

Faclair na Gàidhlig: Manuscripts Encoding User Documentation

Eystein Thanisch (eystein.thanisch@cantab.net)

Introduction

“Ba fortrén mo menma. Basa sáithech imtholtanach. No luind dar Albain.
Rofinnainn cach rét.”

- *Scél Túain Maic Cairill*
(ed. J. Carey; adapted)

Purpose of this Document

This document is a guide to marking up, publishing, and querying manuscript transcriptions made for the *Faclair na Gàidhlig* Manuscripts Project (FnaG MSS Project). It provides:

- [INTRODUCTION] An overview of the structure of the corpus;
- [CHAPTERS 1–4] Detailed guidance on marking up transcriptions and encoding palaeographic, textual, linguistic, and codicological data therein;
- [CHAPTER 5] Encoding prefatory material and commentary on the text and manuscript;
- [CHAPTER 6] Navigating XML files and an introduction to “stylesheets” (computer programs that control how transcriptions and contained data are presented);
- [CHAPTER 7] Options for generating editions from the corpus and for querying it for data;
- [CHAPTER 8] A guide to functions enabling quick querying of and commenting on texts;
- [CHAPTER 9] Standard operating procedures for corpus management and version control.

Alongside examples and illustrations in the text, multimedia tutorials and exercises will also be made available to reinforce understanding of key points.

NB: This document covers only tech-related aspects of the Project. Training and ongoing guidance on selecting and working with texts and manuscripts will be provided separately by the Manuscripts Consultant. This document also does not pretend to provide a full course in any of the languages or technologies discussed. Only aspects relevant to the Project are discussed. Links to additional resources and reference works are provided.

Within the document, “the team” includes the following individuals (as relevant in context):

- The *Faclair na Gàidhlig* Project Coordinator: Lorna Pike.
- The Manuscripts Consultant: Prof. William Gillies.
- Both Manuscript Research Fellows: Drs Eystein Thanisch and Martina Maher.
- DASG staff providing technical support and general advice: Dr Mark McConville and Stephen Barrett.

Purpose of this Introduction

This INTRODUCTION provides an orientation within the tech-based aspects of the FnaG MSS Project. It is designed as a relatively superficial tour; subsequent chapters provide full details on how to mark-up transcriptions within the conventions established for the project and how to work with and develop the Project's various ancillary tools.

Cross-references to sections giving full coverage of topics are provided within the text of the INTRODUCTION, as are links to useful external reference works (see Further Resources, below).

Objectives of the FnaG Manuscripts Project

The FnaG MSS Project aims to produce a linguistic corpus representative of Scotland's surviving medieval and early modern Gaelic manuscript literature that can be used by lexicographers in compiling *Faclair na Gàidhlig*. As such, it should span as wide a range of vocabulary, grammar, registers, orthographies, genres, dialects, and periods as possible. This is nothing less than an effort to reconstruct the pre-modern Gaelic universe, from a bewildering number of components and perspectives.

The transcriptions produced must be as comprehensible as possible to *Faclair na Gàidhlig's* lexicographers, as must transcribers' interpretations of the transcriptions. The transcriptions must also include all relevant palaeographical, codicological, and orthographical information/ambiguities and provide linguistic and literary orientation, as any of this might well inform the lexicographers' use of the material. Furthermore, editorial intervention may be required to clarify readings but lexicographers must always have access to the scribal original.

In short, multiple layers of information, and occasionally even bifurcating paths of interpretation, are involved for every word transcribed. It has thus proved not only expedient but fruitful to create digital editions from the manuscript transcriptions and make full use of the additional tools that this unlocks for retrieving and presenting data from within the transcriptions.

This computer-based solution, however, requires high levels of consistency in terms of the mark-up of transcriptions. Hence the present User Documentation has been produced to ensure that consistency is maintained across the project. It will also no doubt prove possible to streamline and enrich the scheme presented here and this User Documentation also aims to form the basis for doing so.

Basic Principles of TEI

Why TEI?

The core of the Project's methodology for encoding the manuscript corpus is the *TEI Guidelines*. TEI (Text Encoding Initiative) is the standard set of conventions in digital humanities for the encoding of texts and data about texts. Keeping within these conventions where at all possible ensures:

1. Functionality; TEI has been being used and improved since the 1980s (e.g. [CELT](#)), meaning that it is highly and robustly responsive to editors' needs.
2. Cross-compatibility; TEI is well-known, so data encoded therein can be easily used by other programs.
3. Documentation; TEI is often used by scholars with no background in coding, so plentiful resources are available to support its users.

Various other project-specific advantages will be set out in the course of this User Documentation.

What is TEI?

The following is a very brief orientation within the key features of TEI and associated technologies, as used in the FnaG MS Project. To gain competence in using the technologies described here, further study and practice will be required (see [FURTHER RESOURCES](#)).

TEI is an adaptation of XML, which is a mark-up language used to render documents and data both human- and machine-readable. XML is fundamental to the Internet and word processing technology, among much else. The rules of XML also apply to TEI, although TEI involves some additional conventions and constraints.

An XML document (that is, a document in natural language marked up in XML) is effectively a tree, with a single trunk and spreading, forking branches (ignore the roots, for the purpose of this analogy). The idea of the document (the trunk) forms a unified and enclosing point of origin, then the various categories of data it contains (limbs), then sub-categories (branches), then the data itself (twigs) is all subdivided within that. The position of a datum in the tree, or perhaps compared to other data, yields further information about it.

The document is sub-divided by *elements*, also known as *tags*. TEI contains a large “vocabulary” of such elements for sub-dividing and categorising the document in different ways. Some are essential to ensure machine-readability; others can be used or not depending on the sort of analysis to which the document is going to be subject.

Elements can be easily recognised in XML code by the angle brackets that enclose their names (e.g. `<element>`). These can either be *open*, meaning that they contain and thus categorise data and possibly other elements (`<element> ... </element>`), or *closed*, meaning that they simply represent a piece of data in their own right (`<element/>`).

Figure 1 (below) presents a very basic TEI document.

Figure 1

The following short document,

Example Document

Here is a paragraph.

And here is another paragraph.

could be marked up as follows:

```
<XML>
  <heading>Example Document</heading>
    <p n="1">
      <w>Here</w>
      <space/>
      <w>is</w>
      <space/>
      <w>a</w>
      <space/>
      <w>paragraph</w>
      <space/>
      <pc>.</pc>
    </p>
    <p n="2">
      <w>And</w>
      <space/>
      <w>here</w>
      <space/>
      <w>is</w>
      <space/>
      <w>another</w>
      <space/>
      <w>paragraph</w>
      <space/>
      <pc>.</pc>
    </p>
  </XML>
```

NB: The above code has been simplified for the purpose of illustration and does not contain everything required in a valid TEI document.

The document itself is represented by the all-enclosing `<XML>` root element. It is then divided into a heading (`<heading>`) and two paragraphs (`<p>`). The heading contains the document's title as a "string".¹ The two paragraphs are each divided into words (`<w>`), spaces (`<space>`), and punctuation characters (`<pc>`). The `<w>` and `<pc>` elements each contain text strings. The `<space>` element is closed in each instance, as no more information is needed to represent a space. All other elements in the document are open.

¹ "String" denotes text (letters or numerals) that appears in code but is not itself code. Rather, it is data. Strings are also known as a "literals", as they are only human-readable and do not have meanings in the context of the code.

Note also how the layout of the code, which is conventional for XML, visually indicates each element's position in the tree hierarchy.

The language of genealogy is sometimes used to describe elements' relationships. An element enclosed directly within another element are said to be the "child" of the element within which it is enclosed, which is in turn described as the "parent" (e.g. <[heading](#)> is the child of <XML> in Figure 1; <XML> is the parent of <[heading](#)>). An element enclosed within another element but with intervening levels of elements between them is said to be the "descendant" and any of its enclosing elements can be called its "ancestor" (e.g. <[pc](#)> is a descendant of <XML> in Figure 1; <XML> is an ancestor of <[pc](#)>). In many cases, an element will be a parent and a child, and an ancestor and a descendant, at the same time.

Elements can also have *attributes*. These classify or distinguish elements or provide further information about the data they contain. Each <[p](#)> element in Figure 1 has an @n attribute,² a sequential number distinguishing the paragraphs from one another.

Each string, each piece of natural language textual data in this document, now appears not only in its own right and in the order in which it was written but also "marked up" within a tree hierarchy that indicates what sort of data it is and what role it plays in the document. Note that how the document is marked up is to a large extent up to the developer. In this instance (as in the FnaG MSS transcriptions, as it happens), each paragraph was sub-divided by word (<[w](#)>) but the decision could have been taken to sub-divide more closely by character (<[c](#)>) or less closely by sentence (<[s](#)>) or to employ all three levels. This would depend on the purpose of the encoding.

Finally, it is important to note that TEI is not just used to encode continuous texts. It can also be used to create databases and is especially applicable to bibliographies and manuscript catalogues. Use is made of this functionality within the FnaG MSS Corpus.

What can be done with TEI?

A number of things could be done with the document in Figure 1.

[Published](#)

Figure 1 could be "published" in its entirety, with all text strings rendered in a form that is just human-readable (webpage, MS Word document, PDF, etc.). This means transforming the XML/TEI into another language: HTML (more on this below). HTML resembles XML in that it also sets out the document in a tree hierarchy of elements. However, its element vocabulary is different. Also, crucially, HTML documents can be directly rendered by web browsers and word processors.

An HTML rendering of the XML from Figure 1 is presented in Figure 2:

² @ is prefixed to attribute names when referencing them but not when including in the XML code itself.

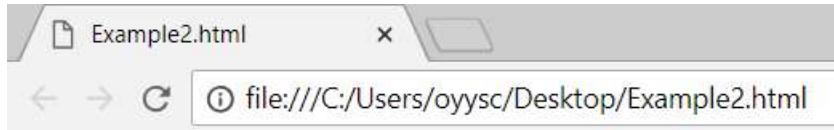
Figure 2

```
<html>
  <body>
    <h1>Example Document</h1>
    <h2>1</h2>
      <p>Here is a paragraph.</p>
    <h2>2</h2>
      <p>And here is another paragraph.</p>
  </body>
</html>
```

The `<h1>` element indicates a top-level heading.

Saving this as an HTML file (extension: .html) and opening it in a web browser, or in MS Word, results in the following being displayed (Figure 3):

Figure 3



Example Document

1

Here is a paragraph.

2

And here is another paragraph.

Note that the browser has automatically formatted the text child of the `<h1>` element in bold and large font and inserted line breaks between the two paragraphs.

However, the transformation from XML (which cannot be rendered as in Figure 3) to HTML (which can) does not happen automatically. The XML/TEI elements used in marking up the document have no inherent meaning or relationship to HTML elements; they are not “predefined”. For example, despite its name, the `<space>` element will not automatically produce a space. And the text content of `<heading>` will not necessarily be treated any differently from text elsewhere.

In HTML, meanwhile, elements are predefined in terms of how they are presented by web browsers and word processing software. The text content of `<h1>` will always be in bold and large font; the text content of HTML `<p>` elements will always be separated by line breaks. This can be altered or overridden, however.

Another document — a “stylesheet” — is needed. The stylesheet will contain machine-readable instructions (declarative computer programs) on how the different elements in the XML document are to be transformed into the “published” HTML version.

In Figure 2, the stylesheet will have instructed that the text string contained in the XML `<heading>` element be placed in an HTML `<h1>` element (bold and large font) but that text descended from each XML `<p>` elements be placed in an HTML `<p>` element (without formatting, separated by line breaks). It instructed that the paragraph numbers (@n) be presented within `<h2>` elements.

Note that the stylesheet does not have to provide for all the content in the XML document. For example, if it had been decided that the paragraph numbers or the heading in the XML document were not needed, and provision for them had not been made in the stylesheet, then they would not appear in the HTML document.

Since only HTML can be directly presented as an easily readable document, one might ask why texts are not just marked up in HTML, without the need for TEI/XML. This is possible. However, the separation of the XML/TEI input, the stylesheet process, and the HTML output yields big opportunities and efficiencies.

The stylesheet file is separate from the XML/TEI file, as is the resulting published HTML output. So, however many stylesheets are created and however many transformations of the XML/TEI to HTML are performed, the original XML/TEI remains the same. As a result:

1. Multiple HTML versions of the same underlying XML/TEI document can be generated, applying different editorial principles, or selecting different content, each time.
2. The HTML output can often be amended via a single edit of the stylesheet rather than a series of changes to every instance of the element involved in the XML (not unlike when using Styles in MS Word).

XSLT and CSS are the two most widely used stylesheet languages. The FnaG MSS Project stylesheets use mainly XSLT, with occasional CSS. This is discussed further in CHAPTER 6.

Queried

The XML document in Figure 1 could also be queried and various pieces of information obtained about the document. For example, a program could retrieve the number of paragraphs containing the word “another”, or the number of occurrences of the word “paragraph”, or a list of all words appearing in the main text of the document ordered alphabetically, with or without repetition of the same word. Obviously, this only becomes seriously useful in the case of documents that are longer and more complex than that presented in Figure 1.

This could be done via another stylesheet. Instead of formatting the whole document, the stylesheet would instruct that only data meeting certain criteria and/or sorted a certain way be published in the HTML output. For example, if an alphabetised list of words contained in the Figure 1 XML document referenced by paragraph were required, the following HTML could be generated via a stylesheet (Figure 4):

Figure 4

```
<html>
  <body>
    <table>
      <tr>
        <th>Word</th>
```

```

        <th>Para Number</th>
    </tr>
    <tr>
        <td>a</td>
        <td>1</td>
    </tr>
    <tr>
        <td>and</td>
        <td>2</td>
    </tr>

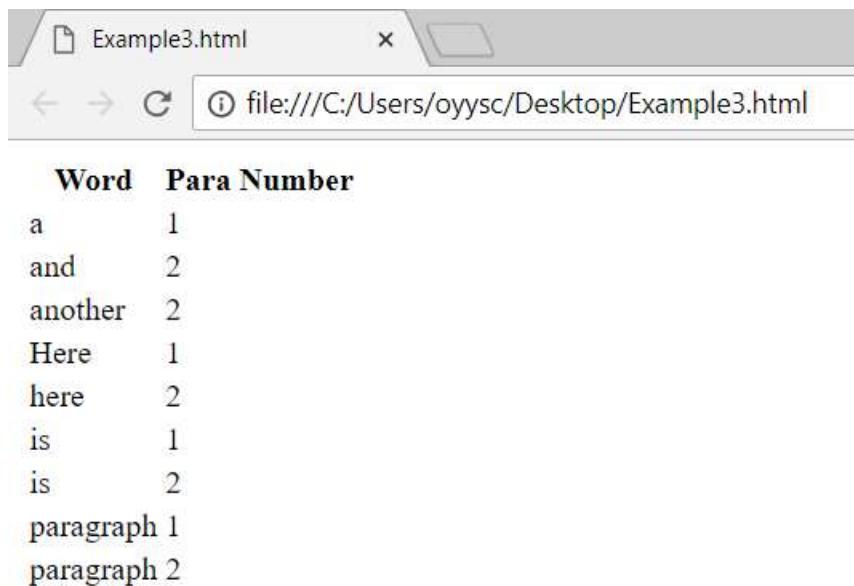
    <tr>
        <td>another</td>
        <td>2</td>
    </tr>
    <tr>
        <td>Here</td>
        <td>1</td>
    </tr>
    <tr>
        <td>here</td>
        <td>2</td>
    </tr>
    <tr>
        <td>is</td>
        <td>1</td>
    </tr>
    <tr>
        <td>is</td>
        <td>2</td>
    </tr>

    <tr>
        <td>paragraph</td>
        <td>1</td>
    </tr>
    <tr>
        <td>paragraph</td>
        <td>2</td>
    </tr>
</table>
</body>
</html>

```

This displays as a table in a web browser, or in MS Word or MS Excel (Figure 5):

Figure 5



A screenshot of a web browser window titled "Example3.html". The address bar shows the URL "file:///C:/Users/oyysc/Desktop/Example3.html". The content of the page is a table with two columns: "Word" and "Para Number". The data is as follows:

Word	Para Number
a	1
and	2
another	2
Here	1
here	2
is	1
is	2
paragraph	1
paragraph	2

The stylesheet language, XSLT, can also be used to calculate stats or produce yes/no and factual answers to questions as well as print and arrange text strings.

However, a quicker way to run some queries is to use xPath or XQuery, which are dedicated languages for querying XML. These do not require stylesheets or transformation to HTML.

The various means of querying are discussed further in CHAPTERS 6 and 7.

Review

To summarise this brief survey of TEI:

- TEI is a widely used and understood framework for encoding texts.
- TEI is an adaptation of XML.
- TEI/XML can be used to mark up (i.e. subdivide) a text to give the editor greater control and concurrent options concerning its publication.
- To be rendered by a web browser/word processor, TEI/XML must be transformed into HTML.
- Publishing a text marked up in TEI/XML as HTML requires one or more stylesheets.
- A marked-up text can also be queried for information, using either a stylesheet or a query language.

The FnaG Manuscript Corpus: A Tour

Having surveyed the basic issues and benefits of using TEI, this section considers the FnaG MSS Corpus and how these technologies are employed in this particular context.

Corpus Contents

Figure 6 contains a summary of the transcriptions currently included in the FnaG MSS Corpus:

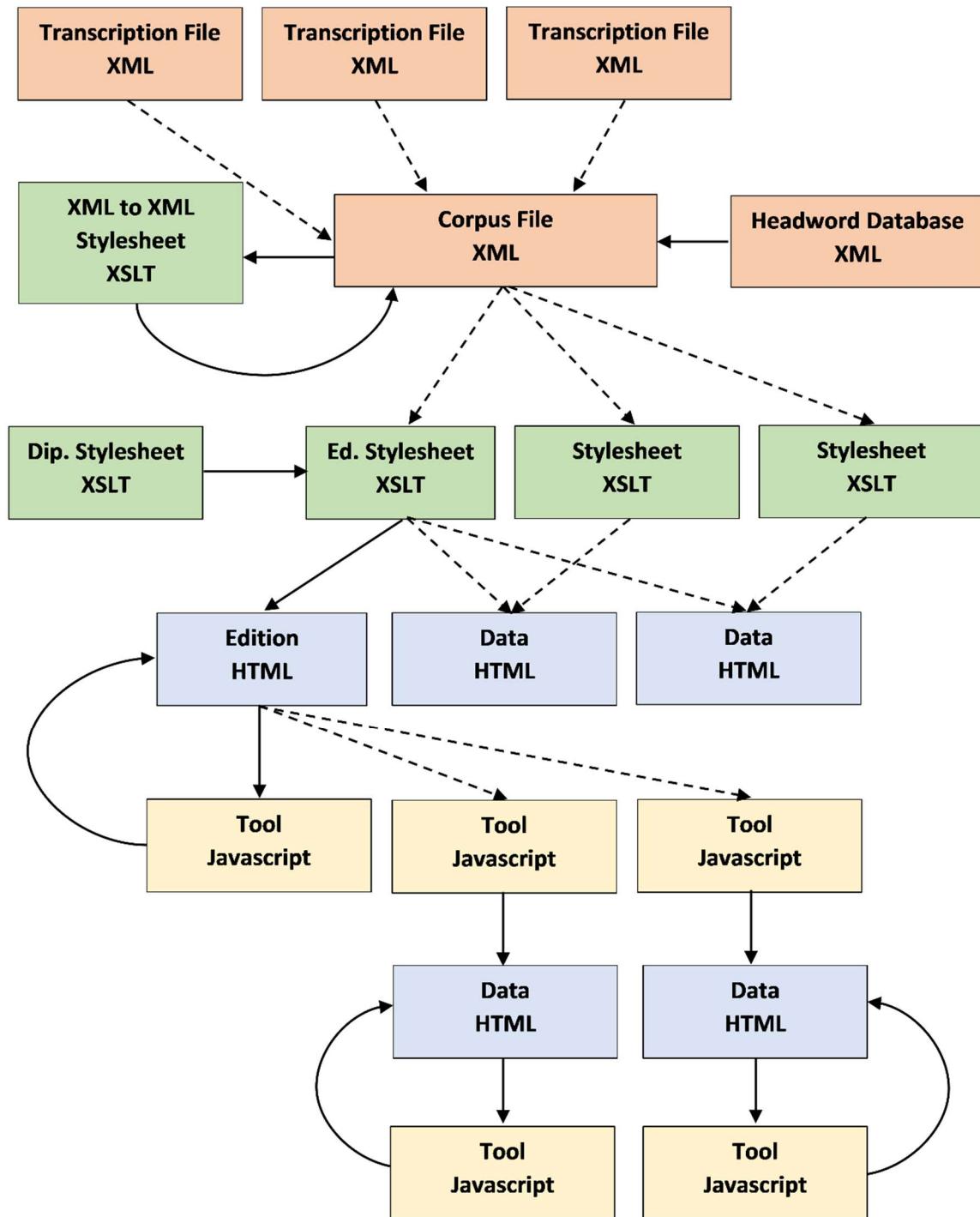
Figure 6

Transcr. File	Manuscript	Text	Comment
transcription1.xml	CUL I.i.6.32	Gaelic Notes from the Book of Deer Gaelic Rubric and Colophon	
transcription2.xml	BL, Egerton 2899	'Aingil Dé dom dhín'	
transcription3.xml	NLS 72.2.11	<i>Cath Fionntrágha</i>	ms. pp.1–4 only
transcription4.xml	NLS 72.1.1	<i>Dialogue between SS Anselm and Mary</i> <i>Extracts from Liber Scintillarum</i>	<i>Dialogue in progress</i>
transcription5.xml	NLS 72.1.33	Medical Charms <i>Extracts from the Colliget</i> <i>MacLean's Letter to his Physician</i>	
transcription6.xml	NLS 72.1.31	Columban Poetry <i>The MacDougall Charter</i>	Columban Poetry only partly legible.
transcription7.xml	NLS 72.2.2	Selection of bardic and sub-bardic poetry	
transcription8.xml	Bod. Lib, Arch. Seld. B24	Gaelic marginal poem in <i>The Kingis Quair</i> manuscript.	
transcription9.xml	NLS 72.1.50	Scottish and Irish historical texts	In progress
transcription10.xml	NLS 1745	Grammatical tract	Not yet begun
transcription11.xml	NAS RH.6/218	Argyll Charter	Not yet begun
transcription12.xml	NLS 72.1.36	'Soraidh slán don aoidhche areir' <i>Laoiadh an Deirg</i>	<i>Laoiadh</i> in progress.
transcription13.xml	EUL Laing III.21	<i>Astrologico-Medical Tract</i>	
transcription14.xml	NLS 72.1.2	<i>Tract on Categories</i>	
transcription15.xml			
transcription16.xml	RIA E.i.3 (778)	Poems by Cathal Mac Mhuirich	In progress

File Structure

The components of the FnaG MSS Corpus and the computer languages involved are set out in Figure 7.

Figure 7



Arrows indicate either transfer of data or application of transformation rules. A solid arrow indicates the only transfer/transformation that can take place; a dashed arrow indicates one of several possibilities. The number of files in the diagram does not necessarily reflect the number of files of that type in the Corpus; for example, there are currently 15, not three, transcription files in the Corpus and that number will rise.

Each of type of component will now be considered in detail.

