press

DOI: <https://doi.org/10.1353/cel.2022.0004>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/853485>

Examining linguistic layers in the St. Gall glosses, again

AARON GRIFFITH

ABSTRACT: The St. Gall Priscian glosses present a well-known puzzle for scholars interested in the language and composition of the corpus. While we know that the glosses were copied, closer examination reveals that they were copied from a variety of sources of different ages. That is, some forms are older, while others are younger. Given the compilatory nature of glossing, it is impossible to find sections of text that are more or less archaic. The sections of text would have to be so small that they would consist of individual glosses scattered through the whole corpus. Thus far, no one has been able to suggest any way around this problem.

This paper offers a proposal that may allow us to bypass this issue by creating a profile for every gloss, based on its linguistic and palaeographic ‘signature’. Once the ‘signatures’ are made, they can be compared with one another and grouped via computer algorithm. These grouping of glosses, with some good fortune, may allow us to detect the layers behind the St. Gall glosses, layers which up to now have eluded us.

KEYWORDS: Old Irish, St. Gall, glosses, orthography, historical phonology, cluster analysis

Aaron Griffith [a.griffith@uu.nl] is Assistant Professor in Celtic languages and culture at Utrecht University, where he teaches on all aspects of early Celtic, especially Old Irish and Middle Welsh languages and linguistics. His research focusses on the synchronic and diachronic grammars of the medieval Insular Celtic languages and on language change more broadly.

North American journal of Celtic studies Vol. 6, No. 1 (2022)
Copyright © 2022 by The Ohio State University

0. The research plan and its broader scientific background

In what follows, I will offer a plan of action outlining a possible way to untangle the complicated multilayered glossing present in the St. Gall Priscian. A ‘plan for action’, as opposed to a solution or partial solution, may sound underwhelming. On the other hand, given that the proposed project involves a significant investment of time, introducing it now and opening it up to feedback at this stage is both logical and sensible, as this will improve the end-product. It also happens to fit in with some broader trends in the scientific community.

Though the issues are much older, there has been an explosion of concern in the last ten years about how scientific work is conducted and disseminated. Some of the key concerns are the lack of reproducibility (i.e., a robust, scientifically valid result should be verifiable by a different researcher using different methods; this is frequently not the case);¹ the publication solely of positive results (i.e., what did not work is frequently not reported, let alone published), and a lack of academic integrity (i.e., data is sometimes ‘massaged’ or even fabricated so that interesting results can be reported; or various analyses not originally planned are undertaken in the search for positive results). The last two categories are most relevant here. The benefit to other scholars of knowing what did not work should be obvious: if one knows what did not work, one will not waste time trying the same thing oneself. As part of the debate on academic integrity, there have been increasing calls to pre-register experiments. The idea is that methods should be fixed BEFORE the study is performed. This makes the experiment part of the public record and protects against ‘being scooped’ by rival researchers, but it also prevents researchers from ‘fudging’ the experiment in an attempt to obtain better or more interesting results. While financial and competitive pressures that lead to these problems are more prevalent in medicine and the natural and social sciences, the issues are also relevant for the humanities and, thus, Celtic studies. I do not intend to linger on this aspect any further here, but I do feel it is important to situate this paper, which serves essentially as a pre-registration of an academic study, within the broader debate on open science and good scientific practice.

A version of this paper was read in September 2020 at the virtual Workshop on Glossing in Medieval Celtic Contexts which was hosted by Virginia Tech and organised by Charlene Eska, Paul Russell, and Joe Eska. I would like to thank the workshop participants for their very useful comments on the oral presentation of this project. The written paper also received a significant amount of critical commentary from two reviewers, and I am quite indebted to them for their work. The usual disclaimers apply.

¹. There are actually more terms in play than simply this. For experimental sciences, repeatability (same team, same experimental set-up), replicability (different team, same experimental set-up), and reproducibility (defined as above: different team, different experimental set-up) are all relevant in slightly different ways. For the definitions, see Association for Computing Machinery 2020. For an overview of some debate on the use of the terminology in one sub-field, see Plesser 2018; more generally in National Academies of Science, Engineering, and Medicine 2019: esp. 39–54.

1. St. Gall Priscian glosses. The history of teasing apart the linguistic layers

The grammar of Old Irish as found in Thurneysen 1946 is based on sources ‘preserved in more or less contemporary manuscripts’ (4 §4). Principal among these manuscripts are marginal and interlinear glosses of Latin material, and the Old Irish glosses on Priscian found in St. Gall (Stiftsbibliothek ms. 904), together with the Würzburg and Milan glosses, are the largest such collections.² While the writing of the St. Gall manuscript and the copying of the glosses into it have been quite precisely dated to 850–851 (Ó Néill 2000), the glosses themselves give contradictory information as to their date of composition. Lash 2017: 148–149 notes, correctly, that a claimed seventh-century layer of glosses (see Hofman 1996: 45) has not been properly substantiated. The glosses represent ‘at least partly, if not mostly, the language of the ninth century’ (Lash 2017: 149). The material that is not in the language of the ninth century must be earlier, though it is unclear how much so. Regardless, it is clear that there are multiple layers of glosses in St. Gall.

Writing over 100 years ago, Strachan 1903a: 470 notes differing opinions as to the date of the glosses, and suggests that they have a heterogeneous origin. That is, they were collected and copied not from a single source, but from multiple sources. As such, it would make sense that they not show a unitary character (i.e., that they show variation in orthography, linguistic age, etc.). To support this idea of multiple sources being visible in the linguistic variation in the St. Gall Priscian, he notes the presence of multiple manuscripts glossing Priscian and the fact that there is variation in these manuscripts (1903a: 471–472). Further, it appears that St. Gall all but united these different collections. Thus, it makes perfect sense for there to be variation in the data. In the rest of his 1903a paper, Strachan then attempts to tease apart the various strata of glosses by examining different features and looking for parts of the manuscript where certain older or younger features can be grouped together. Though he examines a large number of variant features, the most significant are the endings of *i*- and *u*-stem genitive singulars (-o vs. -a at 472–473) and the orthography of schwa between a non-palatalized and palatalized consonant (CiCʲ vs CaiCʲ at 474); other interesting features include ó vs. úa (475–477), é vs. ía (477), -ae and -ai vs. -e and -i (477–478), and the writing of [b d g] as <bb dd gg> (479 & 485–486).

Strachan 1903a: 472–473 presents, mostly on the basis of the variation of -o and -a in the genitive singular, broad sections of text with more archaic and less archaic features. For instance, he notes that the variants are roughly equal on pp. 1–40; that the -a variant is much more common than -o on pp. 50–150; that the reverse is true on pp. 150–202; that -a is more common on pp. 202–210; and that only -o is found on pp. 215–216. The results are not as clear for the orthography of schwa, however. For most of the other features, he notes that there is a clear difference ‘between the early and

² The standard print edition of these gloss collections is found in Stokes & Strachan 1901–1903, while updated digital editions are found in Griffith & Stifter 2013 for Milan, Bauer 2015 for St. Gall, and Stifter et al. 2021 for both. A digital edition of Würzburg is still in progress, but Doyle 2018 provides a starting point.

the later part of the codex (with the exception of some of the final portion)' (1903a: 474), but he does not specify whether this difference is always in the same direction (e.g., all the archaic features are in the early part) or not, nor does he specify whether the division between 'early' and 'late' is consistently located across features, apart from saying that late features are rare between approximately pp. 150 and 200 (so also Hofman 1996: 45–46). In the penultimate paragraph of the main text of the article (i.e., before the lists of features and their loci), Strachan notes that '[f]rom purely linguistic evidence it will hardly be possible to go beyond this rough division' (1903a: 475). This might seem to be a discouraging result, and in some senses it is, but Strachan rightly and importantly notes that the issue probably lies in the nature of glosses as texts. Though one might talk of 'the St. Gall glosses' as a single text, they are in reality a collection of micro-texts and, thus, differ markedly from continuous narrative texts.

After Strachan's publication, the matter rested for over 90 years, until Lambert 1996 re-opens the dossier. While Strachan, indeed, recognizes the limits of his approach, Lambert confirms these limits and adds a few further considerations to the problem of determining the linguistic layers of the St. Gall glosses. He notes that the divisions of 'old' and 'new' that Strachan determines were largely arbitrary and that glosses do not have sections as such. Finally, the 'old' and 'new' sections were not uniform anyway, as they generally contain both older and newer forms. Beside these justified criticisms, Lambert also makes a number of contributions to the discussion by adding some additional dating criteria to the list that Strachan had made and by suggesting that hand B of the St. Gall glosses has more archaic features than hands A and C. Further, he makes the important observation that not only linguistic evidence, but also orthographic evidence should be taken together (1996: 189).

The most recent contribution to the discussion of linguistic layers in St. Gall is Roost 2013, who takes a slightly different tack. She examines parallel glosses with reference to three phonological features: final *-a* for *-(a)e*, the variation *-o* vs. *-a* in the genitive singular of the *i*- and *u*-stems, and *ó* vs. *úa*. Her findings are that Strachan's groupings are not confirmed. Rather, she finds what she terms 'small clusters' of younger or older forms on the same or adjacent pages (2013: 31 & 34). This finding is, in some senses, not surprising, given Lambert's 1996 observation, and also Strachan's suggestion 1903a: 475, that the glosses cannot really be seen as blocks of continuous text. Therefore, smaller clusters would indeed be expected. The limitation of Roost's study, of course, is that it is confined to the smaller number of glosses that are found in parallel manuscripts. It is, thus, unable to tell us anything about other glosses that are found only in the St. Gall Priscian.

It is now time to take stock of where we are with the problem. We know that the St. Gall glosses are not homogenous. Rather, they consist of an unknown number of different layers or groups. In addition, given that each individual gloss is, in a sense, its own micro-text without a necessary connection to the preceding or following micro-text, these different layers or groups do not necessarily consist of consecutive glosses. That is, the mere fact that one specific gloss has a series of features does not mean that an

adjacent one has a greater than average chance of also having those same features. Rather, the different layers or groups consist of a possibly random-seeming distribution of glosses across the entire manuscript. What is necessary is, as Lambert 1996: 192 suggests, a profile for each individual layer of glosses. If we can identify, for some subset of glosses, a profile of the layer they belong to, we can then look for other unknown members of that same layer by looking for the same profile. What such a profile might look like is the problem to which we will now turn.

2. Indicators for linguistic layers in the St. Gall Priscian glosses

The idea behind a profile for various layers of glosses in St. Gall is itself relatively simple. We might imagine, for example, that three distinct groups of glosses were brought together in St. Gall. Each of these could (but need not) have distinct characteristics. If we could determine those characteristics, we might be able assign individual glosses to the different groups, regardless of where those glosses appeared in the manuscript. This theoretical outcome is obviously an optimistic one. Many, even most, glosses are short and would probably not contain enough information to assign them to a group. Nonetheless, the idea is worth exploring further, given that it could, indeed, provide us with valuable information about gloss groupings and, therefore, gloss layers.

In order to make this idea work, we must establish what variation is present in the St. Gall glosses as a whole. Before turning to that, however, it is necessary to discuss some facts and assumptions underlying what is to come. First of all, different layers of glossing do not require different chronological layers. For example, two scribes working at the same time might show diatopic variation (i.e., variation associated with a particular place) or might exhibit different traits because of training in different schools (i.e., different orthographic or standardization practices). Different layers might also not show any differences at all apart from their handwriting. On the other hand, chronological layers can and do correlate with certain linguistic and orthographic traits, for which one might compare the three different hands in the Würzburg glosses. We can, therefore, assign many variables in the glosses a relative chronological status.

For purposes of this project, I propose the following divisions: ‘early’ (roughly contemporaneous to the *prima manus* of Würzburg, ca. 700); ‘Würzburg’ (covering the Würzburg II hand ca. 750, as well as the somewhat later Würzburg III hand; hereafter abbreviated ‘Wb.’); ‘Milan’ (covering the Milan glosses, ca. 800; hereafter abbreviated ‘Ml.’); and ‘late’ (from the time of the copying of St. Gall ca. 850 to the end of the Old Irish period). In earlier versions of this project, I had used a ternary division, with ‘Wb.’ and ‘Ml.’ being collapsed into one ‘middle’ period, but I have come to accept the views of a reviewer who criticized the fewer divisions as insufficient.³ While it will become

³ I do not, however, make allowance here for the third hand of Würzburg. It comprises relatively few glosses (under 100), meaning that there is very little data from which conclusions can be drawn. For that reason, I collapse Würzburg II and III below in any statistics given. Again, this does not mean that I

clear below that splitting the two does not make a difference in the majority of categories considered here, having separate labels for 'Würzburg' and 'Milan' does retain useful information in some cases, and that is clearly desirable.