Input

Transcription Files (XML)

A transcription file consists of one or more marked up transcriptions from a single manuscript, as currently bound and shelfmarked. It may comprise multiple texts and the work of multiple scribes. As set out in CHAPTERS 1–5, there are conventions in TEI for differentiating these within the file.

The file consists of a Header ([`<teiHeader>`](#)) and a Body ([`<body>`](#)).

The Header contains introductory material describing the manuscript (physical characteristics, history, contents etc.) and its scribal hands. It also contains a catalogue of texts transcribed from the manuscript, with commentary on each text's literary and linguistic characteristics. For more information, see CHAPTERS 1 and 5.

The Body contains the actual marked-up transcriptions. It can consist of multiple distinct texts from within the manuscript, which can each in turn contain further subdivisions (e.g. stanzas in poetry); see CHAPTER 1. The basic unit of the transcription is the word ([`<w>`](#)), alongside spacing ([`<space>`](#)) and punctuation characters([`<pc>`](#)); see CHAPTER 3. These are placed within paragraphs ([`<p>`](#)) in the case of continuous prose, or within lines ([`<l>`](#)) and “line groups” ([`<lg>`](#)) in the case of verse or clearly subdivided prose (e.g. collections of aphorisms); see CHAPTER 2.

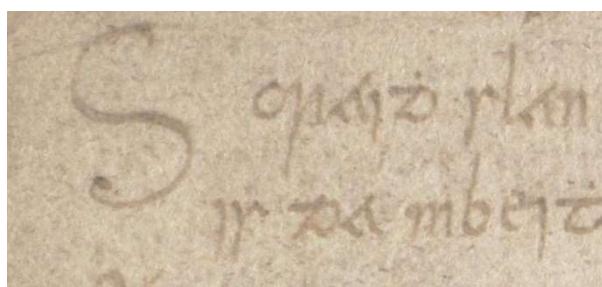
As well as imposing an editorial structure on the text (e.g. imposing line division in poetry), the page layout in the manuscript (lines ([`<lb>`](#)), columns ([`<cb>`](#)), pages ([`<pb>`](#))) is imposed separately and in parallel, meaning both can be represented in the output; see CHAPTER 2.

Code Sample

How all this is done is described fully in the referenced Chapters. For illustrating the interactions between the different components of the corpus, however, a short example of how a word is marked up is required.

Figure 8 shows the first two words of the poem ‘Soraidh slán don aoidche areir’ in NLS Adv. 72.1.36.

Figure 8: NLS Adv. 72.1.36, fol.85v2.



These words and their immediate context is marked up as follows (Figure 9):

Figure 9

1	<TEI>
2	<teiHeader>
3	<!-- INTRODUCTORY MATERIAL HERE -->
4	</teiHeader>
5	<text>
6	<body>
7	<div n="1" type="verse" resp="Hand9" corresp="MS12.1">
8	<pb n="85v" xml:id="MS12.85v"/>
9	<lg n="1" type="stanza" xml:id="MS12.1.1">
10	<l n="a" xml:id="MS12.1.1a">
11	<lb n="2" xml:id="MS12.85v.2"/>
12	<w lemma="soraid" lemmaRef="http://www.dil.ie/3849 4" ana="noun">
13	Sorai
14	<abbr cert="high">
15	d
16	<g ref="g4"> h </g>
17	</abbr>
18	</w>
19	<space type="scribal"/>
20	<w lemma="slán" lemmaRef="http://www.dil.ie/3783 4" ana="adj">
21	slan
22	</w>
23	<space type="scribal"/>
24	<!-- TEXT CONTINUES HERE -->
25	</l>
26	</lg>
27	</div>
28	</body>
29	</text>
30	</TEI>

NB: The above code has been simplified for the purpose of illustration and does not contain everything required in a valid TEI document.

Corpus File (XML)

The transcription files each contain information relating to the corresponding manuscript and texts therein. The corpus file is a single TEI file (root element: <[teiCorpus](#)>) containing data that could potentially be of relevance to any of the transcription files. For example, in Figure 8, l.16, the @ref “g4” is a unique id number in a list of manuscript abbreviations (“glyphs”) stored in the corpus file. It stands for a *spiritus asper*: a lenition dot or tick. Hence the <[g](#)>, which expands the abbreviations, contains “h”. And, at l.7, @resp designates the main (in fact, sole) scribe of the text as “Hand9”. This is the unique id for Eoghan Mac Gilleoin (fl. 1699) in the database of scribal hands stored in the corpus

file.³ The corpus file also includes a TEI bibliography of primary and secondary literature. The glyph and scribal hands databases both contain further information; see CHAPTERS 1 and 5.

The corpus file is drawn upon as needed by stylesheets when generating editions or data reports. In terms of XML structure, it actually forms a super-header for all the transcription files in the corpus (Figure 10). The files are stored separately but the `<xi:include>` element combines the specified files as a corpus. The stylesheet transformation is then executed on the corpus file.

Figure 10

```
<teiCorpus>
    <teiHeader>
        <!-- CORPUS DATA -->
    </teiHeader>
    <xi:include href="Transcriptions/transcription1.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription2.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription3.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription4.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription5.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription6.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription7.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription8.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription9.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription10.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription11.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription12.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription13.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription14.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription15.xml" parse="xml"/>
    <xi:include href="Transcriptions/transcription16.xml" parse="xml"/>
    <xi:include href="hwData.xml" parse="xml"/>
</teiCorpus>
```

Headword Database (XML)

As will be observed from Figure 10, another XML file is also included in the corpus: hwData.xml. This is a database of headwords that have already appeared at least once in one of the transcription files, as identified by the `@lemmaRef` URL within the `<w>` element (e.g. Figure 9, ll. 12, 20). The headword entry in the headword database contains references for the headword to eDIL and/or Dwelly (via *Faclair Beag*). In the case of nouns and adjectives, data on gender and stem is also included. For example, Figure 11 contains the entry on *soraid*, as used in Figure 9:

Figure 11

1	<entryFree corresp="http://www.dil.ie/38494" n="3">
2	<w type="data" lemma="soraid" lemmaRef="http://www.dil.ie/38494" ana="adj" lemmaDW="soraidh" lemmaRefDW="http://www.faclair.com/ViewDictionaryEntry.aspx?ID=E989F7F7C98DEE6DD023B9CA652A935C"/>
3	<gramGrp type="noun">
4	<gen>f</gen>
5	</gramGrp>
6	<gramGrp type="adj">

³ Some hands appear in multiple manuscripts (e.g. Hand8, Eoghan mac Pháil), hence information on them being stored separately.

7	<iType>i</iType>
8	</gramGrp>
9	</entryFree>

When marking up transcriptions, or transcribing directly into TEI, the headword for each `<w>` is located in eDIL if possible or, failing that, in Dwelly (via *Faclair Beag*). The headword and URL are then added as the `@lemma` and `@lemmaRef` respectively within the transcription. If the headword/URL come from eDIL, then their counterparts from Dwelly can be added within the appropriate entry in the headword database (`@lemmaDW` and `@lemmaRefDW`). This means that the modern Scottish Gaelic form need only be retrieved once for each headword, rather than for every instance of that headword in the transcriptions.

Where applicable, morphological data on stem and gender can also be added within the entry in the headword database,⁴ with similar advantages.

As well as being a useful piece of data in its own right, the `@lemmaRef` URL forms a unique ID that links its entry in the headword database to each occurrence of the headword in the transcriptions. This allows the data it contains to be brought in when compiling editions and used as criteria when phrasing queries. For example, words of certain stems or genders can be retrieved by searching for the all the `<w>` elements with the same `@lemmaRef` as entries in the headword database the desired `<gender>/<iType>` element.

Schema (RNG)

The TEI/XML input files also require an XML schema. This is a document that lists the elements and attributes used and the rules that define their proper use. XML development environments like oXygen make use of the schema to ensure that TEI/XML code is valid and consistent.

The FnaG MSS Project uses the schema fnag_mss2.rng; this contains the TEI elements used in the transcription files, with some customisation to meet particular needs. For example, TEI includes `@lemma` and `@lemmaRef` attributes of the `<w>` element for the encoding of a dictionary headword and a URL to a dictionary entry. These attributes have been retained to point to the default headword for a word but `@lemmaED/@lemmaRefED`, `@lemmaDW/@lemmaRefDW`, and `@lemmaSL/@slipRef` have been added on the same model, to specify headwords from eDIL, Dwelly, and the HDSG-A MSS Slips respectively.

Process

Stylesheets (XSLT)

Depending on their purpose, stylesheets make use of various combinations of the three types of input file described above (transcription files, the corpus file, and the headword database file) to produce the desired HTML output. This desired output can be an edition or one of a very wide range of data reports.

Stylesheets are themselves often used in combination. To produce a “full edition” (as currently available within the FnaG MSS Project), two stylesheets are used in combination:

- `stylesheet-ed-pub.xsl` generates a formatted header for each selected transcription file and a text of the transcriptions containing all editorial interventions and laid out according to

⁴ The incompleteness of the data in the entry in Figure 11 is due to omissions in the dictionary entries.

editorial analysis (e.g. verse is presented in metrical lines; regular word division is imposed based on `<w>` elements).

- `stylesheet-dip-comp.xsl` (applied with `stylesheet-ed-pub.xsl` via an `<xi:include>` element) appends a diplomatic text to the editorial version, in which text and layout follow the manuscript as closely as possible (e.g. verse is presented in manuscript lines; word division is derived from `<space type="scribal">` elements (e.g. Figure 9, ll. 19, 23)).

Versions of these stylesheets can also be used separately.

Other stylesheets generate various sorts of reports. These include alphabetised lists of forms with information on context (`stylesheet-ed-ana-general.xsl`), possibly limited by scribal hand, date, headword, part of speech, or suchlike (`stylesheet-ed-ana-specific.xsl`). These are generated using a dedicated stylesheet that also draws on `stylesheet-ed-pub.xsl` to format forms and their context. Yet another sort of stylesheet returns statistical information about the corpus (e.g. `stylesheet-ed-ana-stats.xsl`); this can function independently.

Finally, XSLT stylesheets can also transform XML (including TEI) files without changing their format. This means that they can be used to perform mass changes to the input XML files. For example, `stylesheet-xml-hwData-update_2.xsl` updates the headword database (`hwData.xml`) with new headwords that have been added to the corpus.

Some of these stylesheets involve large numbers of calculations and can take a very long time. For example, applying `stylesheet-ed-ana-general.xsl` to the whole corpus takes approximately 48 hours. This issue is being looked into and more streamlined versions of the same functions will be developed in due course.

Output

The stylesheets surveyed in the previous section, when applied to the XML input, generate a range of outputs, mostly in HTML. Those that contain “pub” in the filename generate some sort of complete text of the selected transcription(s). Those that contain “ana” in the filename generate some sort of data report within the scope of the selected transcriptions.

These various outputs are briefly surveyed below.

[Editions \(HTML\)](#)

The editorial text, generated using `stylesheet-ed-pub.xsl`, of ‘Soraidh slan don aoidhche areir’ from NLS Adv. 72.1.36 is presented in Figure 12.

Figure 12

85v: Eoghan Mac Gilleoin (Hand9) [Collect Slips](#) [Headword Search](#)

1.
1. Soraidh slan don aoidche [a]reir :
fada gear ag dul ar gcuil
2. is da mbeith ansa dom cuir a ccroich :
is truagh nach é anocht a tús

[Add Comment](#)

2.
3. Ata dias istighsi anocht :
ar nach ceillin in rosc rún
4. gioldh nach bfuilid beil ar bheil .
is géir géir a silladh súl

[Add Comment](#)

3.
5. Truagh an chuibhreach do ni an chiall :
re silladh suirghech na súl
6. ni ferda an tost do ni an beil :
sgeil do ni a rosc ar a rún

[Add Comment](#)

7. ~~Ata dias istighsi anocht~~

[Add Comment](#)

4.
8. Cho leig egla lucht na mbréig :
smid ar mo bheil a rosc mall
9. achd tuigsi ani adéir mo shuil :
agus tu san chul ud thail

[Add Comment](#)

5.
10. Cuinamh dhuine an aoidhche anocht :
is truag gan sin mar so go brath
11. na leig an mhaidin astech :
eiridh is cuir amach lá

[Add Comment](#)

6.
12. Och a mhuiре a bhuiime tseng :
os tu is cean as gach cleir
13. tarthuidh agus gabh mo lamh :
soraidh slan don naoidche [a]reir

The diplomatic text generated in conjunction with it (using stylesheet-dip-comp.xsl) is presented in Figure 13:

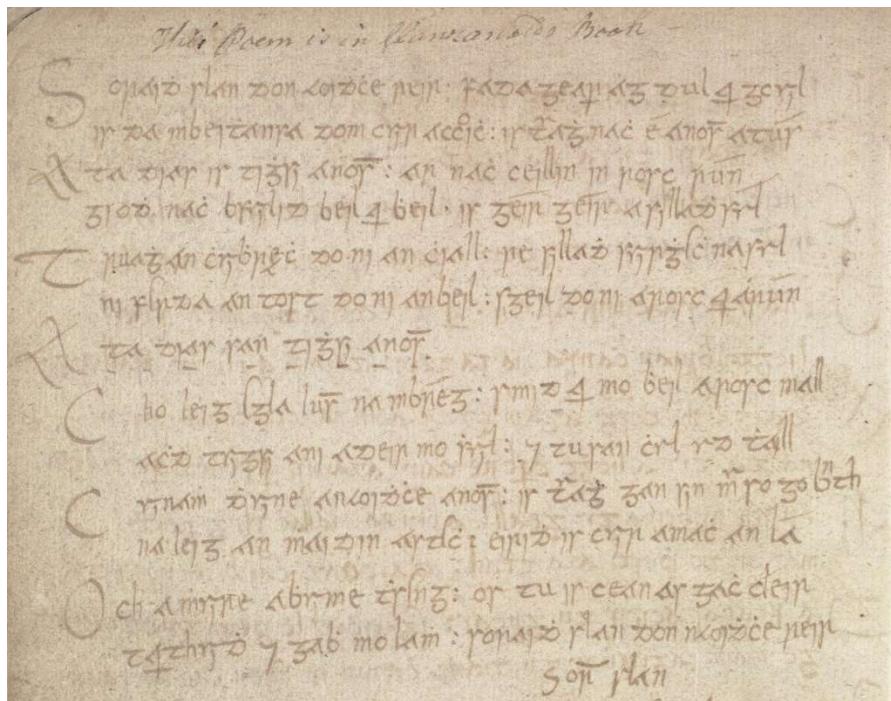
Figure 13

85v: Eoghan Mac Gilleoin (Hand9) [Collect Slips](#)

1. Soraidh slan don aoidche reir: fada gear ag dul ar gcuil
2. is da mbeithansa dom cuir accroich: is truagh nach é anocht atús
3. Ata dias is tighsi anocht: ar nach ceillin in rosc rún
4. gioldh nach bfuilid beil ar bheil. is géir géir asilladh súl
5. Truaghan chuibhreach doni an chiall: re silladh suirghech na súl
6. ni ferda an tost doni an beil: sgeil doni arosc ararún
7. ~~Ata dias is tighsi anocht~~
8. Cho leig egla lucht na mbréig : smid ar mo bheil arosc mall
9. achd tuigsi ani adéir mo shuil : *agus* tusan chul ud thail
10. Cuinamh dhuine anaoidhche anocht : is truag gan sin mar so go brath
11. na leig an mhaidin astech : eiridh is cuir amach lá
12. Och a mhuiре ab/buime tseng: os tu is ceanas gach cleir
13. tarthuidh agus gabh mo lamh : soraidh slan don naoidche reir

Note that, in Figure 13, line division is as per the manuscript and the supplied “a-” in line 1 does not appear, for example. The text is still somewhat adapted, however: abbreviations are still expanded and the deleted text in line 7 is still rendered via strikethrough, not via dots beneath the line, as in the manuscript (see Figure 14).

Figure 14



As well as formatting and printing the transcription in these two ways, based on two different interpretations of the original TEI/XML, these stylesheets also embed data about each word in the text within the HTML. This data is derived both from the TEI/XML mark-up made in the transcription file itself and from the corpus file and headword database. Each word is contained within an HTML element, with data added as attributes of this element.

The element for the first word, “Soraidh”, is presented in Figure 15 (some of this data supports various functions and is not meant to be immediately intelligible; this is covered in more detail in CHAPTER 8).

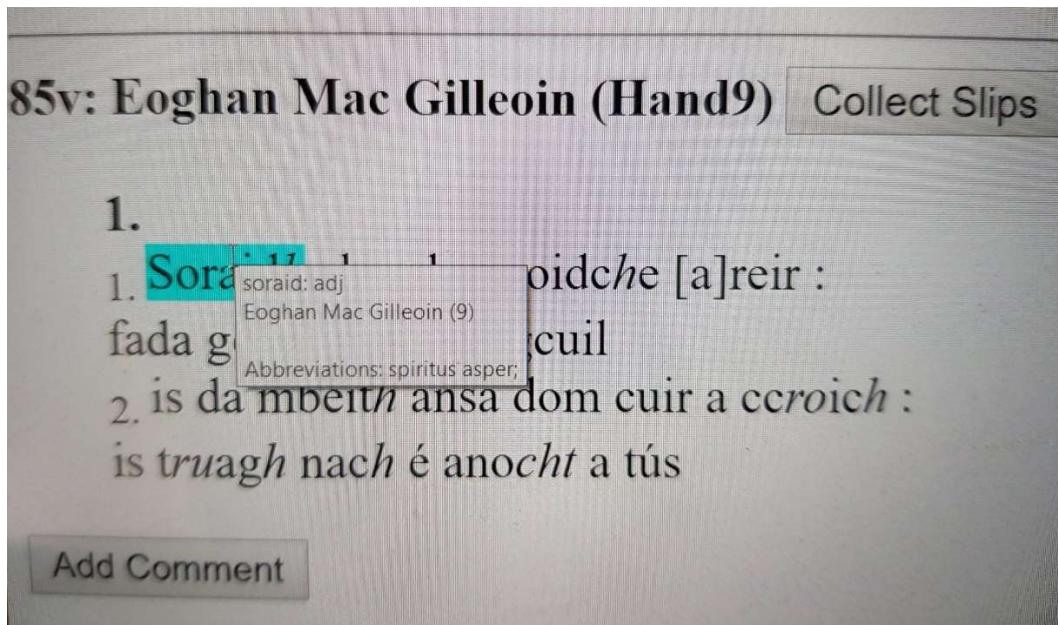
Figure 15

```
<a id="d2e2317" pos="1388" onclick="addSlip(this.id)"  
onmouseover="hilite(this.id)" onmouseout="dhilite(this.id)"  
lemma="soraid" lemmaRef="http://www.dil.ie/38494" lemmaED="soraid"  
lemmaRefED="http://www.dil.ie/38494" lemmaDW="soraidh"  
lemmaRefDW="http://www.faclair.com/ViewDictionaryEntry.aspx?ID=E989F7F7C9  
8DEE6DD023B9CA652A935C" lemmaSL="" slipID="" ana="adj" hand="Eoghan Mac  
Gilleoin (9)" ref="MS12 85v 13" date="17" medium="Verse" cert=""  
abbrRefs="www.vanhamel.nl/codecs/Spiritus_asper" lineID="MS12.85v.13"  
title="soraid: adj &#xA;Eoghan Mac Gilleoin (9) &#xA;&#xA;Abbreviations:  
spiritus asper; &#xA;" style="text-decoration:none;&#xA;&#x9;&#x9;&#x9;&#x9;&#x9;&#x9;&#x9;&#x9;&#x9;&#x9; color:#000000" class="ed">soraid<i id="10" cert="high">h</i></a>
```

This data can be accessed in two ways.

First, some basic information is presented as a “tooltip” when the mouse is placed over the word (Figure 16).

Figure 16



Secondly, the data on the selected word can be exported to a separate HTML file, which also contains hyperlinks to further external resources. Figure 17 shows a table of the first line of ‘Soraidh slán don aoidhche areir’, generated by clicking on each word in the line in “Collect e-Slips” mode.

Figure 17

Form	MS Ref.	MS Context	Text Form	Issues?	Abrevs	Scribe	Scribe Date	Pos	Lemma (eDIL)	Lemma (Dwellys)	HDSG Slip	URI	Del Col	Del Row
Soraidh	MS12 85v 1	Soraidh slán don aoidhche [a]reir	Verse		Spiritus_asper	Eoghan Mac Gilleoin (9)	17	adj	soraid	soraidh		38494	Del Row	Comment
slán	MS12 85v 1	Soraidh slán don aoidhche [a]reir fada ...	Verse			Eoghan Mac Gilleoin (9)	17	adj	slán	slán		37834	Del Row	Comment
don	MS12 85v 1	Soraidh slán don aoidhche [a]reir fada gear ...	Verse			Eoghan Mac Gilleoin (9)	17	prep, art	do			17096	Del Row	Comment
aoidhche	MS12 85v 1	... Soraidh slán don aoidhche [a]reir fada gear ag ...	Verse		Spiritus_asper	Eoghan Mac Gilleoin (9)	17	noun	adaig	oidhche		256	Del Row	Comment
[a]reir	MS12 85v 1	Soraidh slán don aoidhche [a]reir fada gear ag dul ...	Verse	supp. char(s)		Eoghan Mac Gilleoin (9)	17	adv	irair			29082	Del Row	Comment

This is accomplished via a JavaScript function; for more information, see below.

Data (HTML)

The data that can be generated from the underlying TEI/XML transcription files (whether from the whole corpus or from selected transcriptions) is rich and almost infinitely varied. XSLT, as used in the stylesheet, can select data down to a very fine level of precision or be used to phrase complex queries.

A few examples are considered below. Most of these generate tables that can be opened in MS Excel or MS Word (as is evident from the examples in the Appendices).

General Data (stylesheet-ed-ana-general.xls)

A complete word list, with data on each word, can be generated using `stylesheet-ed-ana-general.xsl`. This finds each `<w>` element (i.e. each word) within the defined scope (the whole corpus or selected transcriptions), sorts them by part of speech, and adds a row to the appropriate section of an HTML table containing the form, the headword, context, and other information. For such a table, generated from ‘Soraidh slán don aoidhche areir’, see [Adding a New Text](#)

Certain tasks need to be performed when adding a new text to a transcription file regardless of its form. We will cover these before proceeding to examine what is specific to each form. We will use the “Definitions” on fol. 24^r of NLS Adv. 72.1.4 as our initial example.

Corpus Data

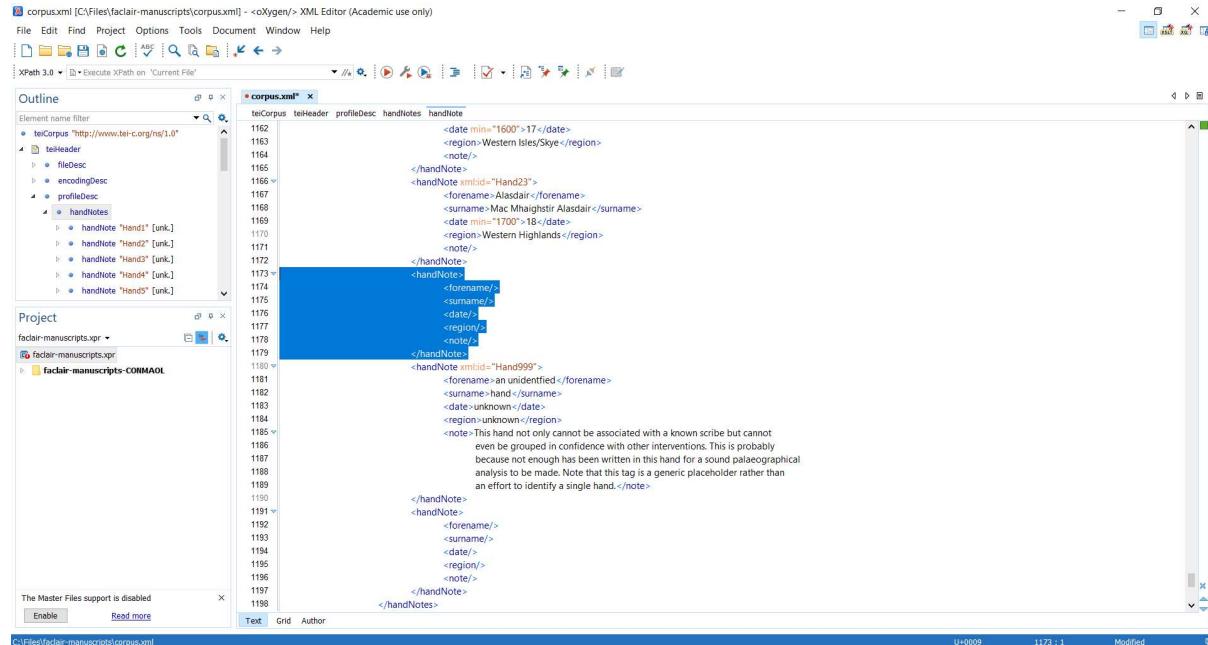
First, identify the main scribe of the text, whether from a catalogue entry or from your own research. In our example, the scribe (according to Black 2011) has not been identified.

In `corpus.xml`, go to `<handNotes>` (Appendix 6a l.156). Check whether this hand already has a record, using the catalogue entry or your own research to identify whether it appears in any other manuscripts from which it might have been transcribed. Our example text is in NLS Adv. 72.1.4’s Hand 2, which Black identifies with Hand 5 of NLS Adv. 72.1.27. This manuscript is not yet included in the corpus so it is safe to assume this is the hand’s first appearance.

If the hand were already recorded in `corpus.xml`, note the value of the `@xml:id` attribute of the corresponding `<handNote>` element.

In this case, however, copy the empty `<handNote>` template element plus all descendants and paste it in within `<handNotes>`, after the most recent identified hand and before the entry for Hand999.

Figure 32



Assign “Hand” plus the next sequential hand number to it as the value of @xml:id. In our example, this would be “Hand24”. The hand is anonymous and the dates of the hand are also thus uncertain; however, what we are calling Hand24 begins a text at fol. 25^r6, after a note in the hand of Níall Óg mac Néill mhic Bheathadh. As the latter has a floruit in 1541 (Bannerman 1998: 137), we can assume provisionally that Hand24 belongs to the 16th century. The Beaton hands that appear in the manuscript are belong to individuals associated with Kilelane, Islay, but we cannot be sure that this is where the other hands conducted their activity.

Our initial record of this hand might thus look as follows (Figure 33):

Figure 33

```
<handNote xml:id="Hand24">
    <forename>unk.</forename>
    <surname>unk.</surname>
    <date type="floruit" min="1500" max="1600">16</date>
    <region cert="medium">Islay</region>
    <note>
        <p>Hand 2 in NLS Adv. 72.1.4 and Hand 5 in NLS Adv. 72.1.5.</p>
    </note>
</handNote>
```

In time, a full description of this hands traits can be written up under the `<note>` element, based on encounters with the hand across the project. For now, the briefest cross references are provided.

Transcription File Data

Next, we create an entry for the new text in the transcription file header. Under `<msContents>` (Appendix 5a, I.34), make a copy of the template for `<msItem>` plus all descendants and paste it in, also as a child of `<msContents>`.

Within this `<msItem>` element, add the details about the text that you have available. This will vary but the following are recommended (line references to Appendix 5b, where the elements are annotated with more information on what is required):

- `<msItem n="itemN" xml:id="itemID" resp="HandXXX">` (I.117):
 - @n is a sequential number, indicating this text’s position of the array of texts transcribed from this manuscript, in the order in which they appear in the manuscript. This often requires some forethought as to which texts from the manuscript are likely to be transcribed. In this case, however, the Definitions on fol. 24^r of NLS Adv. 72.1.4 comprise the first Gaelic text in the manuscript. So @n can safely = “1”;
 - @xml:id is the manuscript’s id, plus @n. So @xml:id = “MS17.1”, in this case;
 - @resp indicates the main scribe responsible for the text; as per the `<handNote>` we just added, @resp = “Hand24”;
- `<locus>` (I.39): fol. 24^r;
- `<incipit>` (I.40): “Re sine”;
- `<explicit>` (I.41): “co haicidach”.

The initial entry in `<msItem>` should thus resemble the following (Figure 34):

Figure 34

```
<msItem n="1" xml:id="MS17.1" resp="Hand24">
    <locus>fol. 24<hi rend="sup">r</hi></locus>
    <title>Definitions</title>
    <incipit>Re sine ...</incipit>
    <explicit>... co haicidach</explicit>
    <!-- Further elements here (can remain empty) -->
</msItem>
```

Finally, in the transcription file <handDesc> element (Appendix 5a, l. 123), add a <handNote> element, with a @corresp attribute with the id of the main scribe of the text as a value: so,

Figure 35

```
<handNote corresp="Hand24">
    <note>
        <!-- Further elements here (can remain empty) -->
    </note>
</handNote>
```

Project Record Data

As a last step before we proceed to add the text proper to the transcription file, open the “Project” Dropbox folder, then open “xmlid_lists.xlsx”. On the “Hands” sheet, if it be new to the corpus, add details of the text’s main hand.

The New Text

The <div> Element

Finally, we can add the first element of the text itself. Within the transcription file’s <body> element (Appendix 5a l.169), at the position corresponding to the text’s position in the manuscript relative to the other texts included in the transcription file, add a <div> element. (Appendix 5a l.170). This is the delimiter of one text from another in the file.

The <div> element requires various attributes:

- @n: this is the sequential number for the text, as in <msItem> @n;
- @type: this is the form of the text (“prose”, “verse”, or “divProse”);
- @resp: the main hand responsible for the text;
- @corresp:= <msItem> @xml:id;

The <div> element we insert for the “Definitions” on fol. 24r of NLS Adv. 72.1.4 would be as follows:

Figure 36

```
<text>
    <body>
        <div n="1" type="divProse" resp="Hand24" corresp="MS17.1">
            <!-- Text continues here -->
        </div>
    </body>
```

```
</text>
```

Form-specific Elements

So far, only `<div> @type` is dependent on the text's form. The elements added within `<div>`, however, will determine the basic editorial structure of the text and so are dependent on its form.

Prose

Prose text is simply contained inside `<p>` (paragraph) element, which is a child of `<div>`. One `<p>` element is used to enclose the entire text. Page (`<pb>`), column (`<cb>`), and line markers (`<lb>`) are then placed within the `<p>` element and the text proceeds from there.

Bruighion Bheg na hAlmhain, from NLS Adv. 72.1.36, would be framed as follows (Figure 37):

Figure 37

```
<text>
  <body>
    <div n="2" type="prose" resp="Hand9" corresp="MS12.2">
      <p>
        <pb n="96r" xml:id="MS12.96r"/>
        <lb n="1" xml:id="MS12.86r.1"/>
        <!-- Text continues here -->
      </p>
    </div>
  </body>
</text>
```

Verse

Verse is structured within an `<lg>` (“line group”) element (in this case, a stanza, defined using `@type`), then a `<l>` (“line”), corresponding to the analysis of the metre applied by the transcriber.

Due to stylesheet requirements, the location of the `<pb>` and `<lb>` elements is important:

- `<pb>` must be placed outside the `<lg>` element at the start of the poem. If a page break occurs within the poem, `<pb>` can be placed wherever is appropriate, although it should never be placed between `<lg>` and `<l>`.
- `<lb>` should always be placed within the next `<l>` element.

‘Is maith mo leaba, is olc mo shuain’, from NLS Adv. 72.1.36, would be framed as follows (Figure 38):

Figure 38

```
<text>
  <body>
    <div n="4" type="verse" resp="Hand9" corresp="MS12.4">
      <pb n="114r" xml:id="MS12.114r"/>
      <lg n="1" type="stanza" xml:id="MS12.4.1">
        <l n="a" xml:id="MS12.4.1a">
          <lb n="1" xml:id="MS12.114r.1"/>
        <!-- Metrical line continues here -->
      </lg>
    </div>
  </body>
</text>
```

</l>
<!-- Stanza continues here -->
</lg>
<!-- Text continues here -->
</div>
</body>
</text>

Divided Prose

Divided Prose is structured like verse, within `<lg>` elements (with `@type` set to “proseDiv”), except that the lines (`<l>`) are determined not by an identifiable metre but by self-contained units of content. These may or may not also be represented in terms of manuscript page layout.

The “Definitions” from NLS Adv. 72.1.4 would be framed as follows (Figure 39):

Figure 39

<text>
<body>
<div n="1" type="divProse" resp="Hand24" corresp="MS17.1">
<pb n="24r" xml:id="MS12.24r"/>
<lg n="1" type="proseDiv" xml:id="MS17.1.1">
<l xml:id="MS12.4.1a">
<lb n="1" xml:id="MS12.24r.1"/>
<!-- Text continues here (first definition (Latin)) -->
</l>
<l xml:id="MS12.4.1b">
<!-- Text continues here (first definition (Gaelic)) -->
</l>
</lg>
<lg n="2" type="proseDiv" xml:id="MS17.1.2">
<l n="a" xml:id="MS12.4.2a">
<!-- Text continues here (second definition (Latin)) -->
</l>
<l xml:id="MS12.4.2b">
<!-- Text continues here (second definition (Gaelic)) -->
</l>
</lg>
</div>
</body>
</text>

In the above text, line breaks continue to occur but, as they do not correspond with a new `<l>` element after the first, they are not represented here.

If the page was structured in columns (it is not), the column would indicated by placing the `<cb>` element after `<pb>`.

Precisely how many `<lg>` elements are used, containing how many `<l>` elements apiece, depends on the text and editorial decisions made in relation to it. In this case, each definition (`<lg>`) is given in Latin (`<l>`), then in Gaelic (`<l>`). The `<lg>` enclosing each version means that the defintion can be retrieved and presented in both languages, for the purposes of comparison. However, a set of dicta in Gaelic could potentially be grouped according to theme or manuscript page layout or suchlike.

Multiple Forms

Some texts contain material in some combinations of the three forms covered above. In such circumstances, some special arrangement of elements in both the text body and the text header is required to structure and present the transcription properly.

Arrangement of Body Elements

A `<div>` element can be placed as the child of another `<div>` element to represent a subdivision in the text (the stylesheet will represent the subdivision with a title and a line break in the edition). This could be used to structure a long and complex text that is in the same form throughout. However, it is **essential** to add a new, embedded `<div>` if the form changes.

For example, *Cath Fionntrágha*, in NLS Adv. 72.2.11, is a prose saga containing embedded verse would be structured as follows:

Figure 40

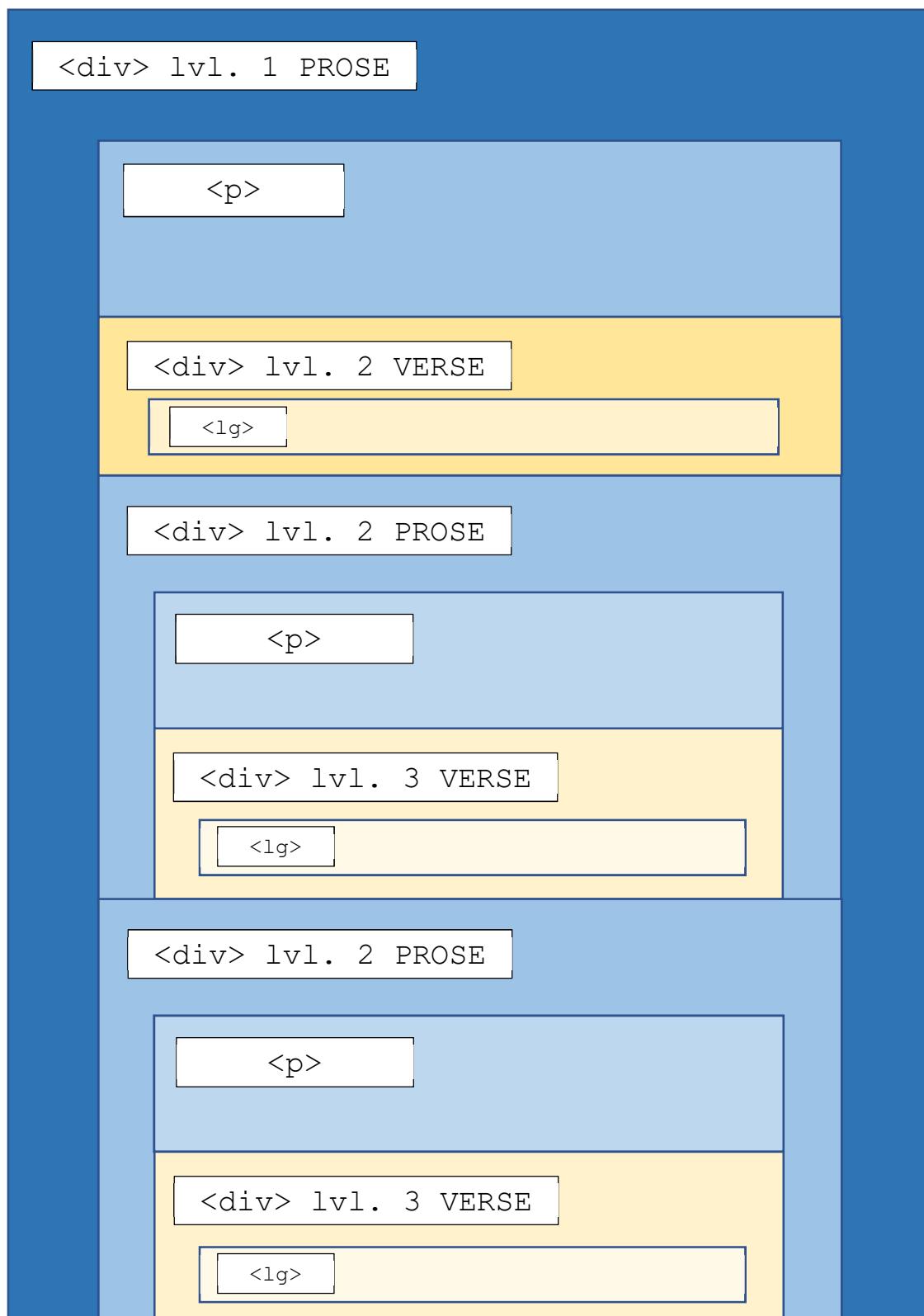
1	<code><text></code>
2	<code> <body></code>
3	<code> <div n="1" type="prose" resp="Hand23" corresp="MS3.1"></code>
4	<code> <p></code>
5	<code> <pb n="1" xml:id="MS3.p1"/></code>
6	<code> <lb n="1" xml:id="MS3.p1.1"/></code>
7	<code> <!-- Prose continues here --></code>
8	<code> </p></code>
9	<code> <div n="1.1" type="verse" resp="Hand23" corresp="MS3.1.1"></code>
10	<code> <lg n="1" type="stanza" xml:id="MS3.1.1.1"></code>
11	<code> <l n="a" xml:id="MS3.1.1.1a"></code>
12	<code> <!-- Metrical line continues here --></code>
13	<code> </l></code>
14	<code> <!-- Stanza continues here --></code>
15	<code> </lg></code>
16	<code> <!-- Verse text continues here --></code>
17	<code> </div></code>
18	<code> <div corresp="MS3.1" type="prose" resp="Hand23"></code>
19	<code> <p></code>
20	<code> <!-- Prose continues here --></code>
21	<code> </p></code>
22	<code> <div n="1.2" type="verse" resp="Hand23" corresp="MS3.1.2"></code>
23	<code> <!-- Verse text continues here --></code>
24	<code> </div></code>
25	<code> </div></code>
26	<code> <div corresp="MS3.1" type="prose" resp="Hand23"></code>
27	<code> <p></code>
28	<code> <!-- Prose continues here --></code>
29	<code> </p></code>
30	<code> <div n="1.3" type="verse" resp="Hand23" corresp="MS3.1.3"></code>
31	<code> <!-- Verse text continues here --></code>
32	<code> </div></code>
33	<code> </div></code>
34	<code> </div></code>
35	<code> </body></code>
36	<code></text></code>

The required structure of the multiple `<div>` elements is perhaps not intuitive. Assuming the text begins as prose, the initial prose `<div>` (line 3) cannot, within TEI conventions, be used to enclose all top-level (i.e. non-embedded) prose in the text. The initial prose `<div>` (line 3) has a `<p>` child (line 4) that encloses the prose until the first second-level `<div>` (line 9), before which the `<p>` child is closed (line 8). After the first second-level `<div>` (line 9), which contains a poem embedded in the text (e.g. ‘Triallim siar go hEirinn áin’, pp.1.25–2.3), another second-level `<div>` (line 18) is embedded with a `<p>` child (line 17), which contains the resumed prose. Its `@corresp` attribute keeps it associated with the initial top-level prose `<div>` (line 3).

Further verse appears; the next verse `<div>` (line 22) is a child of the second-level prose `<div>` (line 18). However, at the end of the verse, both `<div>` elements are closed and a new second level prose `<div>` (line 26) is added as a child of the top-level `<div>` (line 3). This can then contain the next verse `<div>`. In the tradition of *opera geminata*, the text then proceeds in pairs of prose and verse `<div>` elements, the latter a child of the former.

A simpler diagram may help (Figure 41):

Figure 41



There is nothing particular about prose and verse in this context. This model would apply whichever combination of forms were involved. That is, the first form, however little textual material is actually in it, becomes the overarching form for the text. While this might be troubling in theory, in practice

almost all stylesheet operations select material's closest `<div>` ancestor and thus capture the form at that particular point in the text.

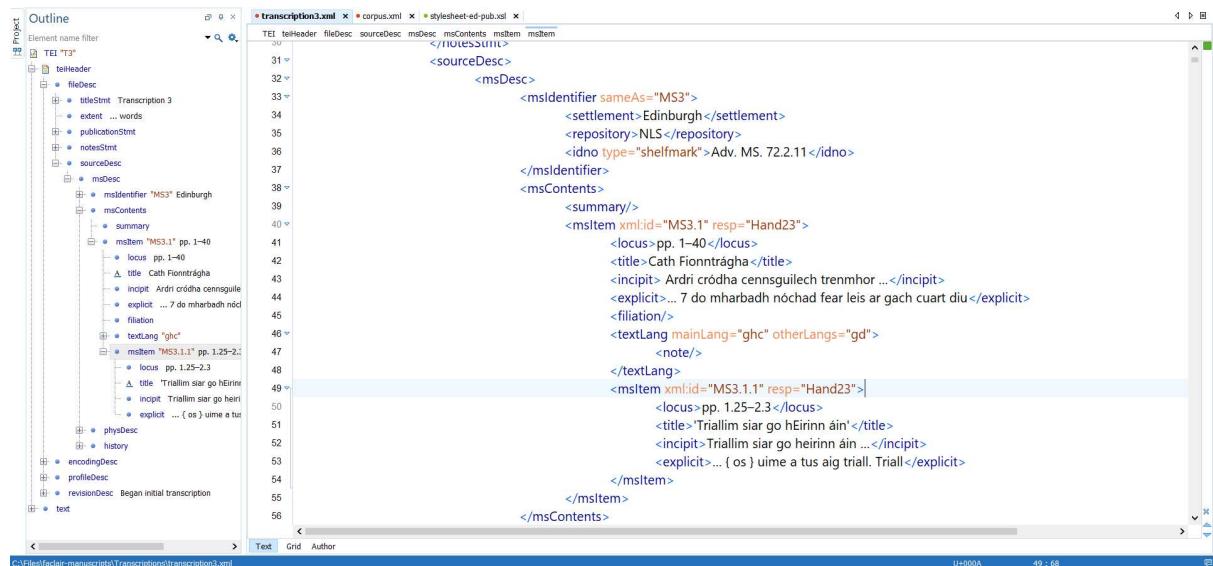
Arrangement of Header Elements

As already stated, each `<div>` child/descendant of the `<body>` element must correspond to an `<msItem>` descendant of `<teiHeader>`. Thus, for each sub-division of a text that you have represented using a `<div>` element **that represents either a deliberate sub-division or a change of form** (rather than a later continuation of the top-level form), you must add an `<msItem>` element plus descendants as a child of the `<msItem>` that represent the overall text. So, for each of the verse `<div>` elements in Figure 40, an `<msItem>` element would need to be added.

If a `<div>` is assigned a new `@n` and `@corresp` value, it should have its own `<msItem>` element in the header; if it reverts to a previous `@n` and `@corresp` value, its `<msItem>` element will already exist.

Figure 42 shows the first poem from *Cath Fionntrágha* represented as a child `<msItem>` element in the header (oXygen).

Figure 42



Revision Log

Once you have added the basic data and structure for the new text, go to the `<revisionDesc>` (Appendix 5a l.164) element in the transcription file header and add a `<change>` element (Appendix 5a l.165) noting the new text that has been added.

Review

- Three textual forms are currently recognised in the corpus: “prose”, “verse”, and “divProse”.
- Texts can consist of combinations of multiple forms.
- A text’s form is indicated by the @type attribute of its enclosing `<div>` element.
- The form determines the text’s overall TEI structure.
- A change of form is represented by embedding a `<div>` element within the text’s enclosing `<div>` element; this can be a complex operation.
- Embedded `<div>` elements within `<div>` must correspond to `<msItem>` elements in the transcription file header unless they stand for text at the top-level in the text (i.e. returning to the top-level `<div>`).

Chapter 3

Basic Transcription

“Thog iad na siùil bhreaca
Bhaidealachà dhònach,
Shìn iad na coilpeinean raga
Teanna rìghne
Ri fiodhannan àrda fada
Nan colg bìth-dhearg.

Cheangladh iad gu gramail snaompach,
Gu neo-clearbach
Troimh shùilean nan cromag iarainn
'S nan cruinn fhailbeag.

Cheartaich iad gach ball den acfhainn
Ealamh dòigheil,
'S shuidh gach fear gu freastal tapaidh
'Bhuill bu chòir dha."

- Alasdair mac Mhaighstir Alasdair, *Birlinn Chlann Raghnaill* (ed. D. Thomson).

Introduction

At long last, we are in a position to consider how to add and mark up actual text transcribed from manuscripts. This chapter covers the general strategy and principles underlying the mark-up scheme devised for the project, then sets out how to mark up straightforward, unproblematic text. Chapter 4, meanwhile, covers the representation in the mark-up of problems, uncertainty, and special manuscript features.