A further difficult point concerns the dating of features more generally. McCone 1985 outlined numerous Middle Irish features found already in the Würzburg and Milan glosses. This shows that a number of sound changes often noted as Middle Irish must have actually occurred already in the Old Irish period, at least in some registers or dialects. To exemplify the general problem, we may consider a hypercorrect form like *húa ambus* 'by attempt' (Ml. 75^d8), with an unetymological spelling -*mb*- (for *ammus* vn. of *ad-midethar*).⁴ This hypercorrect form shows that the assimilation of [nd] and [mb] to [nn] and [mm] had already occurred by the time of the copying of Milan. In fact, it had probably already happened by the time of the copying of Würzburg (cf. acc. sg. *cláinn* for expected *claind* 'descendants' [Wb. 5^b33]).⁵ Nonetheless, while hypercorrection and assimilated <nn> shows that the change had already taken place by Classical Old Irish, it is nonetheless quite uncommon in any of the major gloss collections. For that reason, I would argue that the *writing* of older [nd] and [mb] as <nn> and <mm> is a 'late' feature of Old Irish, while *writing* them as <nd> and <mb> is a feature of the 'early', 'Wb.', and 'Ml.' periods, and this despite the fact that the sound change itself must have taken place already by the 'Wb.' period.

This line of argumentation may seem counter-intuitive, but it attempts to classify forms and variation according to the normal spelling of the different periods. The precise date of the linguistic change underlying the spelling is important, but it is not the primary focus here. The primary focus is textual layers in St. Gall, for which linguistic and orthographic variation within the glosses is critical. For that reason, the approach differs from that found in Stifter 2013, although the material investigated is similar and, thus, a large number of categories discussed here are also discussed by him. Fortunately, the conclusions for the items are all compatible, unless stated below. In general, however, it will be useful to consult Stifter 2013: 173–198 in conjunction with the following list.

At this point, we can turn to a listing, with brief commentary, of the main categories relevant to the dating of the St. Gall Priscian glosses, as well as a brief explanation of how they are interpreted for the purposes of this paper. Note that some of the categories are linguistic in nature, while others (from 12 to 15) are orthographic in nature. Both are relevant for classifying gloss layers or groups:

deny any differences between them. It does mean that I do not think that there are enough data points to make a separation meaningful.

⁴ Whether -*ðm*- should have remained or not (cf. *maidm* 'act of breaking' [Thurneysen 1946: 94 §152 (c)]) is irrelevant here.

⁵ This form does not fall under the rule formulated by Fortson 2009–2010, which states that palatalized ND yielded N between front vowels at a relatively early period. Since *cláinn* (Wb. 5^b33) falls outside that rule (ND does not follow a front vowel), it appears secure that the later assimilation of ND to NN had already occurred in Würzburg, unless this writing is simply an error. Stifter 2013: 183 states that the assimilation probably did not take place until the ninth century, thus placing it after Wb., but he does not discuss the Wb. form mentioned just above.

1. The gen. sg. of *i*- and *u*-stems varies between *-o* and *-a*. McCone 1996: 139 argues that *-o* began to merge with *-a* ‘quite early’. While Würzburg already has both, the *-o* variant is still strongly dominant (115 *-o* vs 65 *-a*; see Strachan 1903a: 473¹).⁶ By Milan, the majority has switched to the *-a* variant (195 *-a* vs. 120 *-o*), but it can hardly be called the ‘normal form’ there (McCone 1996: 139). McCone was likely influenced by his assumption that the archaic nominative singular form *fēda* ‘Lord’ found in Cambrai shows the confusion of *-a* and *-o*, but this form is likely to be phonologically correct and should not confuse the issue at hand (see Sims-Williams 1999; Griffith 2005). The upshot of this discussion is that the *-o* variant is ‘early’ and ‘Wb.’, while the *-a* variant is rather a ‘Ml.’ and ‘late’ trait.⁷
2. The decline of the deponent endings during the course of the Old Irish period has been well documented (Strachan 1893; Griffith 2014). Though active endings are found already in Würzburg, they are rather uncommon, and it has been suggested that that gloss collection was written / copied at a time not far removed from when active endings first began to penetrate into deponent paradigms (Griffith 2014: 69). Even in Milan, the deponent endings are highly dominant (Griffith 2014). It, therefore, makes sense to class the deponent endings as ‘early’, ‘Wb.’, and ‘Ml.’, and non-deponent endings as ‘late’, as it is not until after Milan that the loss of the deponent endings truly gains steam.
3. Though the details are not entirely clear (McCone 1996: 134), *ó* alternates with *úa*. The former is preserved generally in Early Old Irish (Thurneysen 1946: 39–41 §60, where he refers to ‘archaic texts’). The *ó* also survives before all gutturals in Würzburg, while in Milan and St. Gall, it survives before voiced gutturals.⁸ We must, therefore, distinguish several environments. Undiphthongized *ó* before non-gutturals is clearly ‘early’, while *úa* in that

⁶ I have taken Strachan’s collection as definitive for the figures cited here of *-o* vs *-a*. A reviewer points out that *inbetho* (Wb. 10^b1) is not listed in that collection. There may be other omissions, but I assume here that they do not materially affect the overall picture of the distribution of the two endings in Würzburg.

⁷ The *-a* variant actually masks two different endings. The alternation of *-o* and *-a* in Würzburg and Milan, for instance, presumably reflects the difference between phonetic [o] and [a]. By late Old Irish, however, the *-a* variant had very likely acquired the pronunciation [ə] via the general weakening of final syllables. I do not see a principled way of recovering which pronunciation was intended in any gloss, since the spelling of both is the same: <a>. As a result, the spelling variant <a> of the genitive singular is assigned to both ‘Ml.’ (probably in its [a] pronunciation) and ‘late’ periods (in its [ə] pronunciation). No more precision than that is possible.