The scheme presented here is, of course, open to refinement and revision, and feedback on it from both a technical and a philological point of view is welcome. It should be borne in mind that the corpus already contains over 18,000 marked up words, so any changes could require time-consuming revision of the existing transcriptions, but this may, of course, be nonetheless worthwhile.

Prerequisites

This chapter assumes that you have already familiarised yourself with and acted the preceding sections of this user documentation. In particular, you will need to have set up a transcription file and updated the corpus file ready for the new text and, as per the instructions in CHAPTERS 1 and 2. You should also be on an appropriate *GitHub* branch.

You may or may not have already completed the transcription you plan on marking up. It is possible to transcribe directly into TEI and, for a transcriber conversant in the project's TEI conventions, this is certainly faster. However, while learning TEI, it may be adviseable to have a transcription prepared and ready, then mark it up. Note that you will still need access to the manuscript (unless your transcription's apparatus is extraordinarily detailed), as the TEI mark-up will require details not normally included wholesale even in a scholarly edition (e.g. which abbreviations are used; scribal word division; the reason why a reading is uncertain).

The following online resources should also be accessible (internet connection thus required),

- *ISOS* (<https://www.isos.dias.ie>), or an image of the manuscript from elsewhere, or the manuscript itself;
- *eDIL*: <http://www.dil.ie> ;
- *Dwelly* (via *Faclair Beag*): <http://www.faclair.com/> ;
- *Tionscadal na Nod*: http://www.vanhamel.nl/codecs/Project:Tionscadal_na_Nod ;
- *HDSG MSS Slips* [completion due Autumn 2018; DASG staff to provide details] ;
- The file “xmlid_lists.xlsx” from the project *Dropbox* folder;
- The *AHK* script containing the hotkey library running, if in use (advised);

plus any reference works for grammar and palaeography that you find helpful.

For an example of the coding described in this chapter in action, watch FnaG MSS- Basic Transcription Demo

General Principles

Before we look at the details of the mark up, however, it is worth considering the general editorial principles that animate the TEI strategy that has been devised for the project.

Focus on the Manuscript Readings

FnaG's lexicographers, when they use the MSS Corpus, will be interested not just in semantics and range of vocabulary and so forth. They will also be investigating matters such as orthography, and what this might reveal about pronunciation and intellectual influences, or deviations from standard morphology or syntax and the insights these might yield into dialect. Therefore, the default position when transcribing is to give the unamended manuscript form. Where this could be misleading, or when our interpretation of the form must needs be shown, there are facilities in TEI for making editorial interventions and emendations, but these must never be silent. It must always be possible for users of the transcriptions to access the manuscript form.

Focus on the Individual Manuscript

Our primary "informants", on this project, are the scribes of the manuscripts and the corpus offers transcriptions of their work, not editions of texts. Cross-references to other manuscript versions of the same text, where helpful, can be made via footnotes or in the introduction but each transcription file must contain material solely from a single manuscript.

Accountability

Lexicographers, when using the project's transcriptions, will be needing to assess the credibility of each piece of data they provide and make informed decisions about whether and how it can be used in the dictionary. For each form, therefore, the transcriptions must flag up any cause for uncertainty and provide details. These must be as precise as possible, with allowances made for the need to maintain pace when adding transcriptions and for the lexicographers' ability to conduct their own investigations.

In textual editing, editors are often conceived of as taking on decision-making responsibility and readers are expected to place a certain amount of trust in editors, at least pertaining to less controversial decisions. On the FnaG MSS Project, despite our policy of presenting manuscript readings by default, we are required to make interpretative decisions, when expanding abbreviations or marking up a reading with a dictionary headword, for example. Given the future use of the transcriptions we produce, we need to show the decision that was made and explain if and how it was difficult.

Logocentricity

The fundamental unit of the TEI mark-up on the project is the word (<w>). More precisely, a "word" is defined in this context as the smallest unit for which a dictionary headword would customarily be created. Compound words or verbal complexes, when the components are identifiable with existing dictionary headwords, are marked up as a single word but then split into their component words by embedding <w> elements within an enclosing <w>; the stylesheet, when generating the editorial text, will maintain their status as compounds or verbal complexes with appropriate word division.

There is currently no provision in place for marking up sentences, phrases, or tokens, although this may be applied in the future. The focus on indivisible lexical items is to allow queries of the resulting database — the corpus's transcriptions, by XML's very nature, are both editions and databases at the same time — to track usage of lexical items to the widest extent in the corpus.

One Transcription, Two Systems of Reference

In texts, or sub-divisions thereof, that are “verse” or “divided prose” in terms of form, it is important that two systems of reference be maintained in parallel. Every reading from the manuscript must be locatable both on the manuscript page and in the editorially-imposed structure (e.g. metrical lines). This enables both an editorial text and a diplomatic edition to be generated from the same TEI/XML mark-up. Details are given later in the chapter and were covered briefly in CHAPTER 2.

This does not apply to “prose” texts, as, there being no other editorial structure to apply to them, they are by default presented with the line division as it is on the manuscript page. There are other policy differences (e.g. word division) that justify the creation of editorial and diplomatic texts in these cases, of course.

Speed

Priority should obviously be given to producing transcriptions and commentary that are as careful, credible, and accountable as possible. This is, however, a time-limited project with ambitious objectives, so opportunities to achieve time efficiencies in TEI-related work have been taken where feasible. These include:

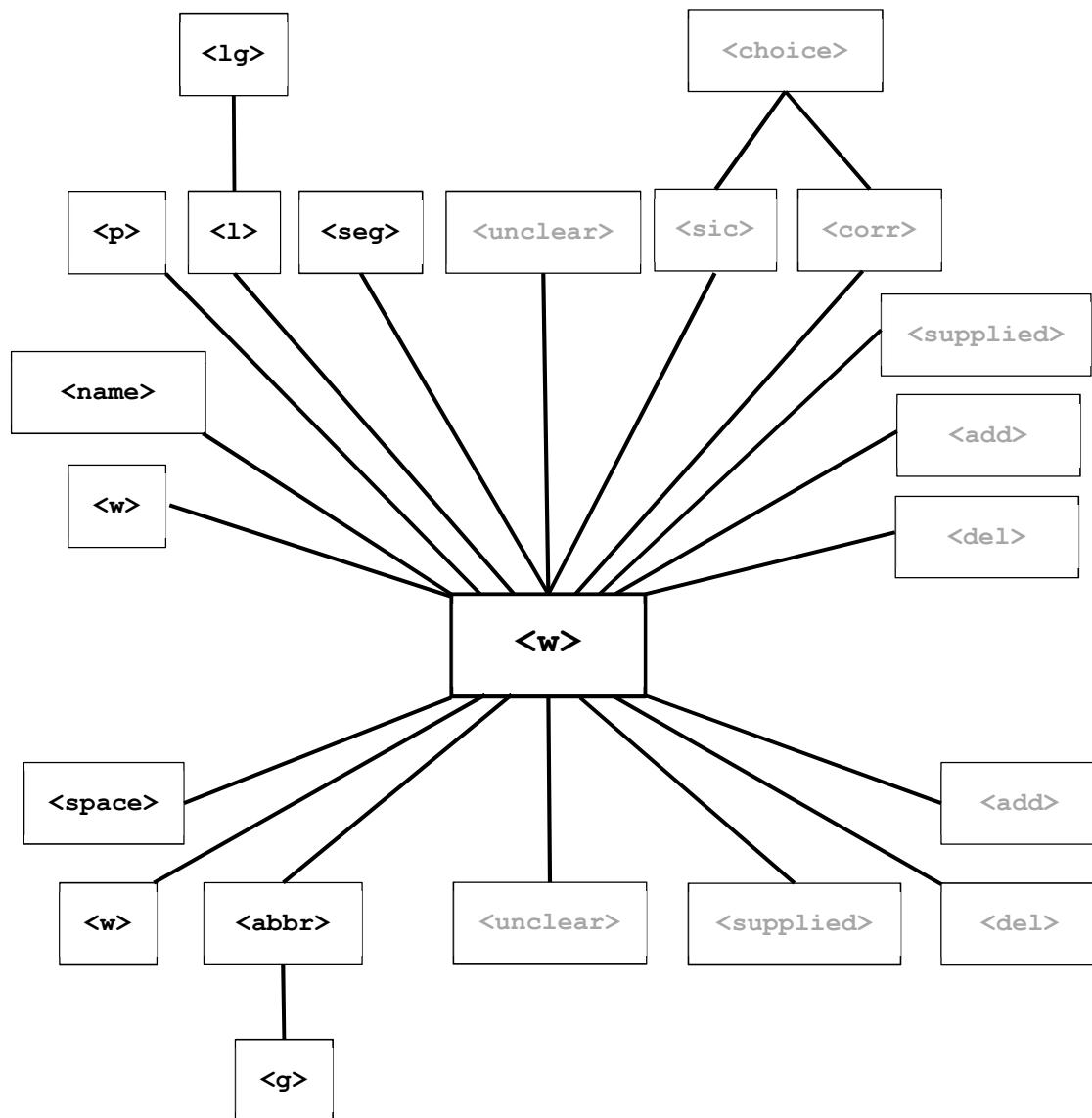
- The creation of a library of hotkeys using the *AHK* software package (see INTRODUCTION);
- The use of the headword database (*hwData.xml*) to centrally store lexical data so that only the minimum needs to be added during mark-up;
- The auto-generation of certain sections of the transcription file, such as the bibliography.

Transcription

The Word

As alluded to above, the representation of a word is at the core of the project’s “logocentric” TEI scheme. The potential parents and children of the `<w>` element are shown in Figure 43 (greyed out elements are covered in CHAPTER 4).

Figure 43



An individual word is marked up as follows,

```
<w lemma="focal" lemmaRef="http://www.dil.ie/22400"
ana="noun">facal</w>
```

The text content, “facal”, is the manuscript form. This should be an exact transcription from manuscript, except in the circumstances discussed in CHAPTER 4. Upper-case letters can be used to indicate capitalisation or display script (e.g. MS12.114^r.1; APPENDIX 8b I.7). For indication of spaces internal to a word, see below.

Three basic attributes are required for each word:

@lemma: the reading's dictionary headword, as identified by the transcriber;

@lemmaRef: a URL for the dictionary headword (@lemma) entry;

@ana: a simple POS (part-of-speech) tag (noun, verb etc.), according to the scheme set out in the "ana" sheet of xmlid_lists.xlsx (*Dropbox*, "Project" folder).

The dictionary headword/URL should be sought first in eDIL, then (if the reading cannot be identified with an entry in eDIL) in Dwelly (via *Faclair Beag*). Every effort should be made to use one of these resources, as various functions are set up to handle URL forms specific to them. If another resource proves consistently useful in this context, then the project team should be informed so that the resource can be considered for formal inclusion and the necessary technical upgrades undertaken.

In addition, to add a reference to this word in the HDSG-A MSS Slips Archive (when ready), use @lemmaSl to add the slip headword and @slipRef to add its URL in the Archive.

The @lemmaRef is very important for subsequent functions in the system. It acts as a unique identifier for the reading's headword, allowing all instances of a given headword throughout a text or file or throughout the corpus to be retrieved accurately, as well as stats to be generated on the corpus's overall contribution and richness. In another important function, it acts as a link to the record for the headword in the headword database (hwData.xml). This is drawn upon to provide a Dwelly headword/URL for headwords also in eDIL, as well as additional morphological information.

For these reasons, the @lemmaRef should be consistent in form (e.g. the "http://" prefix should always be included) and either eDIL or Dwelly (in that order of preference) should be used consistently for a given headword.

Special procedure for using Dwelly (via *Faclair Beag*)

Each Dwelly entry in the digitised version on www.faclair.com has a unique URL, which can be used to link back to that entry. For example, if the entry for *focal* is retrieved, the URL is:

<http://www.faclair.com/ViewDictionaryEntry.aspx?ID=53D7536354783855A07F93735CB692B3>

However, this URL is also unique to this search. A different URL is generated if the page is reloaded *focal* is searched for again:

<http://www.faclair.com/ViewDictionaryEntry.aspx?ID=16ABC9AEE56DE112DC8A4980D3660C82>

Both URLs still point to the entry for *focal*. However, as the @lemmaRef is meant to act as a unique ID for the headword for the purposes of searching the corpus, and not just link to the entry, having different URLs for the same headword is a problem.

It is therefore adviseable, when adding a Dwelly headword/URL, to check the corpus for previous occurrences of that headword. If found, the URL should be copied from those previous occurrences. Methods for searching the corpus are discussed in CHAPTER 7.

In time, it is envisaged that the headword database will contain a URL for every word in the corpus. It can then be searched for the designated Dwelly URL. However, at present, it only contains details for the most frequently occurring words so is unlikely to provide a solution.

While searches and queries can be performed on the corpus with @ana as a criterion, it does not play any role in identifying the headword and can thus be safely varied according to the usage in a

particular instance. That is, the @ana attribute should contain the POS for the manuscript form in context, not for the abstract dictionary headword. So, for example, an adjective would be tagged as a “noun” if it is interpreted as being used in this instance as a substantive even if the dictionary entry makes no reference to a substantive use for that adjective.

There will also be cases where the relationship with the dictionary headword is looser; these will be discussed in CHAPTER 4.

Within the Word

Abbreviations

In most pre-modern manuscripts, abbreviations, suspensions, and contractions are used to a greater or lesser extent. This was brought to a very high state of refinement in medieval Gaelic manuscript culture, meaning that an elaborate and often ingenious repository of abbreviations was potentially available to scribes. In order to provide the most accurate and accountable rendering of a manuscript form, we need to show in transcriptions not only where abbreviations have been expanded but what the abbreviation is and how sure we are that we have expanded it correctly.

As an example, we will consider the first two quatrains of ‘Is maith mo leba, is olc mo shúain’ (MS12.114^r.1–4; see APPENDIX 8). In the second word, “maith”, the lenition of the “t” is indicated by a *spiritus asper* (a “lenition tick”). This is marked up in our TEI follows:

```
<w lemma="maith" lemmaRef="http://www.dil.ie/31438"
    ana="adj">mai<abbr cert="high">t<g ref="g4">h</g></abbr></w>
```

Two elements, each with an attribute, are involved: <abbr> (@cert) and <g> (@ref).

<abbr> encloses every character involved in the abbreviation (“th”, in this instance). The **direct** text content of <abbr> (“t”, in this instance) consists of the character(s) written out on the manuscript page and bearing whatever indicates the abbreviation marker (the *spiritus asper*, in this instance).

@cert: this attribute’s value is the level of certainty with which you have expanded the abbreviation; the options are “high”, “medium”, “low”, and “unknown”.

<g> (“glyph”) encloses the characters that are not directly visible on the manuscript page in their usual form but are expanded out of the abbreviation (“h”, in this instance).

@ref: this attribute’s value is an id number corresponding to an @xml:id in a database of abbreviations stored in corpus.xml (see APPENDIX 6a l.148). This database contains details such as the abbreviation’s name, a link to its entry in *Tionscadal na Nod* (if available), and a short description. For a quick reference list, see the “Glyphs” sheet in xmlid_lists.xlsx (*Dropbox*, “Project” folder). “g4”, for example, is the id for *spiritus asper*.

Some abbreviations require complete transformation when expanded; none of the characters involved in the abbreviation remain in the expanded form. In such instances, <abbr> has no direct text content and the expansion is simply placed within <g>. For example, in MS12.114^r.2, the preposition “ar” is written as a “q” with a stroke through the descender. This is rendered in our TEI as follows:

```
<w lemma="ar" lemmaRef="http://www.dil.ie/3902"
    ana="prep"><abbr cert="high"><g ref="g8">ar</g></abbr></w>
```

Each character bearing one or more abbreviations should have its own `<abbr>` element. While it might be tempting, when faced with a string of abbreviations, to enclose them all the same `<abbr>` element, later functions and queries will need to be able to identify the certainty level of a specific expansion and queries could potentially be composed to retrieve abbreviations rather than words.

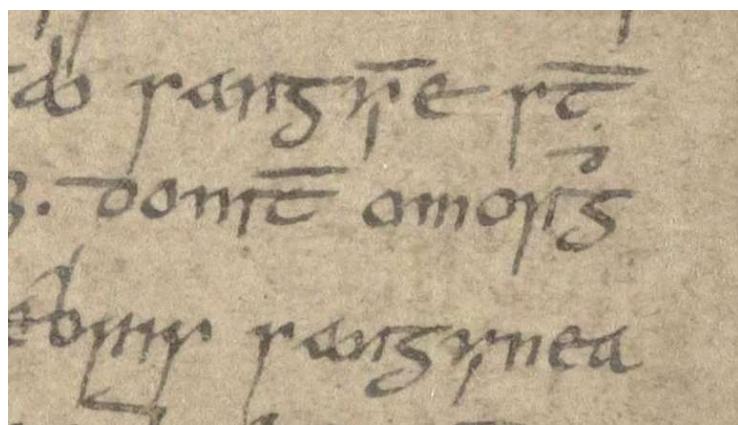
Thus, for example, “chaird” (MS12.114r.3) is rendered as follows in our TEI:

```
<w lemma="càird" lemmaRef="http://www.faclair.com/ViewDictionaryEntry.aspx?ID=007AB404DEE  
A6B0878BD3BEECF89ABF2" ana="noun"><abbr cert="high">c<g  
ref="g4">h</g></abbr><abbr cert="high"><g  
ref="g11">air</g></abbr>d</w>
```

If the same letter carries multiple abbreviations, however, the same `<abbr>` element should be used (it being not possible, or at least not straightforward, to subdivide a letter).

For an example of this, we must leave NLS Adv. 72.1.36 and turn to a form in NLS Adv. 72.1.33 (MS5), p.37.24 (Figure 44): “morgudh”.

Figure 44



The ending, “-udh”, is indicated by a superscript d (standing for “ud”), which itself carries a *spiritus asper* (“h”). This is rendered in our TEI as follows, using only one `<abbr>` element.

```
<w lemma="morgad" lemmaRef="http://www.dil.ie/32573"  
ana="vnoun">mor<abbr cert="medium">g<g ref="g24">ud</g><g  
ref="g4">h</g></abbr></w>
```

If you are unable to match the abbreviation you are looking at against any of the abbreviations listed in `corpus.xml/xmlid_lists.xlsx`, you will need to add a new record of it and create a new id. See APPENDIX 6b I.503 for details.

`<abbr>` elements cannot span `<w>` boundaries in TEI, so the element must be closed before the `<w>` element is closed. Occasionally, scribes will use an abbreviation to give the last character(s) of one word and the initial character(s) of the next word. In these cases, two separate `<abbr>` elements will need to be created in each `<w>` to represent the abbreviations contribution to each word. This means any stats involving counts of abbreviations will be slightly inaccurate.

For the handling of lower certainty abbreviations and representation of other problems with reading, see CHAPTER 4.

Word-Internal Spaces

Scribes will sometimes place deliberate spaces within what we would analyse as a single, indivisible word. A `<space>` element can be placed inside a `<w>` element to represent this. It will be ignored in the editorial text, so the word will be presented there as one word without spacing. The word will be represented with a space in the diplomatic text.

For an example, see “ceadna” / “cead na” at NLS Adv. 72.1.4 fol. 24^r7 (APPENDIX 9b ll.117–119 and the different reanderings in APPENDIX 9c).

For more details on `<space>`, see below.

Words in Languages Other than Gaelic

Words in languages other than Gaelic (of whatever period or register) should still be marked up in `<w>` elements. However, the only attribute required is `@xml:lang`, the value being the standard XML tag for the language. See Appendix 9 for examples of Latin text.

The boundary between a loanword in Gaelic from another language and the rendering of a word from another language in orthography influenced by Gaelic can be very difficult to draw. Since the policy of FnaG is to be inclusive of the diverse different forms of Gaelic that have existed historically in Scotland, the weighting should perhaps be towards Gaelic in such situations.

The tags currently in use are:

- la (Latin);
- sco (Scots);
- hbo (Ancient Hebrew);
- und (unknown).

For a full list of tags that can be used, see <https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>

Around the Word

Word Division

Obviously, one encounters spaces more often between words. This project’s definition of a space is relative. That is, the span between letters does not need to pass any particular measurement to qualify as a `<space>`. Instead, the span between letters is marked up with `<space>` if it is noticeably larger than other spaces between other letters. The `<space>` element is omitted if the scribe could have left more of a space than they did, had they wanted to. Thus, if the scribe is struggling to fit his text into a line and generally has to leave less of a span between words, then a smaller span will qualify as a deliberate space and thus for mark-up as `<space>`.

The treatment of each `<space>` element is determined by its `@type`:

“scribal”: This represents a space in the scribe’s text, according to the above criteria. This will be the most commonly used `<space>` element in most transcriptions;

“editorial”: This defines a space as inserted by a later editor; it is only to be used if the manuscript itself is illegible and an older transcription/edition is being relied upon;

“force”: This will insert a space in the editorial text but **not** in the diplomatic text. This can be useful for representing page layout and indentation;

“em”: This inserts a long space, 3× the length of the standard space. This can be useful for representing page layout and indentation. It will result in a space being inserted in both editorial and diplomatic texts.

It should be noted that word division should continue to be represented at the end of editorial divisions (e.g. at the end of metrical lines), should these (as is often the case) not coincide with manuscript lines. This is so that word division can continue to be represented in the diplomatic text.

See the mark-up around the <1> elements in Appendices 8b and 9b and the results in Appendices 8c and 9c for examples.

Compound Words

To mark up a group of <w> elements as belonging to a compound word, enclose them all in a single outer <w> element. For example, the name “Trénmór” (“threnmhoir”; NLS Adv. 72.1.36, fol.96r3; APPENDIX 7b II.83–88) is represented in our TEI as follows:

```
<w ana="adj, adj">
  <w lemma="tré'n" lemmaRef="http://www.dil.ie/41798"
    ana="adj"><abbr cert="high">t<g ref="g20">re</g></abbr>n</w>
  <w lemma="mó'r" lemmaRef="http://www.dil.ie/32548"
    ana="adj"><abbr cert="high">m<g ref="g4">h</g></abbr>oir</w>
</w>
```

For additional special rules regarding the mark-up of names, see below.

The outer <w> element need only have an @ana attribute, with its value being a comma-separated list of the POSs of the words in the compound. The processing software, with reference to the stylesheet, will use the combination of POSs to determine whether and where word division is to be applied. In most cases, at present, it will run compounds together (compare the treatment of “threnmhoir” in the editorial and diplomatic texts in APPENDIX 7c). This also means that a list of compound words in the corpus, with data on their components, could potentially be generated by a query. The component words otherwise retain their identity as separate words.

<space> elements can still be placed between components of compound words as required (e.g. NLS Adv. 72.1.36, fol.96r2; APPENDIX 7b I.50) and the rules described above will be applied.

Currently, constructions defined as “compounds” for these purposes include:

- Nouns preceded by their adjective outside of a copulative construction;
- Combinations of nouns and adjectives run together by the scribe or seemingly meant to be joined together based on, for example, the lenition of the initial of a component after the first;
- Verbs, nouns, and pronouns with a following emphatic or pronominal;
- Verbal complexes, including particles, infix pronouns, suffixed pronouns, emphatics, and pronominals, as well as the verb itself.

It should be noted that the primary relevance of compounds in the context of this project’s TEI is word division in the editorial text, with the secondary relevance being to querying scribes’ concepts of word boundaries. Compounds should thus be formed with editorial conventions in mind.

Occasionally, a word will have more than one component but this can only be indicated by the value of the @ana attribute, as the word cannot otherwise be split into component words that make sense in themselves, there being too many mutations involved in their conjunction. This policy has been adopted for conjugated prepositions (e.g. APPENDIX 7a fol. 96^r8; APPENDIX 7b l. 246). While some examples, like “fris” could be easily split, in the mark up, into preposition and pronoun/possessive (“fri” + “s”), the results would be misleading in many cases. For example, attempting to mark up “agum” or “leis” would lead to either the preposition “agu”/“lei” or the conjugation “um”/“is” being returned in searches, neither of which would be accurate. Therefore, in such cases, the parts-of-speech involved in the form should simply be put in a comma separated list as the value of @ana, with no enclosed <w> elements required. The stylesheets have been adapted to parse the @ana value and identify affected forms as either a preposition or as a pronoun/possessive, as applicable.

Names

Names be enclosed in <name> elements and then rendered with <w> elements as described above. For example, our previous example of “Trénmór” (“threnmhoir”; NLS Adv. 72.1.36, fol.96^r3; APPENDIX 7b ll.82–89) is represented fully in our TEI as follows:

```
<name type="personal">
  <w ana="adj, adj">
    <w lemma="trénn" lemmaRef="http://www.dil.ie/41798"
      ana="adj"><abbr cert="high">t<g
      ref="g20">re</g></abbr>n</w>
    <w lemma="mórr" lemmaRef="http://www.dil.ie/32548"
      ana="adj"><abbr cert="high">m<g
      ref="g4">h</g></abbr>oir</w>
  </w>
</name>
```

For the simpler example of Fionn, see NLS Adv. 72.1.36, fol.96^r3; APPENDIX 7b ll.66–69.

Names are differentiated by @type:

- “personal”: human individuals;
- “place”: locations;
- “population”: peoples and communities;
- “other”: miscellaneous.

A <name> element can be embedded within another <name> element, if needed. For example, in the *Book of Deer* (MS1.4r.1), we find the placename “Pett Ma[e]l Dúib”. This is a placename, but it contains a personal name, which in turn contains another personal name. This would be rendered in our TEI as follows:

```
<name type="place">
  <w lemma="pet(t)" lemmaRef="http://www.dil.ie/34300"
    ana="noun">pett</w>
  <name type="personal">
    <w lemma="Mael" lemmaRef="http://www.dil.ie/31244"
      ana="noun">mal</w>
    <name type="personal">
      <w lemma="dub" lemmaRef="http://www.dil.ie/18985"
        ana="noun">dúib</w>
```

```
</name>
</name>
</name>
```

Names should be marked up as such as a matter of policy. They offer a qualitatively different sort of lexicographical evidence, such as by fossilising forms that are no longer productive and it is thus important to be able to easily distinguish them in queries and searches.

Other Elements

Non-Word Elements

Elements other than words will often be needed to represent a scribe's rendering of the text fully.

Punctuation

One such case is punctuation. For example, at NLS 72.1.4 fol. 24^r9 (APPENDIX 9b I.120), the scribe appears to use a comma. Punctuation is represented by its nearest equivalent in UTF-8, enclosed in a <pc> element. <pc> elements can also be placed within a <w> element, if the punctuation appears as part of the rendering of a word (e.g. NLS 72.1.4 fol. 24^r2; APPENDIX 9b I.22–23).

If a punctus is used as part of an abbreviation, it is not represented as a separate <pc> element. Sometimes, scribes add points around other abbreviations, perhaps to flag up the proximity of an abbreviation (e.g. NLS Adv. 72.1.4 fol. 24^r7; APPENDIX 9b II.117–119). If this appears to be the sole purpose of a piece of punctuation, it is also not represented, according to the project's current policy, but is considered an innocuous part of the abbreviation.

Punctuation used to separate metrical lines (e.g. throughout the text in Appendix 8) should be represented, as this is part of the evidence used for analysing the poem's metre and thus confirming certain readings.

The <pc> element is also used to render simple decorations or line-fillers (e.g. NLS Adv. 72.1.4 fol. 24^r10; APPENDIX 9b I.147).

Numbers and Dates

Numbers can be marked up as words, even when rendered in Roman and Arabic numerals by the scribe (e.g. NLS Adv. 72.1.36 fol. 96^r8; APPENDIX 8b I.238), as long as the number is equivalent to a single word (i.e. 1–10 or divisible by 10 up to a certain point). If not, the <num> element can be used.

For year dates in the form “1613”, the <date> element should be used. More complex dates should be marked up using <num> and <w> elements, as needed.

New Lines and Other Divisions

The elements used to indicate new pages (<pb>), new columns (<cb>), and new lines (<lb>) have already been noted, and examples of <pb> and <lb> in action can be found in APPENDICES 7–9 (<cb> functions in exactly the same way). These elements can be placed almost anywhere within the mark-up (they cannot be placed within <g> but this should never be needed), except their order in relation to each other should be logical. So the divisions on the manuscript page can always be rendered accurately, even if they split <w> elements.

Annotation

While as many opportunities as is practical have been created to encode formal data within the mark-up, there will inevitably be complex or noteworthy situations requiring special comment, explanation, or cross-reference. There are currently three ways of adding this.

Element Comments

Any element in the mark-up can have a @comment attribute (which has been specially customised for this project), the value being the transcriber's comment on that element. However, the only element set up to display this sort of comment in the output is <w> (the comment displays in the tooltip on mouseover in the HTML). If being able to make visible comments on other sorts of elements seems to be very valuable, this should be raised with the project team and the necessary upgrades can be made.

Footnotes

To add a footnote within the transcription, the following template can be placed almost anywhere in the transcription <body>:

```
<note type="fn" resp="NNN">
    <p><!-- Text of footnote here --></p>
</note>
```

A number will be automatically assigned to the footnote by the processing software and the text of the footnotes placed at the end of each <div>. A reference number hyperlinked to the footnote will be placed at the location of the <note type="fn"> element.

The @resp attribute should be the id of the transcriber adding the note or responsible for its content.

For the mark-up used in commentary and academic prose, see CHAPTER 5.

'altText' Notes

If there is another version of the text being transcribed that could be of interest as the subject of close comparison, this can also be placed in a <note> element as a sibling following the element containing the section of transcription to which the comparison is being made:

```
<note type="altText" resp="NNN">
    <p><!-- Text here --></p>
</note>
```

This text does not need to be marked up in the same way as transcriptions and can be simply typed in within the <p> element. Indeed, the transcription mark-up will not work in this context. Instead, the mark-up designed for commentary should be used here (see CHAPTER 5).

The @resp attribute should be the id of the transcriber adding the note or responsible for its content.

Organisation of Code

As exemplified in APPENDICES 7b, 8b, and 9b, indentation can and should be used to represent the hierarchy of XML elements within the TEI mark-up. If the place in the hierarchy is lost, *oXygen XML Editor* has a function that automatically adds indentation accurately (Ctrl + Shift + P, by default). A new line should be started for each element **above <w> in the hierarchy** by hitting Return. No new lines should be added in this way within <w> or any of its descendants. *oXygen* will interpret this as adding a space within the text content itself and the processing software will then overwrite any rules regarding word division for that <w> element.

For reasons of space, the contents of <w> elements in APPENDICES 7b, 8b, and 9b sometimes straddle line breaks in this document and automatically indenting code in *oXygen* will often have the same effect. However, while producing the same visual effect, this is **not** the equivalent of pressing Return and will cause no problems.

Review

- The project's mark-up strategy emphasises accountability, access to primary evidence, and individual lexical items.
- A range of resources are required in order to add the required data during mark-up.
- The focus of activity and of the element hierarchy is on the word (<w>).
- Only a minority of the information that can be made available about each word is added within the <w> element; much is drawn from elsewhere in the software by the processing software.
- Facilities are available that allow comment to be made on specific points in the transcription.
- Care is required when organising marked up text, even when simply considering visual clarity at the XML level.

Chapter 4

Advanced Transcription

“Do-dechad-sa,’ ol ind óclach, ‘a tírib ingnad, a tírib gnáth, co fesur uait-siu fót forsa mbeba ocus fót for a ngénir fis 7 anfis.””

- From *Immacaldam Choluim Chille 7 ind Óclaig oc Carraig Eolaig* (ed. Carey).

Introduction

This chapter sets out how to handle (in terms of TEI mark-up) more complex situations that can occur on the manuscript page and how to flag up unknowns and points of uncertainty. These fall into three categories:

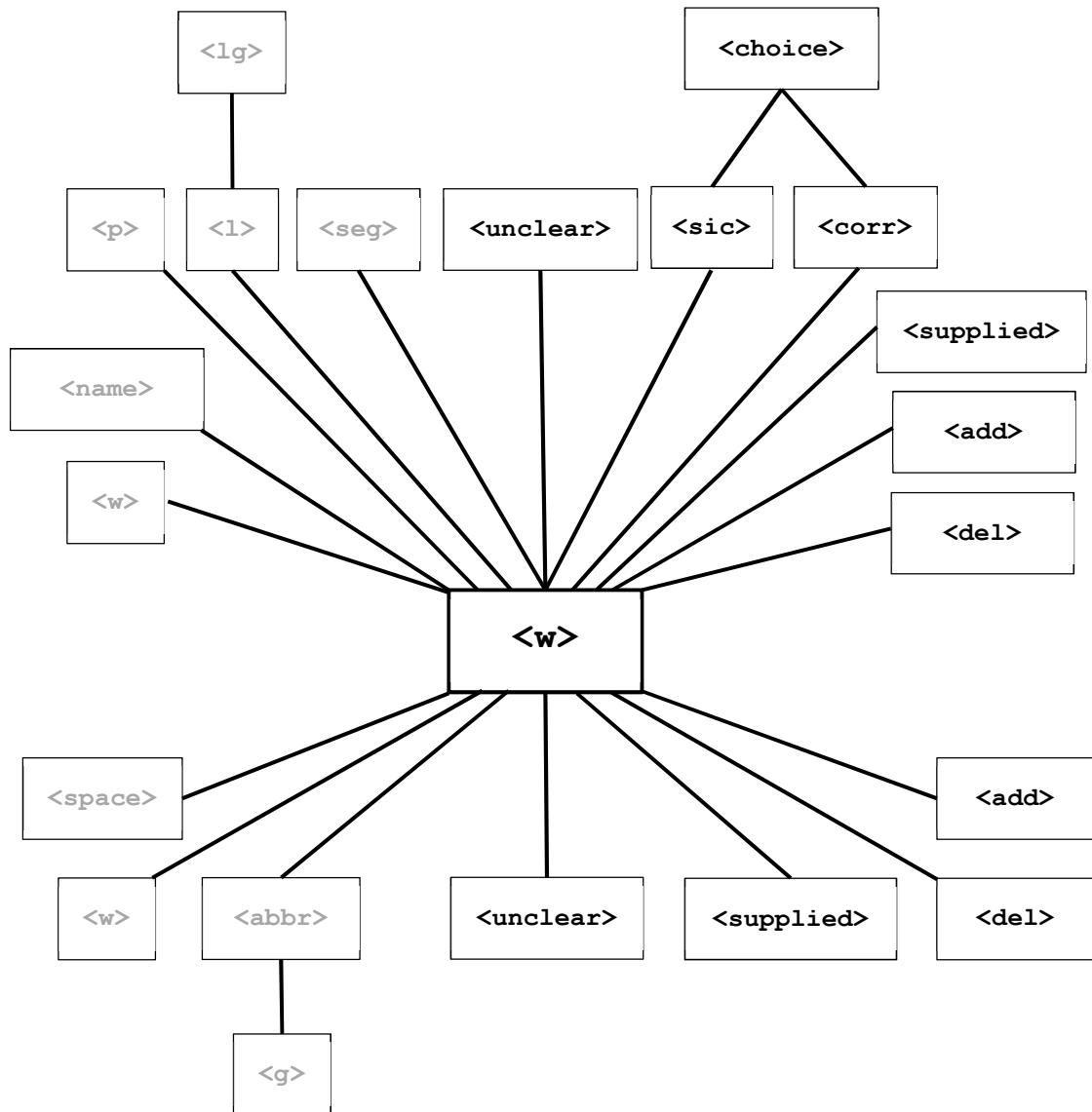
- Special text segments;
- Later scribal interventions;
- Unclear text;
- Editorial emendation.

While this can all be problematic hermeneutically, the general principles and techniques resemble those set out in CHAPTER 3 and there is little involved here that is hugely more difficult conceptually.

Elements Involved

As in CHAPTER 3, the elements involved in marking up more problematic sections cluster around <w>.

Figure 45



Special Text Segments

Sections of text playing particular functions should be marked up as such, to ensure appropriate layout and formatting is applied in the editorial text.

Titles

If a text is provided with a title in the manuscript, then the text of the title should be marked up normally and enclosed in the element `<seg type="title">`. It is envisaged that this will be used only for strings of text set out in some way as a title by the scribe. For an example, see APPENDIX 7b II.6–20. Titles will be presented in bold in the output. ×

Catchwords

Sometimes, what will be the first word of a new folio side or page will be written at the foot of the preceding folio side or page, as a kind of relative foliation. This is a “catchword”. It should be placed marked up as normal then enclosed in `<seg type="catchword">`, which will allow it to be differentiated visually and in searched from the continuous run of the text. If it appears (as it probably will) on a new line, then this should be indicated with an `<lb>` element.

Dúinte

A *dúnad* is the echo of the first line of a poem in the last line, to signal that the poem is ended and complete. Occasionally, *dúinte* can appear within poems to signal a change in topic or at points where the poem originally ended before further material was added.

Scribes will often highlight a *dúnad* by writing it out again, sometimes on a new line. If this has been done, then the repetition of the *dúnad* (**not** the original *dúnad*) should be marked up as normal (c.f. Figure 38) within an `<lg>` element differentiated using the `@type` and then an `<l>`:

Figure 46

```
<div n="1" type="verse" resp="Hand9" corresp="MS12.1">
    <!-- PRECEDING QUATRAINS OF POEM HERE -->
    <lg n="6" type="stanza" xml:id="MS12.1.6">
        <!-- QUATRAIN LINES OF POEM HERE -->
    </lg>
    <lg type="dúnad" xml:id="MS12.1.7">
        <l xml:id="MS12.1.7a"><lb n="14" xml:id="MS12.85v.14">
            <!-- TEXT OF DÚNAD -->
        </l>
    </lg>
</div>
```

Note that the `<lg>` and `<l>` elements used for the *dúnad* still require unique ids as the values of their `@xml:id` attributes, although they should not have numbers (`@n`). A suitable system should be adopted for ensuring that the *dúnad* elements’ ids are unique. In Figure 46, the *dúnad* concludes the poem so its ids are formed from the next number in sequence from the number of the last stanza (as if it were a stanza itself). If the *dúnad* had appeared mid-poem, some alternative suffix would have to be found (e.g. “.dun1”) to form an id that would not recur again in the corpus.

As with catchwords, the *dúnad* will likely be repeated on a new line. If so, this should be indicated using `<lb>`, which should have both an `@n` and an `@xml:id` attribute.

Cinn fo Eitib

To save space, scribes will sometimes place text in gaps at the rightward ends of preceding (or occasionally following) lines. This is usually separated from the line proper by a symbol resembling “//” or suchlike (for which the term is *cenn fo eiti* (“head under wing”)), although it can also be done silently, in which case it is detectable only by considering the text’s sense or by comparison with another manuscript witness.

This is marked up in three steps:

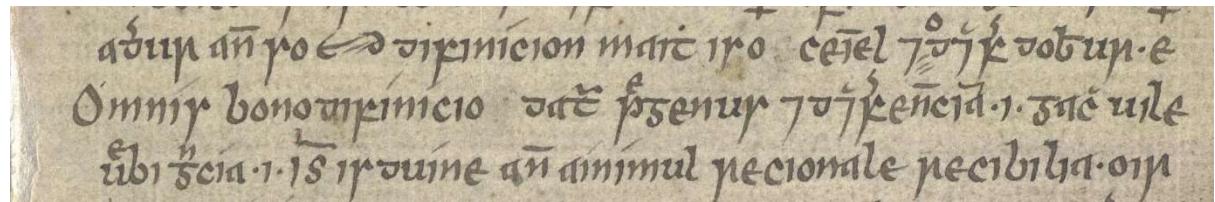
1. An `<lb>` element is inserted at the end of the line where the text ought to be read; instead of an `@xml:id`, it has a `@sameAs` attribute, with the value being the `id (@xml:id)` of the `<lb>` element marking the line where the text was placed by the scribe;
2. The text moved by the scribe is marked up as normal after the `<lb sameAs>` element, with the next true manuscript line begun as normal with an `<lb>` element.
3. Any *cinn fo eitib* included by the scribe should be enclosed in a `<pc>` element, then this `<pc>` element should be enclosed in `<seg type="cfe">`.

For an example, we must turn to NLS Adv. 72.1.2 fol.35r10–12 (Figure 48). The mark-up should thus resemble the following (Figure 47):

Figure 47

<code><div n="3" type="prose" resp="Hand20" corresp="MS14.3"></code>
<code> <!-- PRECEDING LINES OF TEXT --></code>
<code> <lb n="10" xml:id="MS14.35r.10"/><!-- MS FOL. 35r10 TEXT HERE --></code>
<code> <lb n="11" xml:id="MS14.35r.11"/></code>
<code> <!-- MS FOL. 35r11 TEXT HERE --></code>
<code> <lb sameAs="MS14.35r.10"/></code>
<code> <seg type="cfe"></code>
<code> <pc>//</pc></code>
<code> </seg></code>
<code> <!-- TEXT RELOCATED FROM MS FOL. 35r10 HERE --></code>
<code> <lb n="12" xml:id="MS14.35r.12"/></code>
<code> <!-- MS FOL. 35r12 TEXT HERE --></code>
<code></div></code>

Figure 48: *Cenn Fo Eiti* Highlighted



Programming work is still in progress to stabilise the output in such situations. The objective is that the relocated text appear:

- Where it is meant to be read and without any separating symbols in the editorial text;
- Where the scribe located it and with any separating symbols in the diplomatic text.

At present, the relocated text will always display where it is meant to be read, although with an accurate line reference (Figure 49).

Figure 49

10. aderur ann so //
11. Omnis bono difinicio datur per genus et dethferenniam .i. gach uile
10. **difinicion** is o e-ceinel agus \ o / deithfir do berur . e
12. uerbi gracia .i. is ed is duine ann ainimul recionale recibilia . oir

Later Scribal Interventions

Manuscript text can be altered, either by the hand that originally inscribed it or by a later hand, and the alteration can consist of deleting text, adding text, or both. We want to capture as much about scribe's behaviour as possible and mistakes and re-editing can yield particular insights into attitudes to language and style. It is therefore important that sections of manuscript text in which there have been later interventions be carefully marked up so that the original (where recoverable), each intervention, and the final outcome are all made available.

Deleted Text

Sometimes, deleted text is neatly scored out or, in a medieval convention, dots are placed beneath the characters or words that are supposed to be removed (subpunction). In other cases, text is erased or scored out completely, so that what was there originally is irrecoverable.

Identifiable Agent

An example of the former is found at NLS Adv. 72.1.36, fol. 96^r1 (APPENDIX 7b II.22–35). Here, the scribe, Eoghan Mac Gilleoin, wrote out the title of the next tale in the manuscript, *Bruighion Cheisi Choruinn* (see fols 104^v–111^r), not the correct title, *Bruighion Bheag na hAlmhain*. He has then subpuncted the erroneous title and added the correct title to the left.

The `` element encloses the text deleted by the scribe, which is otherwise marked up as normal. This element has a `@resp` attribute, with the value being the id for the hand responsible for the deletion (in this case, “Hand9”). The same ink was used in the subpunction as in the subpunction and in the inscription of the new title, which is also in the same hand. Hence we can identify the hand responsible for the deletion.

The text marked up with ``, in the output, will be struck through (see APPENDIX 7c) but otherwise retrievable through all queries and searches and accessible to the various reading aids built into the digital edition. The tooltip generated by placing the mouse over any of the deleted words will display a notification that the word was deleted and the hand responsible.

Unidentifiable Agent

In cases where text is simply scraped away or a different ink is used to score it out, we will find it much more difficult to identify the hand responsible. If a hand has added new text in place of what was deleted or has somehow taken account of the deletion, we could posit that that hand was responsible. Other evidence relating to the manuscript might supply names of individuals known to have intervened in it who might thus be candidates. However, if no clear identification emerges, the `` element should be used above, with the `@resp` attribute given the value “Hand999”. This indicates an unidentified hand.

Added Text

Added text is split into two categories:

1. Insertions, which are to be read within the main text;
2. Glosses, which expand upon the main text but do not directly fit within it.

The two categories are handled in a very similar way, using the `<add>` element.

Insertions

Insertions are enclosed in an `<add>` element, which has a set of attributes:

- @type: the sort of addition (in this case, “insertion”);
- @place: where the scribe has placed the text to be inserted, from the following options:
 - “above”;
 - “below”;
 - “margin, left”;
 - “margin, right”;
 - “margin, top”;
 - “margin, bottom”;
 - “inline”.
- @resp: the id for the hand responsible for the insertion. This attribute is required even if the same hand is also responsible for the main text.

The intended location of the insertion, which is where it should be marked up within the `<add>` element, may be obvious where the scribe placed it on the page. Alternatively, the scribe may have added callmarks to place it. If its location is uncertain, the material should be included as close as possible to its physical location, outside any elements below `<p>` and `<lg>` level, and the issue explained in a footnote.

Within the `<add>` element, the inserted text should be marked up as normal. In the output, it will be enclosed in punctuation marks that indicate the location of the text for insertion on the manuscript:

- above: “\.../”;
- below: “/...\”;
- margin, left: “>...>”;
- margin, right: “<...<”;
- margin, top: “\\...//”;
- margin, bottom: “//...\\”;
- inline: “|...|”.

The tooltip will display a message indicating the hand responsible for the addition and this information will also be retrieved if an e-slip is generated for the word.

For an example, see APPENDIX 7b II.6–20.

If only characters within a word have been added, then the `<add>` element can be placed within `<w>` and the added characters can be its direct text content. This will not change which hand is recorded as responsible for the word but the addition and another hand (if applicable) will be noted in tooltip note.

Glosses

Glosses are marked up in almost exactly the same way, except that the value of @type is “gloss”. In output, the gloss will likewise have its location on the manuscript page indicate but it will also be preceded by the abbreviation “gl.”, in bold.

If the gloss clearly relates to a particular word or short phrase, then both the <add> element containing the gloss and the word/phrase to which it relates can be placed within <seg type="gloss">. This will mean that lemma and gloss can be retrieved and related to each other by queries and that tooltip messages will highlight their relationship. If the gloss cannot be attached to a very specific point in the text, however, it can simply be placed in a <add type="gloss"> element at an appropriate location.

As there are no examples of glosses in the corpus at the time of writing, the following fictional example has been created. It has been imagined that another scribe of NLS Adv. 72.1.4 (APPENDIX 9) has glossed “sine” with “.i. gan” at fol. 24r1.