⁸ The evidence is somewhat limited, but it accords with the description in the main text. Undiphthongized *ócht* ‘cold’ (3 examples in Wb.) vs. diphthongized *úacht* (5 examples in Ml.) shows the split before voiceless gutturals (Ml. also has *úachtar* / *úachtarach* ‘upper’ with *úa*; 4 examples). Forms and derivatives of *óg* ‘whole, complete’ always have the monophthong (13 examples in Wb.; 3 examples in Ml.; 15 examples in Sg.). For *trúag* / *trúagdae* ‘wretched’ and *trúagae* ‘wretchedness’, Milan generally has *ó* (15 examples), though *úa* also appears (4 examples). Würzburg has only *ó* (2 examples), while St. Gall has one example each of *ó* and *úa*. Forms of *tróaire*, despite the voiceless guttural (see Thurneysen 1946: 87 §137.2), always have *ó*, even in Ml.: (5 examples in Wb. and 17 examples in Ml.). As this is a compound of

- position is 'Wb.', 'Ml.', and 'late'. Before voiced gutturals, *ó* is 'early', 'Wb.', and 'Ml.', while *úa* is 'late'. Finally, before voiceless gutturals, *ó* is 'early' and 'Wb.', while *úa* is 'Ml.' and 'late'. See also Roost 2013: 14 for brief discussion.
4. Similar to *ó*, but with clearer details, is the case of *é* and *ía* before non-palatalized consonants. The Classical Old Irish situation is that *é* is found before palatalized consonants and *ía* before non-palatalized ones (McCone 1996: 134). In Early Old Irish, however, *é* was found everywhere. We can, thus, classify instances of *é* before non-palatalized consonants as 'early' and cases of *ía* as 'Wb.', 'Ml.', and 'late'. Before palatalized consonants, there is no change and, thus, nothing to note.
 5. A similar distribution applies to the voicing of voiceless [t] and [θ] when on the word boundary and not next to an accented vowel (McCone 1981 & 1996: 132–133). The voiceless variants are 'early' while the voiced variants are 'Wb.', 'Ml.', and 'late'. Scribes were aware of this change and hypercorrect spellings are found already in Würzburg II (*búaith* for *búaid* 'victory' [Wb. 11^a7]),⁹ but, in line with the argument offered above, we will classify according to the orthography.
 6. This voicing rule also applies to dentals between unaccented vowels, though McCone 1996: 133 notes that its occurrence is later. Given the only rough chronological differences applied here in this paper, the voiceless variants are still 'early' while the voiced ones are 'Wb.', 'Ml.', and 'late', but the distribution is different, and the features should, thus, be separated here.
 7. Though McCone 1996: 133 takes the alternation of final [x] and [χ] as collapsible chronologically with that of intervocalic [θ] and [t], it is better kept separate here. Though the similarity to the above is notable, the environments are not precisely the same.
 8. Though McCone 1996: 141 treats the assimilation of [ln] arising through syncope to [ll] as a Middle Irish change, it is fairly clearly present in Milan (41 examples of -ll- in *comalnaithir* and related forms vs. 2 with -ln-), while it is largely absent in Würzburg (68 examples retaining -ln- in *comalnaithir* and related words vs. 0 with -ll-). We can, therefore, take <ln> as an 'early' or 'Wb.' feature and <ll> as a 'Ml.' or 'late' feature.
 9. McCone 1996: 141 also treats the assimilation of [ld] arising from syncope to [ll] as Middle Irish. It is clearly a later feature (not appearing in Milan),¹⁰

tróg / *trúag*, the *ó* of *tróaire* could have been retained analogically to the simplex. This is uncertain, but it is the working hypothesis here, which means that Milan regularly shows *úa* before voiceless gutturals.

⁹ My thanks to a reviewer for pointing this out to me, and also for his or her note on two cases of the *prima manus* with *báeth* 'foolish' spelled *baid* (Wb. 8^d8 & 12^d35), seeming to show an overapplication of the voicing rule.

¹⁰ Note, however, *lase nad-reildisem-ni* 'when we did not pollute' (Ml. 63^d15), with the hypercorrect spelling <ld> for expected <ll> = [ll] < [ln]. The verb is *as·lena* 'pollutes'. This indicates that the change -ld- > -ll- must have occurred already to some extent by 800. The principle for dealing with hypercorrection was discussed above and is followed here: the orthographic norm is treated as the standard.

and we can thus treat [ld] as an ‘early’, ‘Wb.’, and ‘Ml.’ feature and [ll] as a ‘late’ one.

10. As noted already by Thurneysen 1946: 315 §495.2 (b), special relative forms, e.g., *canas* ‘who sings’, begin to show lenited initials in the later Old Irish period, e.g., *chanas* ‘who sings’. This allows us to classify mutated special relatives as ‘late’, while non-mutated ones are ‘early’, ‘Wb.’, and ‘Ml.’.
11. The change of [nd] and [mb] to [nn] and [mm] is found already in Würzburg, though it is quite rare both there and in the Milan glosses (see Uhlich 2021). Therefore, we can classify the writing as <nd mb> as ‘early’, ‘Wb.’, and ‘Ml.’, while writing them as <nn mm> is ‘late’ (see also the discussion above).
12. The writing of voiced stops [b d g] with geminates <bb dd gg> is not a universal late trait, as it is found occasionally in Würzburg and Milan, but it appears more often in later texts. Therefore, the geminate writings are classified as ‘late’.¹¹ Non-geminate writings, both post-vocalic <p t c> and the post-consonantal <b d g>, are ‘early’, ‘Wb.’, and ‘Ml.’.
13. The writing of nasals with a punctum delens as <ṁ ṇ> in nasalizing contexts is said to be ‘frequent’ (Thurneysen 1946: 24 §33.1). It also appears between consonants within a word. There is no indication of frequency for the feature, in general, but a quick search of the three main gloss databases yielded the following figures: Würzburg has 25 examples of <ṁ> and 75 examples of <ṇ>; Milan has 34 examples of <ṁ> and 166 examples of <ṇ>; St. Gall has 35 examples of <ṁ> and 223 examples of <ṇ>. Without a complete collection, the relative frequencies cannot be entirely certain, but, given that St. Gall is the smallest of the three major gloss collections yet contains the greatest absolute numbers of <ṁ ṇ>, it seems all but certain that it has the highest ratio of appearance. The feature thus appears to be ‘late’, while its absence is ‘early’, ‘Wb.’, and ‘Ml.’. Some additional research is necessary here to be certain that this is the correct division.
14. The use of a punctum delens to denote lenited [s] and [f] is also a late feature, as the writings <ṣ> and <ṑ> do not generally appear in Würzburg or Milan (Thurneysen 1946: 24 §33.3).¹² The same goes for the indication of lenition on [p] (written as <ph>), which is infrequently indicated in the earlier gloss collections. Thus, if any of these features appears, it is classified as ‘late’. The absence of such features is ‘early’, ‘Wb.’, and ‘Ml.’.
15. The final two categories involve vowels. The first of these concerns ‘helping’ vowels used to indicate the quality of surrounding consonants. As noted in Stifter 2013: 188–189, the *a*-glide is, in general, useful as a chronological

¹¹ Note that Lambert 1996: 188–189 suggests that this trait is archaic, rather than late. I am not sure why he makes this suggestion, but, given that geminate spellings are more frequent in the later gloss collections, I believe that the opposite is true.

¹² There is one questionable case at Wb. 33^c9 (Stern 1908: 545; contra *Thes. Pal.* i 710). The punctum could be used early (see Anderson & Anderson 1961: 128 for an example from codex A of Adomnán’s *Life of Columba*), but the fact is that it is quite rare. I thank reviewer 2 for pointing this out to me.

guide. Its absence is 'early' and 'Wb.' (e.g., in 3. sg. pres. subj. *berid* 'he may carry'), while its presence is 'Ml.' and 'late' (e.g., in *beraid* 'he may carry'). For details and examples, see Thurneysen 1946: 61–67 §§97–105. The *i*-glide, while less common in older sources, is still quite regular, and, as a result, cannot be used to indicate older or younger texts. It will not be considered further here.

16. The next vocalic category concerns the collapse of distinct word-final vowels in Late Old Irish. Though some examples are found already in Würzburg (albeit rarely) and Milan (more frequently; see Strachan 1903b: 51–52), indicating that the sound change had occurred already by Classical Old Irish, it is not a common occurrence in either gloss collection. The change of word-final vowels to schwa (variously spelled) is, thus, categorized as a 'late' change, while the retention of the original vowel quality is 'early', 'Wb.', and 'Ml.' (see also Thurneysen 1946: 62 §99 [c]). Note that examples like *imde* 'many' without a helping vowel and *imdae* are classified as 'early' / 'Wb.' and 'Ml.' / 'late', respectively, under the previous category, which deals with spelling. Examples like *imda* involve a sound change and are counted in this category as 'late'.

This list of features is not quite complete. There are two other features relevant to gloss layers in the St. Gall Priscian glosses: parallel glosses and hands. In some ways, these are the most obvious features relevant to layers. For instance, a different hand guarantees that there is an additional layer to the text, and there are three different Irish-language glossing hands in St. Gall: A, B, and C (Stokes & Strachan 1901–1903: ii xix–xx; Hofman 1996: i 17–19).¹³ Hand A provides the bulk of the Irish material, with hand B responsible for a couple of folios and hand C a couple of glosses. Similarly, the status of parallel glosses for Priscian has been investigated, and it has been seen that glosses in other manuscripts on Priscian are frequently found in St. Gall (Strachan 1903a: 471–472; Hofman 1996: i 42–43). This suggests that these glosses formed parts of collections that were copied into St. Gall. These different collections can thus also be seen as different layers. Despite the fact that parallel glosses and different hands are strong indicators of different layers in the St. Gall glosses, I think it advisable to set these two features aside for now. We already know about these (possible) layers, so we do not need help discovering them. It is other possible layers that are of interest. Also, the information about hands and parallel glosses is a given and can, thus, be used to help test or check results from the current project. That is, if the grouping algorithm (see below) suggests groupings that correspond in some way to the different hands or to parallel glosses, that can be taken as confirmation that the algorithm is working.

¹³ Ó Néill 1998: 2 & 26²³, citing information from Gearóid Mac Eoin, notes without further specification that there are dry point glosses in the St. Gall manuscript, as well, which would make a fourth hand. I have no other information about these glosses, such as their content or even the language that they are written in, so I leave them aside here.

3. Establishing a profile for linguistic layers

Now that we have established what features and parameters will be examined, it is time to explore what to do with them. The most straightforward thing is to analyze each gloss and assign it a ‘score’ for each of the 16 variables and each of the four periodizations, ‘early’, ‘Wb.’, ‘Ml.’, and ‘late’. To give a short example, we can examine the following gloss:

i. *tre chomaisideis do inchosc óen-cheillæ*¹⁴ Sg. 27^{b3}
 ‘i.e., by apposition, to signify a single conception’

The only relevant feature is the writing of [n] between consonants as <ñ>, a ‘late’ feature. The *i*-glide in *cheillæ* (= *-chéille*) is argued above to be non-indicative and, thus, is not considered further. This profile can most easily be represented in a matrix (see Table 1 below). A slightly more complicated example is the following:

níbbu machdath betis grecdi ⁊ nothath foraiþ linni ⁊ dano it latindi amal Sg. 6^{a9}
sodain ut dixit prius. in latinis verbis placuit .f usque facio

‘it were no wonder that they were Greek since we call them loan-words; and yet they are Latin in that case, as he said before: ‘In Latin words it seemed proper (. . . to write) ‘f’ until ‘facio’ (do)’¹⁵

Here, there are more numerous features. The writing of [b d g] has one ‘later’ feature (*níbbu*).¹⁶ The lenited [s] is written with a punctum delens, which is a ‘late’ feature. The writing of helping vowels shows two ‘early’ and ‘Wb.’ features (one in *grecdi* = *grécdai* and one in *latindi* = *laitindai*) and two ‘Ml.’ and ‘late’ features (*foraiþ* and *sodain*). Finally, it shows two instances of voiceless [θ] beside an unaccented vowel (*machdath* and *nothath*), which is ‘early’. The full matrix for this gloss is illustrated below (see Table 2).