Figure 50

```
<seg type="gloss">
    <w xml:lang="la">sine</w>
    <add type="gloss" resp="Hand25" place="above">
        <w lemma=".i." lemmaRef="http://www.dil.ie/26892"
            ana="abbr">.i.</w>
        <space type="scribal"/>
        <w lemma="cen" lemmaRef="http://www.dil.ie/8581"
            ana="prep">gan</w>
    </add>
</seg>
```

This produces the following output, in the editorial text:

Figure 51

1.
1. { re sine \gl. .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis } 3. gan resun { is } -gap- 4. { ni is { cuis } } -gap-

The tooltip messages clarify the text and object of the gloss and the hand responsible:

Figure 52

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.

1. { re sine \gl .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis }

Add Comment

Language: Latin:
unk. unk. (24)
- this word is difficult to decipher
- this word is in a language other than Gaelic

Abbreviations: none

[Back to MS contents](#)

A gloss has been added by Maol-Sheachlainn mac lollainn Mhic an Leatha Ruaidh (Hand25): ".i. gan"

ap-

Figure 53

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.

1. { re sine \gl .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis } 3. g

Add Comment

cen: prep
Maol-Sheachlainn mac lollainn Mhic an Leatha Ruaidh (25)
- this word is difficult to decipher

Abbreviations: none

This has been added as a gloss on "sine" by Maol-Sheachlainn mac lollainn Mhic an Leatha Ruaidh (Hand25).

Change of Hand

The hand rendering the main text might change. That is, a hand other the one which began the text may take over for a greater or lesser period. If this occurs, the change in hand is indicated with a closed `<handShift/>` element, placed wherever the change occurs, with the `@new` attribute having as value the id of the hand that is taking over. The text is then marked up as normal. If the first hand takes over again, then this needs to be indicated with another `<handShift>` element. If a new `<div>` (whether a sibling or child of the current `<div>`) is added, then its `@resp` attribute will update the hand again.

Whether to use `<handShift>` or `<add type="insertion" place="inline">` is determined mainly by the nature of the new hand's contribution. If it is brief and seems like maintenance of the main text, it is an insertion. If it is longer and seems integral to the main text, it is a change of hand.

Unclear Text

Manuscripts vary in their physical condition and authors/scribes vary greatly in their adherence to orthographies, dialects, and standards that are recognisable to us. The resulting issues need to be documented within our mark-up for the information of FnaG's lexicographers.

If the resulting issues could render the text unhelpfully misleading for users, editorial emendation should be considered instead (on which, see the following section).

Problematic Text

Unclear Text

Text that cannot be read and interpreted with certainty should be placed in an `<unclear>` element. `<unclear>` can be used to enclose specific characters within a `<w>` element or multiple elements at `<w>` level and above. It should be as focused as possible on the section of the text which is unclear.

Use of `<unclear>` assumes that some sort of approximation can be made of the text enclosed in it. If the text cannot be reproduced, a `<gap>` element should be used instead (see below).

`<unclear>` takes a number of attributes:

- `@reason`: what is making the enclosed text unclear, of the following values:
 - “text_obscure”: characters are visually difficult to identify (e.g. APPENDIX 8b ll.49–56);
 - “damage”: loss of writing surface means that characters are missing;
 - “fold”: the writing may well be legible but the arrangement of the page in the digital image obscures it;
 - “interp_obscure”: there is no visually apparent problem but a word cannot be identified satisfactorily (e.g. APPENDIX 9b ll.41–44);
 - “abbrv”: an uncertain expansion of an abbreviation is hampering interpretation;
 - “char”: a specific character is ambiguous, hampering interpretation;
 - “met”: a line of verse is metrically irregular;
- `@cert`: how certain the text enclosed in `<unclear>` is, of the following values:
 - “high”;
 - “medium”;
 - “low”;
 - “uncertain”;
- `@resp`: who is responsible for the reading of the text enclosed in `<unclear>`; this might be the transcriber, a reviewer of the transcription, or a bibliographical reference to an edition of the text outside of the project.

Placement of `<unclear>` will depend on `@reason`. `<unclear reason="char">` is designed to enclose a single character, for example, and `<unclear reason="met">` only makes sense in the context of poetry. `<unclear>` elements can be placed within other `<unclear>` elements and

thus build a more detailed account of the problems the transcriber has experienced. For example, at APPENDIX 9b II.47–57, the form “cuis” (?) has been placed in <unclear reason="text obscure">. On account of the difficulties of reading this word, the phrase of which it is part (“ni is cuis”) has been placed in <unclear reason="interp obscure">. The problems with making out a key word have thrown into question the interpretation of other words which themselves can be made out visually.

<unclear reason="damage"> is distinct from all the others. All the others indicate a problem with reading the text that could theoretically be resolved by someone with better perception, knowledge, equipment, luck, or suchlike. Meanwhile, <unclear reason="damage"> indicates that text has been lost irrecoverably and that what is enclosed in the element is hypothesised by an editor (see also ILLEGIBLE OR LOST TEXT, below). <unclear reason="damage"> implies reconstruction of text that the scribe once did produce but its use should be reserved for where the writing surface has been completely removed.

In terms of output, the @cert value will generate a colour code for the enclosed text indicating the level of difficulty and messages will display in a word's tooltip detailing any problems affecting that word:

Figure 54

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.
1. { re sine \gl. .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis } 3. gan resun { is } -gap- 4. { ni is { cuis } } -gap-

Add Comment

[Back to MS contents](#)

ni: noun
unk unk. (24)
- the interpretation of this word, or its context, is doubtful; there is a particular issue with "cuis"
Moderate problems with this reading
Abbreviations: none

Nature

Figure 55

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.
1. { re sine \gl. .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis } 3. gan resun { is } -gap- 4. { ni is { cuis } } -gap-

Add Comment

[Back to MS contents](#)

cuis: noun
unk unk. (24)
- the interpretation of this word, or its context, is doubtful
- this word is difficult to decipher
Moderate problems with this reading
Abbreviations: none

Nature

Unidentifiable Words

If a word cannot be identified with any known dictionary headword, regardless of confidence level (doubtful identifications can be indicated using `<unclear reason="interp obscure">`), then the `@lemma` of the relevant `<w>` should be set to “UNKNOWN” (case sensitive). `@ana` can be given if the POS is obvious.

New Words

If a word does not match a dictionary headword but can be interpreted using a dictionary headword (e.g. an otherwise unattested adjectival form of an attested noun), then the attributes in `<w>` are slightly different:

- `@lemma`: the proposed dictionary headword for the new word, prefixed with “*”;
- `@source`: the existing dictionary headword used to interpret it or on which it seems to be based;
- `@lemmaRef`: the url for `@source`;
- `@ana`: the POS of the word as used in the transcription.

This can also be used for items that are not necessarily new to Gaelic philology but whose interpretation requires explanation. For example, see APPENDIX 7b ll. 125–131. Here, it is hypothesised that “caithe” is a past participle of the verb “caithid”. Normally, a participle could be understood as a form of the verb and the verb simply given as the headword. In this case, the interpretation is more doubtful, so both `@lemma` and `@source` have been given.

Problematic Abbreviations

As already noted in CHAPTER 3 (ABBREVIATIONS), the `<abbr>` element also has a `@cert` level that can be used to indicate uncertainty (or the lack thereof). Adding “medium” or “low” as values will produce colour coding and messages flagging up the problem with the abbreviation, which can then be further examined via the e-slip table.

Figure 56

Form	MS Ref.	MS Context	Text Form	Issues?	Abbrevs
cumsgaigthe	MS17 24r 9	... 9. tosach cumsanaid agus cumsgaigthe maille 10. fein ni co ...	Prose	moderate;	M_stroke Suspension_stroke

Figure 57

Form	MS Ref.	MS Context	Text Form	Issues?	Abbrevs
cumsgaigthe	MS17 24r 9	... 9. tosach cumsanaid agus cumsgaigthe maille 10. fein ni co ...	Prose	moderate;	M_stroke Suspension_stroke

Note that, unlike with `<unclear>`, a “high” value for `@cert` in `<abbr>` denotes an absence of uncertainty, not “slight” uncertainty.

Illegible or Lost Text

Gaps

If a section of text cannot be made out at all and cannot be hypothesised with sufficient confidence, a closed `<gap/>` element should be placed there, with the following attributes:

- `@reason`: what has caused the gap, of the following values:
 - “text obscure”: the text is illegible, although the writing surface is intact”;
 - “damage”: the writing surface has been lost;
 - “fold”: the writing may well be legible but the arrangement of the page in the digital image obscures it;
 - “text omitted”: there is no visually apparent problem but disjunction in the text suggests that material has been omitted;
- `@extent`: the length of the gap, using whatever unit seems most appropriate (e.g. “3 chars”, “1 ms. line”); “unknown” is also an option, particularly where loss of writing surface is involved;
- `@resp`: the id for the individual responsible for measuring and assessing the gap.

See APPENDIX 9b II. 45, 58 for examples.

Characters

If some characters can be made out, but not an entire identifiable word, then these characters can be enclosed in a `<c>` element (no attributes required), with `<gap>` being used where the word cannot be made out. However, note that `<c>` elements cannot, in TEI, contain `<abbr>/<g>` elements. So, if the word fragment includes an abbreviation, it will have to be placed in a `<w>` element for the abbreviation to be expanded. The `@lemma` can then be set to “UNKNOWN” and `<gap>` elements placed to indicate the lost portions of the word.

Editorial Emendations

Transcriptions in the FnaG MSS Corpus should in general reflect the forms used in the manuscript at hand as closely as possible. However, where a form deviates severely from what is expected, to the extent that interpretation is obscured, editorial interventions can be made. This is both to avoid causing users needless confusion and delay and to “show our working”, in terms of how the form came to be identified. At the same time, it is important the the original manuscript reading be always retrievable.

This is done in a number of ways.

Supplying Characters

If the emendation simply involves adding characters to a word, then the characters should be included within the `<w>` element and then enclosed within `<supplied>`, which should have a `@resp` attribute indicating who supplied the characters. For an example, see APPENDIX 8b I.63.

In the editorial text, the added characters will be enclosed within square brackets and a tooltip note will flag up the fact that an editor has intervened in the form. In the diplomatic text, these characters will be omitted silently.

`<supplied>` cannot be used inside `<g>`. If characters need to be supplied inside the expansion of an abbreviation, then the abbreviation should be expanded in amended form and the `@cert` level of `<abbr>` adjusted downwards, should the resulting expansion be unusual.

Supplying Words

Entire words (`<w>`) can also be enclosed by `<supplied>`, again with a `@resp` attribute identifying the editor responsible. `<w>` elements enclosed by `<supplied>` should be otherwise marked up as normal. Whether such `<w>` elements should contain `<abbr>/<g>` elements depends on why the form is being supplied (perhaps so if the omission was accidental, perhaps not if the text is inherently faulty) and whether any clear models exist for how that scribe would have rendered that word. It is ultimately up to the transcriber.

Emended Forms versus Original Forms

If a more complex emendation of a word is required, which involves either deleting characters or re-arranging their order, then a `<choice>` element is used (without attributes). The `<choice>` element encloses `<sic>` and `<corr>`. The `<w>` element, with all attributes, containing the original manuscript reading is placed in `<sic>` and the `<w>` element, with all attributes, containing the emended form is placed in `<corr>`, which has a `@resp` attribute indicating the editor responsible. Any `<space/>` elements or suchlike should be placed outside of `<choice>`.

The model is therefore as follows:

Figure 58

<code><choice></code>
<code> <sic></code>
<code> <w><!-- MS reading here --></w></code>
<code> </sic></code>
<code> <corr></code>
<code> <w><!-- Emended form here --></w></code>
<code> </corr></code>
<code></choice></code>

The output will display the emended form, followed by “*/alt*”. Both will be enclosed in curly brackets. Tooltip messages displayed by each will indicate the original and emended forms and an e-slip generated from “*/alt*” will display the original reading.

Competing Readings

Functionality was developed to allow the transcriber to encode two alternative readings of short strings of words in parallel to one another, if there really is no way of choosing between them. This would not imply that one was correct/emended but that both were possibilities. The functionality consisted of placing two *<unclear>* elements within a *<choice>* element and then using a system of reference numbers to link the counterpart *<w>* elements in each *<unclear>* element to one another.

However, the results proved very difficult to process coherently and the whole concept could be characterised as an abdication of editorial responsibility. For these reasons, this functionality has been deprecated for now, although examples can still be found in the corpus. Instead, in such situations, it is suggested that one reading be chosen by whatever means, the *<unclear>* element used if needed, and any alternative readings set out in a footnote.

If this functionality or something like it seems worthwhile and worth the effort, then this matter can be re-visited.

Review

- Numerous aspects of page layout, textual apparatus, and scribal behaviour can be marked up in TEI.
- TEI mark-up can also be used to track the hand producing text at any given point.
- Unclear or problematic text can be marked up and assessed in terms of cause and confidence of interpretation.
- There are various ways to amend transcriptions while still retaining the original manuscript reading.

Chapter 5

Transcription Introductions

“Redig dam a De do nim
cindead coir ar chomaimsir
a lis cech leirfhilead lind
fis na reigided romaind.

Indeosad co seim solas
cach re head na amaros
soirb cach daigen co dagar
doirb aiged ar an dalad.”

- [?Giolla Íosa Mór?] Mac Firbhisigh
(*Book of Lecan*, fol. 190^ra33–38).

Introduction

As well as the transcriptions themselves and the metadata embedded in the mark-up, the transcription files also require introductions. These contextualise them in historically, culturally, and socially and highlight their major linguistic pitfalls and points of interest. Like the texts, the introductions are marked up in TEI/XML.

However, the mark-up is slightly different from that used for the transcriptions themselves. This chapter sets out the main techniques involved in producing scholarly writing and attendant apparatus in the project's adaptation of TEI. Note that it does not cover the specific content or structure of different sections of the introductions nor what is to be dealt with where. This is summarised in APPENDIX 5b but is primarily to be decided under the direction of the Manuscripts Consultant.

The facilities set out in this chapter can also be used in footnotes to the transcriptions proper.

Unfortunately, at the time of writing, only two files in the corpus (T2 and T8) currently have viable introductions. Adding more introductions is a high priority for the project.

Formatting Text

When writing introductory material (i.e. within <teiHeader>), text can mostly be inputted directly, without the word-by-word, space-by-space mark-up used in the transcriptions. The basic text formatting (bold, italics, etc.) available in most word processors can be reproduced in our TEI. Input within some elements, such as titles, will be formatted automatically but certain tools are at the general disposal of the writer of the introduction. As previously, text is formatted by enclosing it within a particular element. That element is <hi> ("highlight"). The formatting is then determined by the value of @rend ("rendering"):

Bold: "bold";

Italics: "italics";

Underline: "underline";

Subscript: "sub";

Superscript: "sup";

Thus, to produce the manuscript reference, "fol. 24^r10", within a paragraph in an introduction:

```
... fol. 24<hi rend="sup">r</hi>10 ...
```

<hi> elements can be placed within each other to format text multiple ways simultaneously or to add additional formatting to a particular substring of text. For example,

```
<hi rend="italics"><hi rend="bold">This text is in italics and in bold.</hi></hi>
```

This text is in italics and in bold.

```
<hi rend="italics">This text is mostly in italics, although with occasional use of <hi rend="bold">bold</hi> for emphasis.</hi>
```

This text is mostly in italics, although with occasional use of bold for emphasis.

It is also possible to colour text. To do so,

1. Enclose the text in <hi rend="colour:"> ... </hi>;
2. Visit https://www.w3schools.com/colors/colors_picker.asp and select the name or code for the desired colour (omit the prefixed "#" from the code);
3. Add the name or code after "colour:" in the @rend value.

Thus,

```
<hi rend="colour:red">This text is red.</hi>
```

This text is red.

TEI/XML — and HTML, into which it will be transformed — offer a lot of options for formatting text (this is the same underlying technology that powers most word processors, after all), so, if more functionality is needed than is set out here, this should be made known to the project team, as probably can be added.

Lists

To form a list within a `<p>` element,

1. Add a `<list>` element;
2. For each item, add an `<item>` element as the child of `<list>`, with the text as its content.

If needed, sub-headings within the list can be added by adding a `<head>` element as a child of `<list>`, with the sub-heading as its direct text content.

As an example, the REVIEW section from the previous chapter would be marked up as shown in Figure 59.

Figure 59

<code><list></code>
<code> <item>Three textual forms are currently recognised in the corpus: "prose", "verse", and "divProse".</item></code>
<code> <item>Texts can consist of combinations of multiple forms.</item></code>
<code> <item>A text's form is indicated by the @type attribute of its enclosing &lt;div&gt; element.</item></code>
<code> <item>The form determines the text's overall TEI structure.</item></code>
<code> <item>A change of form is represented by embedding a &lt;div&gt; element within the text's enclosing &lt;div&gt; element; this can be a complex operation.</item></code>
<code> <item>Embedded &lt;div&gt; elements within &lt;div&gt; must correspond to &lt;msItem&gt; elements in the transcription file header unless they stand for text at the top-level in the text (i.e. returning to the top-level &lt;div&gt;).</item></code>
<code></list></code>

These steps will produce a bullet-pointed list. To produce a numbered list, add `@n` to each `<item>/<head>` element, with whatever sequential value is required under the chosen system (letters, Arabic numerals, Roman numerals, etc.). For example, the two-step process set out at the beginning of this section would be marked up as in Figure 60.

Figure 60

<code><list></code>
<code> <item n="1"> Add a &lt;list&gt; element;</item></code>
<code> <item n="2"> For each item, add an &lt;item&gt; element as the child of &lt;list&gt;, with the text as its content.</item></code>
<code></list></code>

`@n` can be omitted from any element. For example, it can be omitted from `<head>` and the numbering restarted in the new section of the list.

Tables

Adding a table in TEI/XML is somewhat complex, as one needs to map out the table using simply an element hierarchy rather than any sort of visual interface.

1. A new paragraph (<p>) should be created for the table;
2. Add a <table> element as child of <p>, which will enclose the entire table;
3. If the table is to have column headers, add a <row> element as child of <table>, which will enclose the header row of the table. It should have the attribute: @role: "label";
4. For each column in the table, add a <cell> element as a child of <row role="label">; each <cell> element should have the attribute @role: "data" and contain as text content the header for each column (if applicable);
5. For each data row in the table, add a <row> element with @role "data", then, for each column in that row, add a <cell role="data"> element, with the appropriate data as its text content.

A header row (<row role="label">) can be added at any point, with a new set of column headers. The number of columns can also be changed within the table. In a <cell> element, add the attribute @cols, with the value being the number of columns the cell should take up.

Thus, if a table began with five columns, the code in Figure 61 would begin a new section of the table with only two (wider) columns, each with a new heading:

Figure 61

```
<row role="label">
    <cell role="label" cols="1">
        <!-- column heading here -->
    </cell>
    <cell role="label" cols="4">
        <!-- New column heading here -->
    </cell>
</row>
```

Of course, it may well prove simpler to add a new table in a separate <p> element.

As an example, Figure 62 presents a table summarising the texts used as examples in this manual:

Figure 62

```
<p>
    <table>
        <row role="label">
            <cell role="label">Manuscript Shelfmark</cell>
            <cell role="label">Text Title</cell>
            <cell role="label">Form</cell>
            <cell role="label">Corpus Transcription Number</cell>
        </row>
        <row role="data">
            <cell role="data">NLS Adv. 72.1.36</cell>
```

```

<cell role="data"><hi rend="italics">Bruighion Bheag na
hAlmhain</hi></cell>
<cell role="data">Prose</cell>
<cell role="data">12.2</cell>
</row>
<row role="data">
    <cell role="data">NLS Adv. 72.1.36</cell>
    <cell role="data">' Is maith mo leaba, is olc mo
shúain'</cell>
    <cell role="data">Verse</cell>
    <cell role="data">12.4</cell>
</row>
<row role="data">
    <cell role="data">NLS Adv. 72.1.4</cell>
    <cell role="data">Definitions</cell>
    <cell role="data">Divided Prose</cell>
    <cell role="data">17.1</cell>
</row>
<row role="data">
    <cell role="data">NLS Adv. 72.2.11</cell>
    <cell role="data"><hi rend="italics">Cath
Fionntrágha</hi></cell>
    <cell role="data">Multiple Forms</cell>
    <cell role="data">3.1</cell>
</row>
<row role="label">
    <cell role="label" cols="2">Transcriber</cell>
    <cell role="label" cols="2">Text(s)</cell>
</row>
<row role="data">
    <cell role="data" cols="2">EPT</cell>
    <cell role="data" cols="2">12.2, 12.4, 17.1</cell>
</row>
<row role="data">
    <cell role="data" cols="2">MMA</cell>
    <cell role="data" cols="2">3.1</cell>
</row>
</table>
</p>

```

This renders the HTML output as shown in Figure 63.

Figure 63

Manuscript Shelfmark	Text Title	Form	Corpus Transcription Number
NLS Adv. 72.1.36	<i>Bruighion Bheag na hAlmhain</i>	Prose	12.2
NLS Adv. 72.1.36	'Is maith mo leaba, is olc mo shúain'	Verse	12.4

NLS Adv. 72.1.4	Definitions	Divided Prose	17.1
NLS Adv. 72.2.11	<i>Cath Fionntrágha</i>	Multiple Forms	3.1
Transcriber		Text(s)	
EPT		12.2, 12.4, 17.1	
MMA		3.1	

Quotations

Brief quotations can simply be placed in quote marks ("...") within the paragraph, following the usual syntactic conventions (for adding references, see REFERENCES, below). If you are quoting a transcription from the same file in brief, bear in mind the automated facility described in REFERENCES, below, as this may render quotation of context unnecessary.

Longer quotations should be placed in a `<quote>` element (child of `<p>`). This will present the quotation in a separate, indented paragraph. After the `<quote>` element has been closed, the enclosing `<p>` element should also be closed and a new `<p>` element opened if need be. The structure thus follows the *opus geminatum* format that emerges from embedding `<div>` elements in other `<div>` elements (see MULTIPLE FORMS).

Figure 64

```
<p>
    <!-- Prose commentary here -->
    <quote>
        <!-- Long quote here -->
    </quote>
</p>
<p>
    <!-- Next paragraph here -->
</p>
```

If the long quotation is in prose, it can simply be inputted as the direct text content of `<quote>`. If it is verse, then `<l>` elements can be placed as the children of `<quote>` and each line added therein.

Figure 65

```
<p>
    <!-- Prose commentary here -->
    <quote>
        <l><!-- Verse quote: line 1 --></l>
        <l><!-- Verse quote: line 2 --></l>
        <l><!-- Verse quote: line 3 --></l>
        <l><!-- Verse quote: line 4 --></l>
    </quote>
</p>
<p>
    <!-- Next paragraph here -->
</p>
```

References

Form of Citation

Citations in introductions and footnotes to FnaG MSS transcriptions are by author-date. Thus, the first page of the following work,

William Gillies, ‘The Gaelic of Niall MacMhuirich’, *Transactions of the Gaelic Society of Inverness* 65 (2011), 69–95,

would be cited as follows: “Gillies 2011a: 69”.

If, as is the case here, author + date does not uniquely identify the work in question within the corpus (i.e. the author has multiple publications in that year that appear in the FnaG MSS Project bibliography), letters of the alphabet should be added to the dates in the citations to render them unique.