¹⁴ The text and translation follow Bauer 2015, who follows Thes. ii 77, except that the final word is *óencheillæ* rather than *óencheillae*. This is clear from the manuscript, but it also makes sense in that -æ should not normally follow a palatalized consonant. I am assuming here that -æ is an alternate spelling of -e, which follows from their identity in post-Classical pronunciation (see Thurneysen 1946: 18 §24.1 for the principle, though citing exactly this form as a rare example of spelling <ae> for [e] is not correct).

¹⁵ The translation is that of Hofman 1996: ii 29 under 6^{a18}.

¹⁶ Though the etymology and, thus, the pronunciation of *machdath* is unsure (cf. *machthad* [Wb. 18^{c6}], *macdath* [Wb. 17^{c9}] *prima manus*), the <d> probably represents [ð] here. My thanks to a reviewer for notes on this form.

Table 1: Features of Sg. 27^{b3}

	EARLY	WB.	ML.	LATE
1. -o / -a	0	0	0	0
2. deponents	0	0	0	0
3. ó / úa	0	0	0	0
4. é / ía	0	0	0	0
5. [-t -θ] / [-d -ð]	0	0	0	0
6. [-θ-] / [-ð-]	0	0	0	0
7. [x] / [ɣ]	0	0	0	0
8. [ln] / [ll]	0	0	0	0
9. [ld] / [ll]	0	0	0	0
10. canas / chanas	0	0	0	0
11. <nd> / <nn>	0	0	0	0
12. <p b> / <bb>	0	0	0	0
13. <ñ> / <n>	0	0	0	1
14. <s f> / <ś f̃>	0	0	0	0
15. a-glides	0	0	0	0
16. -V > -ə	0	0	0	0

Table 2: Features of Sg. 6^{a9}

	EARLY	WB.	ML.	LATE
1. -o / -a	0	0	0	0
2. deponents	0	0	0	0
3. ó / úa	0	0	0	0
4. é / ía	0	0	0	0
5. [-t -θ] / [-d -ð]	2	0	0	0
6. [-θ-] / [-ð-]	0	0	0	0
7. [x] / [ɣ]	0	0	0	0
8. [ln] / [ll]	0	0	0	0
9. [ld] / [ll]	0	0	0	0
10. canas / chanas	0	0	0	0
11. <nd> / <nn>	0	0	0	0
12. <p b> / <bb>	0	0	0	1
13. <ñ> / <n>	0	0	0	0
14. <s f> / <ś f̃>	0	0	0	1
15. a-glides	2	2	2	2
16. -V > -ə	0	0	0	0

That the majority of cells in both matrices is empty is expected, as most features will not be present in most glosses. It is also not a problem that there appear to be contradictory ‘early’ and ‘late’ features in gloss 6^{a9}. This point, however, deserves further discussion.

All of the features in the list given above in §2 were assigned to ‘early’, ‘Wb.’, ‘ML’, or ‘late’ based on the relative chronology of the relevant sound changes and palaeographic developments. Given the conservative nature of the copying of texts, it is inevitable that early features will appear beside later ones, especially since the entire manuscript was copied in the later Old Irish period. This issue is a familiar one to scholars of Early Irish and does not need further elaboration here. What does need further explanation is how this may be useful to us.

What we are looking for is different layers in the glosses. Layers in this context could result from differences arising out of the habits of different copyists at the most surface level, i.e., hand A vs. hand B. They could result from a copyist’s faithful copying of different source texts into the St. Gall Priscian, i.e., they could indicate differences in the source texts. The orthographic or paleographic practices of different schools or scribes, as well as chronological differences stemming from different times of copying, could also arise. There could also be myriad different combinations of all the above. These different possibilities can explain how early features, which were unmodified during the copying process from an earlier exemplar, could be present beside late features, which were modified during the copying process. Whether we will be able to detect different layers probably depends on whether the data is robust enough and whether the differences between the layers are great enough to detect. As noted above,

layers can, indeed, be present without any means of detecting them if there are no outward differences between them. Therefore, we must carefully consider how to analyze the collected and tabulated data.

4. Comparing profiles and finding similarities

Now that we have argued for a profile as a way to represent critical features of the different glosses, we can try to compare the profiles in a meaningful way. The goal is to identify groups of glosses that ‘belong together’. Given the profiles above, this can be equated to finding profiles that can be grouped together because they are similar and then examining the network that they produce. The best example of this kind of work in Celtic studies is Ó Muircheartaigh (2015: esp. 48–112), which examines the similarities between modern Irish and Scottish Gaelic dialects using mathematical methods. He presents a synchronic analysis of the relationships of the modern dialects based on a measure of how similar they are. The dialect divisions that he presents correspond in large part, though not entirely, to what scholars have posited based on philological work. Ó Muircheartaigh’s work, though based on a synchronic analysis of the modern dialects, also has a historical interpretation. The parallels to what is being proposed here are strong: by finding similarities on a synchronic level, i.e., the Old Irish glosses on the St. Gall Priscian, we can arguably recover historical information, i.e., what glosses come from a common source or belong to a similar / the same chronological period. In order to achieve this, we must be able to assess the difference between every gloss pair as represented in Tables 1 and 2. There are various ways of measuring difference, but probably the most straightforward manner is simply to take a distance measure that is the difference between values in every cell of the table.¹⁷ For the two glosses given above, the difference matrix is Table 3:¹⁸

¹⁷ More technically, the similarity matrix *S* of two matrices *A* and *B* is the square root of the square of the difference for each cell:

$$S(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (a_{ij} - b_{ij})^2}$$

As noted above, other definitions are also possible, but I do not see a particular advantage to using anything more complicated than this. This definition simply ensures that there are no negative values in the difference matrix.

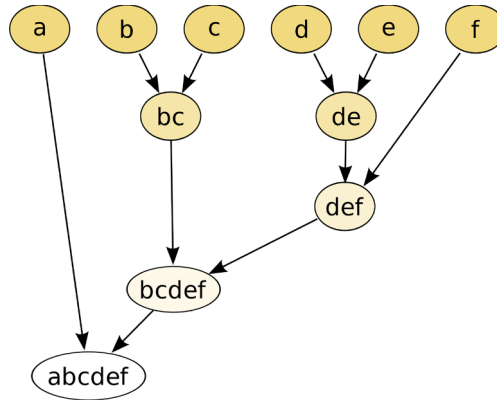
¹⁸ This data is not normalized to a scale between 0 and 1, which means that it will tend to give more precedence to longer glosses at the expense of shorter ones. It is unclear what effect this may have on the analysis. I believe that it may make the cohesion of groups that are discovered (see below) stronger, but perhaps at the cost of missing subtle differences among the shorter glosses.