Citations of manuscripts that appear in the corpus should be by their corpus code (e.g. “MS12”). Citations of other manuscripts should be in the shortest intelligible form. Citations of texts from the corpus can be by title or by corpus number (e.g. “Bruighion Bheag na hAlmhain” or “MS12.2”).

Mark-Up

When you cite a manuscript, a piece of primary or secondary literature, or the transcription you are introducing, or a known scribal hand, a reference can be added to it within the XML document as well as by the conventional, human-readable means. This allows further details to be retrieved and displayed at the point at which the citation is made, for the convenience of users.

Enclose the key component of the reference (e.g. “2011a”, in the above example) in a `<ref>` element, with the following attributes:

`@type`: the type of reference being made, of the following values:

- “bib”: A piece of primary or secondary literature listed in the project bibliography;
- “ms”: A manuscript listed in the project bibliography;
- “text”: A text transcribed within the present transcription file (i.e. a `<div>`);
- “hand”: A hand registered within the corpus;
- “text_ed_line”: A line from a text in the present transcription file, according to an editorially analysed structure (e.g. a line of verse);
- “text_dip_line”: A manuscript line from a text in the present transcription file;

`@target`: the `@xml:id` for whatever is being cited.

Care should be taken to ensure that the right `@type` is used for the right citation, particularly in the case of “text_ed_line” versus “text_dip_line”.

For `@target`, the id should be sought:

- In `<listBibl>` in `corpus.xml` for “ms” and “bib”;
- In `<msContents>` in the present file for “text”;
- In `<handNotes>` in `corpus.xml` for “hand”;

- Within the marked-up text for “text_ed_line” (<1>) and “text_dip_line” (<1b>).

The result, from the above example, would be:

Gillies <ref type="bib" target="Gillies2011a">2011a</ref>: 69

As a result, in the HTML output, clicking on the reference will generate inline:

- A long version of the citation in the case of “bib” or “ms”;
- Further details in the case of “text” or “hand”;
- A copy of the metrical or topic-based line in the case of “text_ed_line”;
- A copy of the manuscript line in the case of “text_dip_line”.

This also allows the citations of manuscripts and primary and secondary literature in each file to be tracked and relevant bibliographies to be generated automatically in each instance.

Below, Figure 66 shows a paragraph from the introduction to T2 with some inline citations of secondary literature and manuscripts as they first appear. Figure 67 shows all the citations expanded.

Figure 66

Other Activities

The same scribe is also the sole hand in the first fragment (fol. 1–5) bound into the composite manuscript, Edinburgh, NLS Adv., MS 72.1.31. This fragment preserves at least 13 poems in his hand about or attributed to Colum Cille, although its poor condition frustrates close examination of many of them. The genre as whole has been examined in detail by Herbert (2005). Four of these poems, as well as a text of ‘Aingil De dom dhin’, appear in Oxford, Bodleian Library, MS. Laud Misc 615, which Herbert and O’Sullivan (1973: 175) date to 1532–1545 and associate with the circle of scholars working on Columban material for the Ó Neills of Tyrone, NW Ireland. Furthermore, the page layout of the first fragment in NLS Adv. 72.1.31 — oblong pages with long lines, often one quatrain per line — resembles that of Laud Misc. 615. Indeed, the quatrains of ‘Aingil De dom dhin’ in Egerton 2899 are each also set out one to a line. This might suggest that the present scribe had intellectual contact with scholars in the north of Ireland and drew on similar literary resources.

Add Comment

Figure 67

Other Activities

The same scribe is also the sole hand in the first fragment (fol. 1–5) bound into the composite manuscript, Edinburgh, NLS Adv., MS 72.1.31 [Edinburgh, NLS Adv., MS. 72.1.31]. This fragment preserves at least 13 poems in his hand about or attributed to Colum Cille, although its poor condition frustrates close examination of many of them. The genre as whole has been examined in detail by Herbert (2005) [Herbert (2005), Maire, ‘Becoming an Exile: Colum Cille in Middle-Irish Poetry’, in Joseph F. Nagy (ed.) and L. E. Jones (ed.), *Heroic Poets and Poetic Heroes in Celtic Tradition: A Festschrift for Patrick K. Ford*, CSANA Yearbook 3–4, Dublin: Four Courts Press, 131–40]. Four of these poems, as well as a text of ‘Aingil De dom dhin’, appear in Oxford, Bodleian Library, MS. Laud Misc 615, which Herbert and O’Sullivan (1973) [Herbert (1973), Maire, and Anne O’Sullivan, ‘The Provenance of Laud Misc. 615’, *Celtica* 10, 174–92] — oblong pages with long lines, often one quatrain per line — resembles that of Laud Misc. 615. Indeed, the quatrains of ‘Aingil De dom dhin’ in Egerton 2899 [London, British Library, MS. Egerton 2899] are each also set out one to a line. This might suggest that the present scribe had intellectual contact with scholars in the north of Ireland and drew on similar literary resources.

Add Comment

Figure 68 shows a paragraph, also from T2, on some of Hand7’s traits. Figure 69 shows the citations of the text, some of which are to the editorial text and some of which are to the diplomatic text, in their expanded form.

Figure 68

Other traits

- As noted, the scribe not infrequently uses pronunciation spellings (e.g. “na b[p]earn” (q20b)). On other occasions, however, the forms that he renders imply a misunderstanding of the text’s sense. For example, he appears to be referring to an eighth (otherwise unknown) archangel when he writes “scaphel blath | a bharr” (fol. iii¹5–6; “Scaphael, fine his helmet”); other manuscript witnesses have “scathfhial” (“sheltering veil”) at this point. In another example, he suspiciously repeats “uriel” (fol. iii¹10) in the verse on Archangel Uriel at a point where an adjective would be expected (= “ur[fh]iel”, air- + fial?), although there is confusion regarding this line throughout the manuscript witnesses (O’Nolan 1912: 255).

Add Comment

Figure 69

Other traits

- As noted, the scribe not infrequently uses pronunciation spellings (e.g. "na b[p]jeann" (q20b

parthalon na bpeann

)). On other occasions, however, the forms that he renders imply a misunderstanding of the text's sense. For example, he appears to be referring to an eighth (otherwise unknown) archangel when he writes "scaphel blaith | a bharr" (fol.iii^f5

airCor dibh michelmora gloir nigloir andraphael doraidh rumscaphelblaith

-6

a bharrCor dibhuriel ámrénacgadromfeithcor dibh sariel smuthbanniam de aradhreich

; "Scaphael, fine his helmet"); other manuscript witnesses have "scathfhal" ("sheltering veil") at this point. In another example, he suspiciously repeats "uriel" (fol.iii^f10

Uriel fial na feart~~dibh uriel~~ na dlochtsachedain famchorpnaeg again antole

) in the verse on Archangel Uriel at a point where an adjective would be expected (= "ur[fb]iel", *air-* + *fial*?), although there is confusion regarding this line throughout the manuscript witnesses (O'Nolan 1912: 255).

Add Comment

The text content of <ref> is not relevant from a processing point of view; the key data is contained in its attributes. Thus, any sort of reference to the resource or entity in question can be enclosed in <ref>, not just a formal citation. Indeed, <ref> might be useful for providing clarification under such circumstances.

Automated Features

It should be noted that the transcription file header, once processed using stylesheet-ed-pub.xsl, will be slightly different from the raw TEI/XML:

- The sections will be in a slightly different order;
- Some generic material will have been imported from corpus.xml;
- Some additional data will have been generated automatically; e.g.:
 - A word count will be calculated in <fileDesc>;
 - A list of hands active in the text, and their respective roles, will be added for each <msItem>;
 - A *Works Cited* list, generated from <ref> elements in the file, will be added as the last child of <body>.

It is adviseable to make each section of the transcription file header as self-contained, in terms of content, as possible, while avoiding gross repetition of material. This is not only because the header will be re-structured in the output but because different sections of the header might end up being viewed in isolation from each other, depending on how the corpus is eventually made available.

That being said, it is also adviseable to acquaint oneself with the structure of the transcription file header output before planning the content for the input. More introductions to files in the corpus will be made available in the near future and these can be studied to this end.

Review

- Each transcription file must have an introduction (<teiHeader>) covering the literary, historical, and linguistic context of the file's contents.
- The content of the introductions is set out in APPENDIX 5b and in instructions from the Manuscript Consultant.
- Various facilities are available to replicate scholarly apparatus in the transcription file introductions.
- Some of these will generate interactive reading aids in the HTML output.
- The HTML output will re-structure the introduction, as authored in the original TEI/XML file.

Appendices

Appendix 1. The same table could be generated for the entire corpus or for any combination of texts or transcriptions.

Specific Queries ([stylesheet-ed-ana-specific.xsl](#))

A list of words selected according to specific criteria, with data on each word, can be generated using [stylesheet-ed-ana-specific.xsl](#). This consists of largely the same code as [stylesheet-ed-ana-general.xsl](#), except it is set up to allow easy inputting of search criteria.

For a table of all instances of “oidhche” (i.e. `<w>` elements tagged with the eDIL headword, [*adaig*](#)) in the corpus, generated using [stylesheet-ed-ana-specific.xsl](#), see Appendix 2.

Gender/Stem Data ([stylesheet-ed-ana-stemGen.xsl](#))

[stylesheet-ed-ana-stemGen.xsl](#) likewise generates a word list from either the corpus or selected transcriptions. However, in addition, it looks up headwords in the headword database ([hwData.xml](#)) and adds data on gender and stem, where applicable.

For a table of all nouns in ‘Soraidh slán don aoidhche areir’ for which stem and gender data is currently available,⁵ see Appendix 3.

Corpus Stats ([stylesheet-ed-ana-stats-detailed.xsl](#))

As well various sorts of word list, statistical or factual answers to various sorts of query can also be generated. For example, Appendix 4 contains a miscellaneous body of information about the corpus (generated 23/07/2018) that has been generated by applying [stylesheet-ed-ana-stats-detailed.xsl](#) to all transcription files. Further pieces of information could certainly be obtained through further work devising queries.

Re-Processing

Once an edition has been generated, the user can interact with it in various ways to aid their reading, collaborate with others, or export data for future reference and analysis. This is accomplished through a series of tools written in JavaScript (file extension: .js), a programming language that can be used to query and manipulate HTML.

For a full demonstration of these tools, see the accompanying video: [FnaG MSS – Editions Demo](#).

Tools (JavaScript)

Word Highlighting ([hilites.js](#))

When the mouse is placed over a word in the transcription in the HTML edition, the word is highlighted **in blue** to show the boundaries of the word, as marked up. This can be useful if additions or deletions have been made to the word, or if it is bisected by a line or page break, for example.

Reference Viewing ([ref.js](#))

In the introductions to the transcriptions, citations are made of primary and secondary literature. Reference is also made to specific points in the text being introduced. This function expands citations of external sources (i.e. primary and secondary literature in-line), within the introduction, for convenience (they are also presented in a Bibliography for each transcription file).

⁵ Stem and gender data has, at time of writing (23/07/2018), only been added for the 200 most frequently occurring words in the corpus.

In the case of references internal to the transcription file, expanding the reference generates a hyperlink that links to the cited location in the file.

[Comments \(textComm.js\)](#)

A high level of accuracy is required in transcriptions for the FnaG MSS Project. Therefore, an extensive review process between the Manuscript Research Fellows, the Manuscript Consultant, and subject specialists (as appropriate). Developers may also need to examine and feedback on the outputs in the course of scrutinising the tech-related aspects of the Project.

For this reason, a comment function has been developed that allows users of the HTML editions of the transcriptions to add and within the introduction and text and within the data generated via WordFile.js (see below). The relevant file(s) can then be saved, allowing the comments to be shared with collaborators.

[Generate e-Slips \(WordFile.js\)](#)

This tool performs two functions:

1. A word is selected from an HTML edition of a transcription; an “e-slip” **for that word** is added to a table (see Figure 17).
2. A word is selected from an HTML edition of a transcription; an “e-slip” **for every instance** of that headword in the current transcription file is added to a table.

The generated tables form separate HTML files, which can be saved separately for future reference or analysis. For their internal workings, see below.

[e-Slip Data \(HTML\)](#)

Once a table of e-Slips (see Figure 17) has been generated (see above), the user can interact with it in a number of ways (Figure 18).

- Comments can be inserted (“Comment”), using a version of textComm.js embedded automatically in the table (see Figure 19).
- Unwanted rows can be removed via the “Del. Row” button.
- The table can be copied and pasted into an MS Excel spreadsheet (Figure 20), so as to gain access to MS Excel’s powerful sorting and filtering features; the buttons in the rightmost column can be removed via the “Del Col.” button.
- The table can also be saved as an HTML file (if opened in Firefox) or printed as a PDF (all browsers); only the former retains the comment function for subsequent input from collaborators.

Figure 18

Lemma (eDIL)	Lemma (Dwellys)	HDSG Slip	URI	Del Col
				Del Row
soraid	soraidh		38494	Del Row
				Comment
slán	slán		37834	Del Row
				Comment
do			17096	Del Row
				Comment
adaig	oidhche		256	Del Row
				Comment
irráir			29082	Del Row
				Comment

Figure 19

Form	MS Ref.	MS Context	Text Form	Issues?	Abbrevs	Scribe	Scribe Date	PoS	Lemma (eDIL)	Lemma (Dwellys)	HDSG Slip	URI	Del Col
													Del Row
Soraidh	MS12 85v 1	Soraidh slan don aoidc/e [a]reir ...	Verse		Spiritus_asper	Eoghan Mac Gilleoin (9)	17	noun	soraid	soraidh		38494	Comment
	Mon, 23 Jul 2018 14:58:12 GMT	Name: EPT		Comment:	Is this a noun or an adjective, in the present context?								
slán	MS12 85v 1	Soraidh slán don aoidc/e [a]reir fada ...	Verse			Eoghan Mac Gilleoin (9)	17	adj	slán	slán		37834	Del Row Comment
don	MS12 85v 1	Soraidh slán don aoidc/e [a]reir fada gear ...	Verse			Eoghan Mac Gilleoin (9)	17	prep, art	do			17096	Del Row Comment
aoidc/he	MS12 85v 1	Soraidh slán don aoidc/he [a]reir fada gear ag ...	Verse		Spiritus_asper	Eoghan Mac Gilleoin (9)	17	noun	adaig	oidhche		256	Del Row Comment
[a]reir	MS12 85v 1	... Soraidh slán don aoidc/he [a]reir fada gear ag dul ...	Verse	supp. char(s)		Eoghan Mac Gilleoin (9)	17	adv	irráir			29082	Del Row Comment
	Mon, 23 Jul 2018 14:57:03 GMT	Name: EPT	Comment:	Double check for lenition on the "d".									Del. Edit
				Comment:	Elision required for metre.								

Figure 20

	Form	MS Ref.	MS Context	Text Form	Issues?	Abbrevs	Scribe	Scribe Date	PoS	Lemma (eDL)	Lemma (Dwellys)	HDSG Slip	URI	
1														
2														
3														
4	Soraidh	MS12 85v1	Soraidh slan don aoidch hə (ə)þeur	Verse			Spiritus asper	Eoghan Mac Gilleoin (9)	17 noun	soraidh	soraidh		38494	
5														
6	Mon, 23 Jul 2018 14:58:12 GMT	Name:	EPT	Comment:	Is this a noun or an adjective, in the present context?									
7														
8	slan	MS12 85v1	Soraidh slan don aoidch hə (ə)þeur fada ...	Verse			Eoghan Mac Gilleoin (9)	17 adj	slán	slán			37834	
9														
10														
11														
12	don	MS12 85v1	Soraidh slan don aoidch hə (ə)þeur fada geor ...	Verse			Eoghan Mac Gilleoin (9)	17 prep, art	do				17096	
13														
14														
15														
16														
17	aoidech	MS12 85v1	Soraidh slan don aoidch hə (ə)þeur fada geor ag ...	Verse			Spiritus asper	Eoghan Mac Gilleoin (9)	17 noun	adaig	cidhche		258	
18														
19	Mon, 23 Jul 2018 14:57:03 GMT	Name:	EPT	Comment:	Double check for lenition on the 'd'.									
20														

Review

This section has provided an overview of the main technical components and output types of the FnaG MSS Corpus. The input comprises a relatively manageable and intuitive quantity of TEI mark-up of transcriptions, as well as maintenance of two TEI/XML databases (the corpus file and the headword database). This can then be transformed into usable HTML files in a variety of ways to meet a variety of needs using a range of XSLT stylesheets, alone or in combination. HTML editions can then be further manipulated via JavaScript to enhance usability or to export selected data.

Moving as much of the processing work as possible away from the individual transcription files to common TEI/XML databases and XSLT stylesheets is designed to speed up the process of transcribing and/or marking up of transcriptions, as well as to maximise control and options in terms of output. However, this strategy relies on the underlying mark-up being accurate and consistent. It also means that understanding the corpus means understanding multiple interlocking components. This User Documentation sets out how both can be achieved.

Software Required

Various software packages are required to work on the FnaG MSS Project. These are listed below. While alternatives to some may well exist, this User Documentation assumes that the reader has access to those on the list. It is therefore recommended that these be acquired in the first instance.

oXygen

oXygen XML Editor is the main software package used on the Project. It is an editing package that supports the writing of XML, XSLT, HTML, and JavaScript code (and much else besides). Features include:

- Automatic highlighting and indentation of code, for ease of reading;
- A checker that identifies syntax errors;
- A validator that reports most nonsensical code;
- Onmouseover reminders of the function of each element;
- In-line drop-down menus containing all possible (i.e. valid) elements that can be inserted at the cursor's location;
- Interfaces and validators for xPath/xQuery;
- A “Grid” view, which transforms XML files into tables; this effectively renders working with the headword database like working with a spreadsheet, even though it is written in XML.
- A processor that can be used to apply XSLT stylesheets to XML.

For more information: <https://www.oxygenxml.com/>

oXygen XML Editor is expensive. A license will be provided to each Manuscripts Research Fellow by Faclair na Gàidhlig.

Saxon

Saxon, like *oXygen*, can be used to transform XML documents into HTML or other XML documents. However, in the context of the FnaG MSS Project, it has proved to be faster than the in-built *oXygen* processor. Also, *oXygen* cannot perform some functions while it is performing a transformation, so performing long transformations in *Saxon* keeps *oXygen* fully available.

Saxon can be downloaded for free here: https://www.saxonica.com/download/download_page.xml

Unlike *oXygen*, *Saxon* has no user interface; it is operated from the command line. Users unfamiliar with operating software in this way, or unfamiliar with *Saxon*, should seek guidance from other members of the team.

ROMA

ROMA is an online tool for tailor-making customising TEI schemas: <http://roma.tei-c.org/>

All changes to the schema (fnag_mss2.rng) should be discussed in advance. Any customisations that go beyond TEI conventions should only be performed if absolutely necessary and if they have been very carefully considered.

AHK

AutoHotKey (AHK) is scripting software; it allows users to design their own hotkeys to perform particular functions (not unlike macros in MS Office). This is useful while transcribing in TEI/XML, as common words, with their mark-up, can be stored under a hotkey. This means that the headword, dictionary URL, and so forth do not need to be looked up for each occurrence of the word.

For example, the following command saved under “faclair.ahk” in AHK (Figure 21),

Figure 21

```
::%dul::  
    Send, <w lemma="dul" lemmaRef="http://www.dil.ie/19189" ana="vnoun">dul  
Return
```

means that, when “%dul” + return is keyed in in *oXygen* (or anywhere else) while the “faclair.ahk” script is running, then the following will be inserted (Figure 22):

Figure 22

```
<w lemma="dul" lemmaRef="http://www.dil.ie/19189" ana="vnoun">dul</w>
```

If entered in *oXygen*, the end-tag (“</w>”) will be added automatically (hence it is omitted in AHK). The actual form may well need to be edited for context.

An existing list of hot-keys for common words and TEI elements has been compiled by the first Manuscript Research Fellow (EPT), which others are welcome to use. It might make sense for transcriber to adapt their own list to the materials they are working with or to the TEI/XML constructions which prove time-consuming.

AHK can be downloaded and used for free: <https://autohotkey.com/>

Notepad++

Offering a less intense and less complex, although obviously also less functional alternative to *oXygen*, *Notepad++* is a text editor that can be used to work with XML, HTML etc. It provides basic syntax highlighting and error spotting. *oXygen* is known to slow down when handling very large files and *Notepad++* may be a good back-up for such circumstances.

It is free to download and use: <https://notepad-plus-plus.org/download/v7.5.7.html>

Zotero

Zotero is prominent in academic contexts as a bibliography management and cite-as-you-write package. It can also combine with a browser plug-in to enable bibliographical entries to be added from library catalogues and other databases. This is useful in any scholarly endeavour. Its particular relevance in the context of the FnaG MSS Project is that *Zotero* can also be used to export bibliographies as TEI/XML files. It can thus be used to compile a bibliography relating to a particular transcription. This can then be turned into TEI and simply pasted in to the master bibliography in corpus.xml, with only minimal editing of the TEI required (see CHAPTER 5).

Zotero is free to download and use: <https://www.zotero.org/>

GitHub

GitHub is a version control, collaboration, and back-up tool that is widely used in the software development industry. It allows multiple developers to work on the same code at the same time, helps each developer track the others’ work, and manages the process of merging the various changes. Also, it is possible to use *GitHub* to wind back time and reload any earlier version of the code, which can

be invaluable for identifying or circumnavigating bugs. There is even a facility to alter a document's history to avoid merge conflicts.⁶

This level of control and complexity naturally means that *GitHub* can be challenging.

GitHub can be used online for free (although with all files accessible to the public): <https://github.com/>. The FnaG MSS Project pays for use of a private repository, to which all team members can have access.

Use of *GitHub* can be made slightly easier via use of a GUI (graphical user interface). Several are available but this user documentation assumed that *GitHub Desktop* is being used. It is also free to use and download: <https://desktop.github.com/>

Dropbox

Dropbox is a very well-known and highly user-friendly cloud storage and collaboration tool. Files can be uploaded to *Dropbox*'s servers from a computer and then accessed online (password protected) from any location. A *Dropbox* folder can also be created on a computer and it will automatically sync with the online storage. Since *GitHub* does not update the date stored online automatically, and since, to view each other's work, team members would need to manually switch *GitHub* branches (see below), a shared *Dropbox* folder is used on the project to share information to which multiple team members need simultaneous access.

Dropbox is free to use and download: <https://www.dropbox.com>

MS Office

As well as being useful for ancillary activities like producing reports relating to the Project, MS Office is of specific relevance in a couple of ways. HTML tables, generated either directly from the underlying TEI/XML or by JavaScript functions from an HTML edition, can be opened in MS Excel and searched/filtered/sorted as required. These tables are largely static when generated as HTML, so this can be invaluable when analysing the texts. Also, the comment feature in the HTML files is homemade and experimental, and MS Word's comment feature may well end up being used as a fallback. HTML editions can be opened as MS Word documents, although with the loss of a large proportion of the added functionality.

⁶ I (EPT) would like to note that I am indebted to Dr Christopher Yocum for his expert training in the use of *GitHub* in the context of the *IrishGen* project.