Table 3: Difference of Sg. 27^b3 and 6^a9

	EARLY	WB.	ML.	LATE
1. -o / -a	0	0	0	0
2. deponents	0	0	0	0
3. ó / úa	0	0	0	0
4. é / ía	0	0	0	0
5. [-t -θ] / [-d -ð]	2	0	0	0
6. [-θ-] / [-ð-]	0	0	0	0
7. [x] / [ɣ]	0	0	0	0
8. [ln] / [ll]	0	0	0	0
9. [ld] / [ll]	0	0	0	0
10. <i>canas</i> / <i>chanas</i>	0	0	0	0
11. <nd> / <nn>	0	0	0	0
12. <p b> / <bb>	0	0	0	1
13. <ṅ> / <n>	0	0	0	1
14. <s f> / <ś f̃>	0	0	0	1
15. <i>a</i> -glides	2	2	1	1
16. -V > -ə	0	0	0	0

On its own, this table cannot tell us anything meaningful. Only in relation to all of the other tables is it possible to glean useful information. Therefore, this calculation needs to be done for every pair of glosses in the collection. As there are approximated 3,500 Old Irish glosses in the St. Gall Priscian, this will mean roughly six million difference tables. Though the matrix of features for each gloss (as illustrated in Tables 1 and 2 above) needs to be created manually, the calculation of the difference tables (as illustrated in Table 3) can fortunately be done automatically. Once the difference tables have been made, what remains is to see what groups can sensibly be made from them.

For this, hierarchical cluster analysis can be used. For an example of such work within Celtic studies, see Yocum 2020 on the authorship of texts within the Book of Leinster. A cluster analysis uses the difference between items (as defined in the previous paragraph) to determine groupings automatically based on how far away they are from one another. There are various freely available programs that do this clustering and, as of yet, no decision has been made on which to use. One of the interesting aspects of such cluster analysis is that the number of groups is not fixed, but can vary from one to the maximum number of items in the collection (in this case, approximately 3,500). The open nature of the number of groups is critical, because how many groups there are depends on the time-depth being examined. A simple example with possible interpretation can help illustrate this.

FIGURE 1: Hierarchical cluster example¹⁹

In this figure, we can imagine that a cluster analysis has been run on the glosses. If there is one cluster (found at the bottom of the figure), it represents the entire collection of Old Irish glosses in the St. Gall Priscian. This is not particularly interesting, but it is certainly a valid group at the shallowest time-depth.

If we look at a slightly higher / deeper level, there are two groups: {a} and {b, c, d, e, f}. This could represent, for example, one group, {a}, with all the glosses that contained no useful criteria according to the 16 categories given above in §2 vs. all other glosses. This is a particularly likely outcome, i.e., that there will be one very large group in which the only commonality among its members is the fact that none of them contains any data useful for assisting a grouping of them together with anything else. At a still higher / deeper level, there are three groupings: {a}, {b, c}, and {d, e, f}. Group {b c} could represent, for instance, glosses copied by hand B, while {d, e, f} represents glosses in hand A.²⁰ Higher levels would contain more groups, until eventually there would be approximately 3,500 groups, each consisting of an individual gloss. The example given here is entirely hypothetical, but it represents a possible outcome of the type of analysis being proposed here. The precise interpretation of the groupings would entail looking carefully at the members and considering what led the items to be grouped together.

5. Outlook for future work

As can hopefully be gleaned from the above description, this is a large research project and the completion of it will require significant work inputting the data into tables

¹⁹ The figure, licensed under the Creative Commons License <https://creativecommons.org/licenses/by-sa/3.0/deed.en>, was made by Stathis Sideris and has been reproduced here unchanged.

²⁰ In both cases, the glosses that contained no useful criteria would still be in {a}.

and then analyzing it.²¹ What is unclear is how much the work will deliver in terms of results. For that reason, the next step must be to analyze a sub-set of the data to see whether the methods and analysis as described above indeed function as intended. To that end, I would suggest looking at two sections from the St. Gall Priscian: from 65^b6 to 67^b21 and from 24^a1 to 26^b34. The former group represents most glosses written in hand B (plus a couple of extra hand A glosses interspersed among those from hand B), while the latter group is in hand A and, according to Lambert 1996: 188, shows a mixture of archaic and later features.

One could argue that handpicking the data to test the project biases the pilot project unfairly toward success. That is, including all of the glosses in hand B, as well as a section of text with both archaic and newer features maximizes the differences present in the data, thus increasing the probability of finding something interesting. Though it is true that the pilot project seems to have an advantage in terms of finding positive results, there is no guarantee that this is the case, and the idea is to test the theory and methods proposed here. The analyses proposed here will produce groupings of glosses, but in the end, any such proposals will have to be judged by scholars as to whether the results are interpretable in a way that is meaningful and plausible. If they are not, we can deem the groupings as unsuccessful or perhaps uninteresting. If the results are interpretable and scholars agree that new groupings have been discovered, we can deem the project successful.

We already know, in a very real sense, all of the relevant facts. The St. Gall glosses themselves are a given: their form and meaning are essentially certain. The 16 criteria given above are not new. We know them from comparative linguistic and palaeographic work. Their division into various periods is also not particularly controversial. What we do not have is a sense of whether these facts can be connected in ways that inform us about underlying groups of glosses, i.e., layers of glossing. If all goes well, this project may enable us to see some connections that right now are not visible.

ABBREVIATIONS

Ml.	=	Stokes & Strachan 1901–1903: i 7–483
Sg.	=	Stokes & Strachan 1901–1903: ii 49–224
Thes.	=	Stokes & Strachan 1901–1903
Wb.	=	Stokes & Strachan 1901–1903: i 499–712

²¹ Despite the use of statistics, which is not common in medieval Celtic studies, this project is fairly traditional in its approach, as it involves hand-coding of a significant database followed by analysis. It is, thus, quite different from the approach to text dating by machine found in Toner & Han 2019, as the project proposed here does not involve machine learning.

References

- Anderson, Alan Orr, & Marjorie Ogilvie Anderson (eds.). 1961. *Adomnan's Life of Columba*. London: Thomas Nelson and Sons.
- Association for Computing Machinery. 2020. *Artifact review and badging*. Available on-line at <https://www.acm.org/publications/policies/artifact-review-and-badging-current> (accessed 14 July 2021).
- Bauer, Bernhard. 2015. *The online database of the Old Irish Priscian glosses*. Institut für Sprachwissenschaft, Universität Wien. Available on-line at <http://www.univie.ac.at/indogermanistik/priscian/> (accessed 14 July 2021).
- Doyle, Adrian. 2018. *Würzburg Irish glosses*. Available on-line at <https://wuerzburg.ie> (accessed 14 July 2021).
- Fortson, Benjamin W., iv. 2009–2010. On 'double-nasal' presents in Celtic and Indo-European and a new Irish sound law. *Zeitschrift für celtische Philologie* 57: 48–78.
- Griffith, Aaron. 2005. *-n(C)s in Celtic. *Die Sprache* 44: 44–67.
- . 2014. The decline of the Old Irish deponent. In *Linguistic and philological studies in Early Irish*, ed. David Stifter & Elisa Roma, 53–78. Lampeter: Mellen Press.
- , & David Stifter. 2013. *A dictionary of the Old-Irish glosses in the Milan Codex Ambrosianus C 301 inf.* Institut für Sprachwissenschaft, Universität Wien. Available on-line at <http://www.univie.ac.at/indogermanistik/milan-glosses.htm> (accessed 14 July 2021).
- Hofman, Rijcklof. 1996. *The Sankt Gall Priscian commentary. Part 1*. Münster: Nodus.
- Lambert, Pierre-Yves. 1996. Les différents strates de gloses dans le ms. de Saint-Gall no. 904 (Priscien). In *Irland und Europa im früheren Mittelalter. Bildung und Literatur / Ireland and Europe in the early Middle Ages. Learning and literature*, ed. Próinséas Ní Chatháin & Michael Richter, 187–194. Stuttgart: Klett-Cotta.
- Lash, Elliott. 2017. A quantitative analysis of *e/i* variation in Old Irish *etar* and *ceta*. *Ériu* 67: 141–167.
- McCone, Kim. 1981. Final /t/ to /d/ after unstressed vowels, and an Old Irish sound law. *Ériu* 32: 29–44.
- . 1985. The Würzburg and Milan glosses. Our earliest sources of 'Middle Irish'. *Ériu* 36: 85–106.
- . 1996. *Towards a relative chronology of ancient and medieval Celtic sound change*. Maynooth: Department of Old Irish, Saint Patrick's College.
- National Academies of Sciences, Engineering, and Medicine. 2019. *Reproducibility and replicability in science*. Washington, DC: The National Academies Press.
- Ó Muircheartaigh, Peadar. 2015. *Gaelic dialects past and present. A study of modern and medieval dialect relationships in the Gaelic languages*. Ph.D. dissertation, University of Edinburgh.
- Ó Néill, Pádraig P. 1998. The earliest dry-point glosses in Codex Usserianus Primus. In *A miracle of learning. Studies in manuscripts and Irish learning. Essays in honour of Williams O'Sullivan*, ed. Dáibhí Ó Cróinín & Toby Barnard, 1–28. Aldershot: Ashgate.
- . 2000. Irish observance of the three lents and the date of the St. Gall Priscian (MS 904). *Ériu* 51: 159–180.
- Plesser, Hans E. 2018. Reproducibility vs. replicability. A brief history of a confused terminology. *Frontiers in neuroinformatics* 11: article 76.
- Roost, Juliana. 2013. *Stages of language development in the Old Irish glosses on Priscian*. M.A. thesis, University College Dublin.
- Sims-Williams, Patrick. 1999. Old Irish *feda* (gen. *fedot*). A 'puzzling' form in the Cambrai Homily and its implications for the apocope of /i/. In *Studia Celtica et Indogermanica. Festschrift für Wolfgang Meid zum 70. Geburtstag*, ed. Peter Anreiter & Erzsébet Jerem, 471–474. Budapest: Archaeolingua.
- Stern, Ludwig Christian. 1908. Bemerkungen zu dem Würzburger Glossencodex. *Zeitschrift für celtische Philologie* 6: 531–545.

- Stifter, David. 2013. Towards the linguistic dating of early Irish law texts. In *Medieval Irish law. Text and context*, ed. Anders Ahlqvist & Pamela O'Neill, 163–208. Sydney: Celtic Studies Foundation, The University of Sydney.
- , Bernhard Bauer, Elliott Lash, Fangzhe Qiu, Nora White, Siobhán Barrett, Aaron Griffith, Romanas Bulatovas, Francesco Felici, Ellen Ganly, Truc Ha Nguyen, & Lars Nooij. 2021. *Corpus PalaeoHibernicum (CorPH) v1.0*. Available on-line at <http://chronhib.maynoothuniversity.ie> (accessed 14 July 2021).
- Strachan, John. 1893. Contributions to the history of the deponent verb in Irish. *Transactions of the Philological Society* 22: 444–566.
- . 1903a. On the language of the St. Gall glosses. *Zeitschrift für celtische Philologie* 4: 470–492.
- . 1903b. On the language of the Milan glosses. *Zeitschrift für celtische Philologie* 4: 48–71.
- Stokes, Whitley, & John Strachan. 1901–1903. *Thesaurus Palaeohibernicus*. Cambridge: Cambridge University Press.
- Thurneysen, Rudolf. 1946. *A grammar of Old Irish*, rev. ed. D. A. Binchy & Osborn Bergin. Dublin: Dublin Institute for Advanced Studies.
- Toner, Gregory & Xiwu Han. 2019. *Language and chronology: text dating by machine learning*. Leiden: Brill.
- Uhlich, Jürgen. 2021. When did *nd* become *nn* in which Early Irish environments? Handout for *Comhhdháil lae in onóir do Damian McManus*, Coláiste na Tríonóide, 21 May 2021.
- Yocum, Christopher Guy. 2020. Text clustering and methods in the Book of Leinster. In *Morphosyntactic variation in medieval Celtic languages. Corpus-based approaches*, ed. Elliott Lash, Fangzhe Qiu, & David Stifter, 85–112. Berlin: De Gruyter Mouton.