Further Resources

While this User Documentation aims to provide all the guidance needed to operate within the FnaG MSS Project, it does not pretend to provide full training in all of the languages and technologies involved and further study will be required to fully understand many aspects. The online resources listed below have proved useful to this end on the Project thus far.

The TEI Guidelines

The TEI Guidelines, maintained by TEI-C (the TEI Consortium), is the ultimate authority on the conventions of TEI. The document also contains a wealth of examples, commentary, and cross-referencing, which is of considerable help in understanding how the system fits together.

A searchable version is available online: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

A TEI element can be inputted into the “Search” field at the top-left to retrieve the section for that element. Within that section, hyperlinks lead to the sections on related elements.

TEI By Example

This is a series of tutorials on TEI designed for scholars in the humanities at absolute beginner level in terms of coding. It includes in-depth, interesting examples, tutorials, and close cross-referencing to the *TEI Guidelines*.

<http://teibyexample.org/>

Note that some modules are of much more relevant to the Project than others. There are, unfortunately, no plays in early Gaelic manuscripts, so the *Drama* module is less relevant. The most relevant are:

- *Introduction*
- *Common Structure and Elements*
- *The TEI Header*
- *Prose*
- *Poetry*
- *Primary Sources*

Note also that, while *TEI By Example* (and the *TEI Guidelines*) usually provide a wealth of options for what to encode from textual sources and how to encode it, choices have already been made on most of these matters in the course of the first year of the FnaG MSS Project and the whole system designed accordingly. Improvements and expansions can still be made but simply making them at the TEI/XML level, without also adapting the XSLT stylesheets and other components, will mean they at best have no effect and at worst corrupt the data. Innovations in the TEI should thus be discussed by all parties involved and a coordinated approach should be adopted.

W3 Schools

This is a series of free online tutorials and reference materials on programming and mark-up languages (including XML, XSLT, HTML, and JavaScript; **but not** TEI in particular). These are designed for the beginner and include exercises to test knowledge. They can be very useful, but they focus on key concepts and techniques rather than on running an entire coding project.

<https://www.w3schools.com/>

Stack Overflow

Stack Overflow is a forum/community where developers discuss problems in coding. The objective is that the developer raising the problem has their issue solved and that, in so doing, a resource is created to help those with similar problems in the future.

<https://stackoverflow.com/>

It can be very useful for reference when devising new code or when troubleshooting, although these activities should always be conducted in coordination with the rest of the team.

Occasionally, questions arising from the Project have been submitted and resolved on *Stack Overflow*:

<https://stackoverflow.com/users/9022913/eystein-thanisch>

If this is being considered, bear in mind that the *Stack Overflow* community has strict rules about what constitutes a valid question: <https://stackoverflow.com/help/on-topic>

Chapter 1

Setting Up a New Transcription File

“bhiodh iad mar Marina
Tsvetaeva a’leum o gheug
gu geug mar mo

chànan

a bha mas fhìor ionlan
le na bàird a’ cothromachadh
an t-saoghal cho grinn sgiobalta”

- Aonghas Phàdraig Caimbeul, ‘Marina Tsvetaeva’.

Introduction

We shall begin by going through the process of adding a new transcription file to the corpus. By way of reminder, a transcription file corresponds to a manuscript as currently bound. This is obviously essential for contributing material from new manuscripts. However, it should also help with orientation within transcription files currently in the corpus.

Topics covered in this section are as follows:

- Pre-requisites (*GitHub* and *oXygen*);
- Adding a new transcription file;
- Adding essential data to the transcription file header;
- Adding data to the corpus file.

As a running example, a manuscript not yet in the corpus will be used: Edinburgh, NLS Adv. 72.1.4. This is the 17th manuscript to be added to the corpus.

Prerequisites

Cloning the GitHub Repo

You will need to work on the FnaG MSS Project *GitHub* repository (“repo”). This is so that you can begin working on the latest version of the corpus and that your own work can be easily tracked and merged with the corpus in due course. You will need to create your own “branch” from the corpus; this is your own version of the files which you can work on and later merge back into the master branch.

If you do not already have a branch you can use, then create one via the following steps,

1. Log into *GitHub*;⁷
2. Go to <https://github.com/conmaol/faclair-manuscripts>;
3. Click on “Branch: master”;
4. Enter a distinctive name for your new branch (e.g. “NLS72134”⁸);
5. Click “Create branch”.

Figure 23

A repository for the Faclair na Gàidhlig manuscript transcription project

4 commits	1 branch	0 releases	1 contributor		
Branch: master	New pull request	Create new file	Upload files	Find file	Clone or download
Switch branches/tags	maol/EPT	Latest commit c6530b9 3 hours ago			
NLS7214	... Corrections and re-arrangement	3 hours ago			
Branches Tags	Corrections and re-arrangement	3 hours ago			
Create branch: NLS7214 from 'master'	Corrections and re-arrangement	3 hours ago			
Transcriptions	Corrections and re-arrangement	3 hours ago			
corpus_files	Corrections and re-arrangement	3 hours ago			
README.md	Initial commit	10 months ago			
corpus.html	Corrections and re-arrangement	3 hours ago			
corpus.xml	Corrections and re-arrangement	3 hours ago			
faclair-manuscripts.xpr	Corrections and re-arrangement	3 hours ago			
hwData.xml	Files from EPT's fork repo	3 days ago			

You have now created your own version (“branch”) of the repo and thus your own version of the corpus. You can make your own changes and additions to this. Your work will be backed up on *GitHub*

⁷ Details on creating an account and gaining access to the project’s private rep will be circulated separately.

⁸ The name will depend on how the branch will be used. Branches can be created for a particular piece of work, on completion of which they are deleted, or they can be used long-term and used for multiple contributions by merging and diverging repeatedly viz-a-vis the master branch. In this instance, we are imagining that this branch is being created specifically so as to work on Edinburgh, NLS Adv. MS 72.1.4.

and a comparison between your branch and the master branch will be maintained so that they can be merged again later and your work included in the master branch.

It is possible to work with the repo online. However, in order to use *oXygen* and for general convenience, we need to “clone” and download the repo and then connect the downloaded clone on your machine to *GitHub*.

1. Ensure that *GitHub Desktop* is successfully installed on your machine;⁹
2. When you click “Create branch”, you will be taken automatically to the new *GitHub* page for this branch (Figure 24).

Figure 24

A screenshot of a GitHub repository page. At the top, it shows the repository name 'conmaol / faclair-manuscripts' (Private), with options to Unwatch (3), Star (0), Fork (1). Below the header, there are tabs for Code, Issues (0), Pull requests (0), Projects (0), Wiki, and Insights. The main content area is titled 'A repository for the Faclair na Gàidhlig manuscript transcription project'. It displays statistics: 4 commits, 2 branches, 0 releases, and 1 contributor. A dropdown menu labeled 'Branch: NLS7214' is open, and a 'Clone or download' button is highlighted with a green box. Below this, a message says 'This branch is even with master.' To the right are links for 'Pull request' and 'Compare'. The commit history table lists the following entries:

File	Message	Time
JS_Files	Corrections and re-arrangement	3 hours ago
Schemas	Corrections and re-arrangement	3 hours ago
Stylesheets	Corrections and re-arrangement	3 hours ago
Training Materials	Corrections and re-arrangement	3 hours ago
Transcriptions	Corrections and re-arrangement	3 hours ago
corpus_files	Files from EPT's fork repo	3 days ago
README.md	Initial commit	10 months ago
corpus.html	Corrections and re-arrangement	3 hours ago
corpus.xml	Corrections and re-arrangement	3 hours ago
faclair-manuscripts.xpr	Corrections and re-arrangement	3 hours ago
hwData.xml	Files from EPT's fork repo	3 days ago

On this page (Figure 24), click “Clone or download”;

3. In the dialog box presented (Figure 25), click “Open in Desktop”;

⁹ *GitHub Desktop* is a GUI that allows user-friendly, offline interactions with *GitHub* repos. There are various other GUIs available that perform similar functions but these instructions assume you are using *GitHub Desktop*.

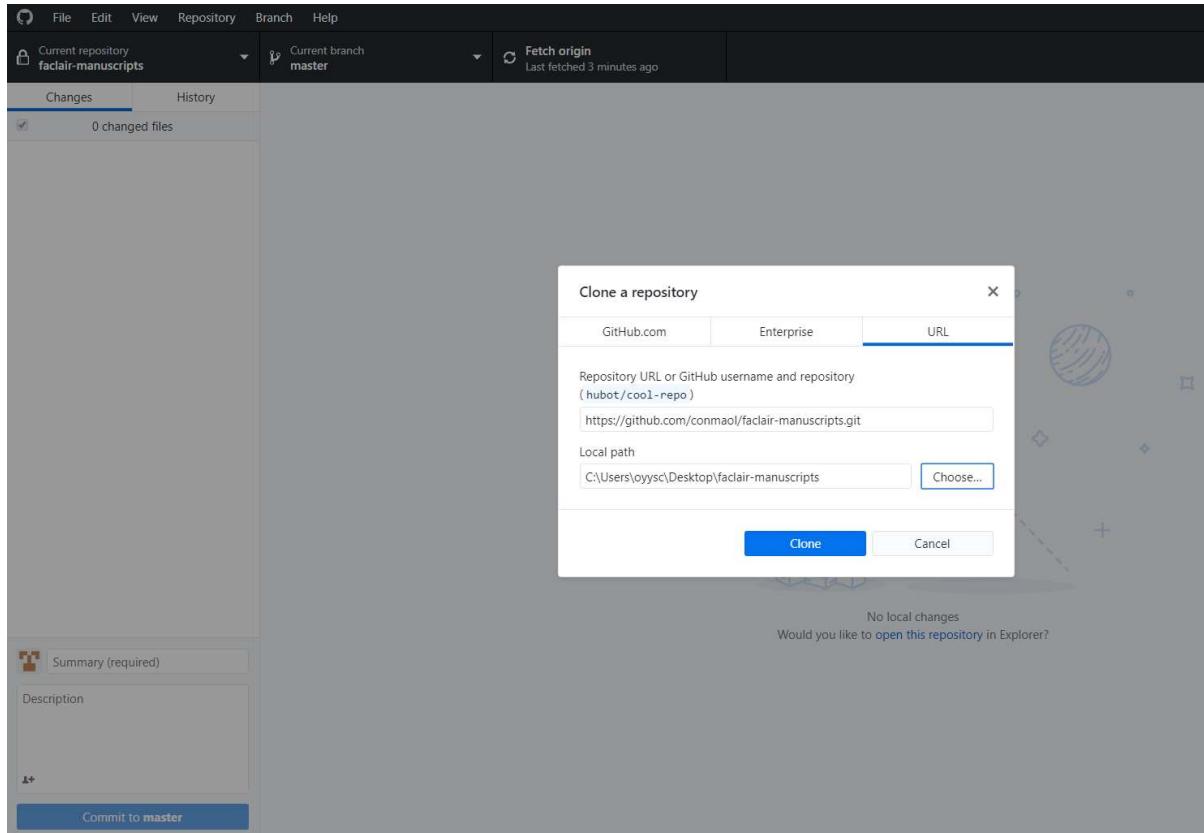
Figure 25

The screenshot shows a GitHub repository page for 'conmaol / facclair-manuscripts'. The repository is private, has 3 issues, 0 pull requests, 0 projects, 0 wiki pages, and 1 contributor. It contains 4 commits, 2 branches, 0 releases, and 1 contributor. The branch is set to 'NLS7214'. A pull request from 'Eystein-Thanisch' titled 'Merge pull request #2 from conmaol/EPT' is visible. The repository contains files like JS_Files, Schemas, Stylesheets, Training Materials, Transcriptions, corpus_files, README.md, corpus.html, corpus.xml, facclair-manuscripts.xpr, and hwData.xml. A modal window is open for cloning the repository via HTTPS or SSH, with the HTTPS URL highlighted.

File	Description	Last Commit
JS_Files	Corrections and re-arrangement	3 hours ago
Schemas	Corrections and re-arrangement	3 hours ago
Stylesheets	Corrections and re-arrangement	3 days ago
Training Materials	Corrections and re-arrangement	10 months ago
Transcriptions	Corrections and re-arrangement	3 hours ago
corpus_files	Files from EPT's fork repo	3 days ago
README.md	Initial commit	
corpus.html	Corrections and re-arrangement	3 hours ago
corpus.xml	Corrections and re-arrangement	3 hours ago
facclair-manuscripts.xpr	Corrections and re-arrangement	3 hours ago
hwData.xml	Files from EPT's fork repo	3 days ago

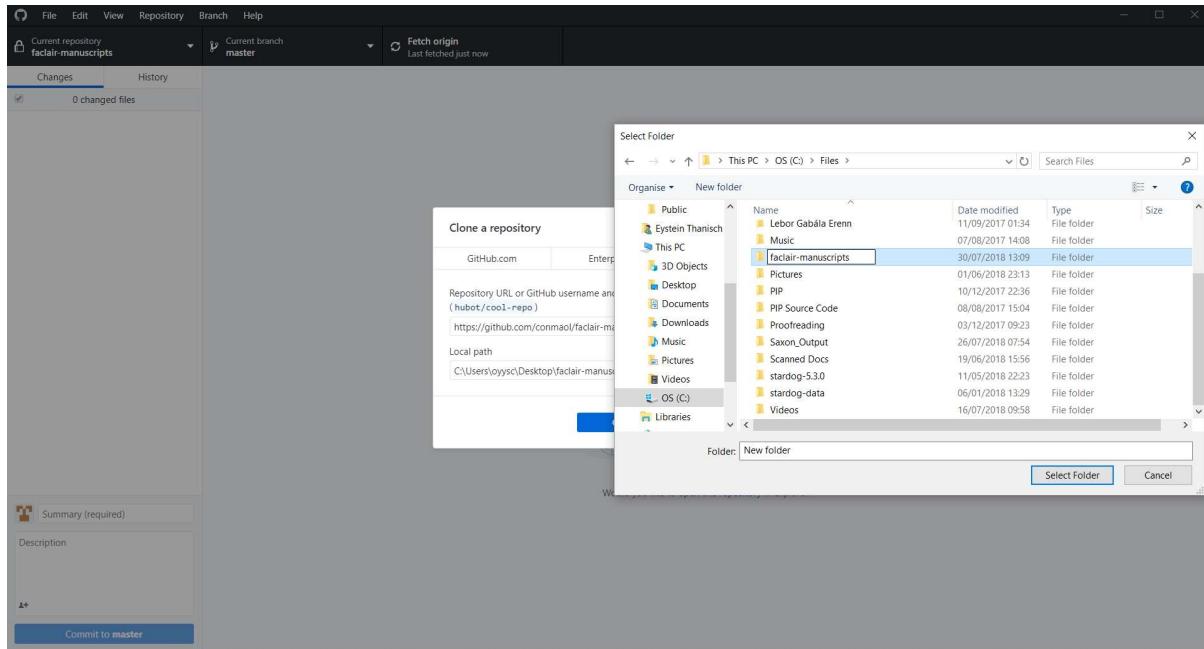
4. *GitHub Desktop* will open and you will be presented with a dialog box containing two fields (Figure 26): a “Repository URL” (displaying the site you just came from) and the “Local path” (either empty or pointing to some default location on your machine);

Figure 26



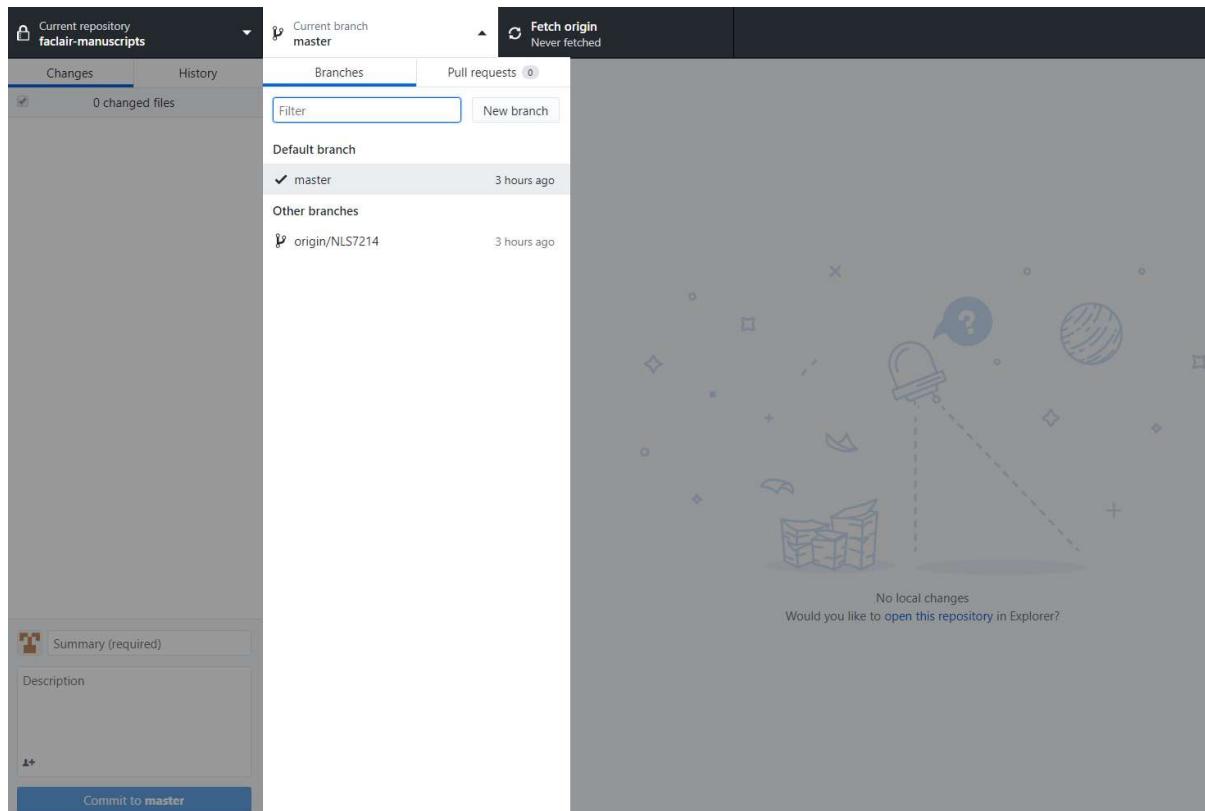
- In the “Local path” field, enter the folder location on your machine where you would like your FnAG *GitHub* repo folder (“faclair-manuscripts”) to be stored and accessed (Figure 27) – you will be visiting this folder very frequently so locate it accordingly. The repo folder will be created in the location you specify, so you do not need to add a new dedicated folder for it;

Figure 27



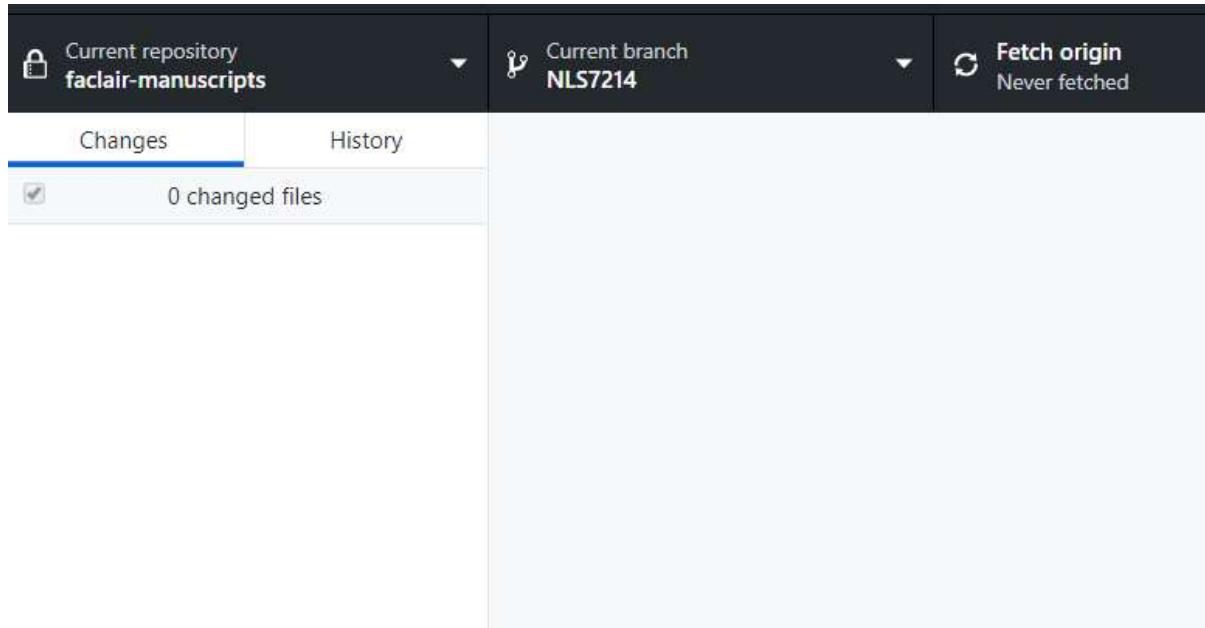
6. Click “Clone” in the dialog box and the repo folder can now be accessed wherever you located it;
7. *GitHub Desktop*, by default, will select the master branch of the repo; to select the branch you created, click “Current branch: master” in the toolbar and select the branch you created from the list (Figure 28; there may be others alongside your new branch and master, so be careful);

Figure 28



8. You can check which branch you have selected via the “Current branch” tab (Figure 29).

Figure 29



If you move between branches on the *GitHub Desktop* GUI, the contents of the repo folder will change. This is not a cause for alarm. This folder updates its contents to those of the branch you have selected, without any loss of content. Its contents will revert when you select the branch you were looking at initially. However, it is **very important** that you monitor which branch you have selected via *GitHub Desktop*, as that is the branch on which you will work if you then open any files from the repo folder.

While the master branch can be accessed via *GitHub Desktop* and its version of the repo displayed in your repo folder, it is adviseable for the purposes of tracking and managing the project for all work to be carried out on branches of master rather than on master itself. Therefore, do not do any work with the master folder selected in *GitHub Desktop*.

When you have added to or changed any files in the repo folder (in Figure 30, a file called "test.txt", has been added to the repo folder), return to *GitHub Desktop* and ensure you are viewing your branch. You will see the change you have made registered in the left-hand sidebar. Clicking on it will display the content involved in the change (e.g. Figure 31).

Figure 30

The screenshot shows a Windows File Explorer window with the following details:

Path: This PC > OS (C:) > Files > faclair-manuscripts >

File List:

Name	Date modified	Type	Size
corpus_files	30/07/2018 13:10	File folder	
JS_Files	30/07/2018 13:10	File folder	
Schemas	30/07/2018 13:10	File folder	
Stylesheets	30/07/2018 13:10	File folder	
Training Materials	30/07/2018 13:10	File folder	
Transcriptions	30/07/2018 13:10	File folder	
corpus.html	30/07/2018 13:10	Chrome HTML Do...	55 KB
corpus.xml	30/07/2018 13:10	XML Document	53 KB
faclair-manuscripts.xpr	30/07/2018 13:10	oXygen XML proje	1 KB
hwData.xml	30/07/2018 13:10	XML Document	505 KB
README.md	30/07/2018 13:10	Markdown file	1 KB
test.txt	30/07/2018 13:43	Text Document	1 KB

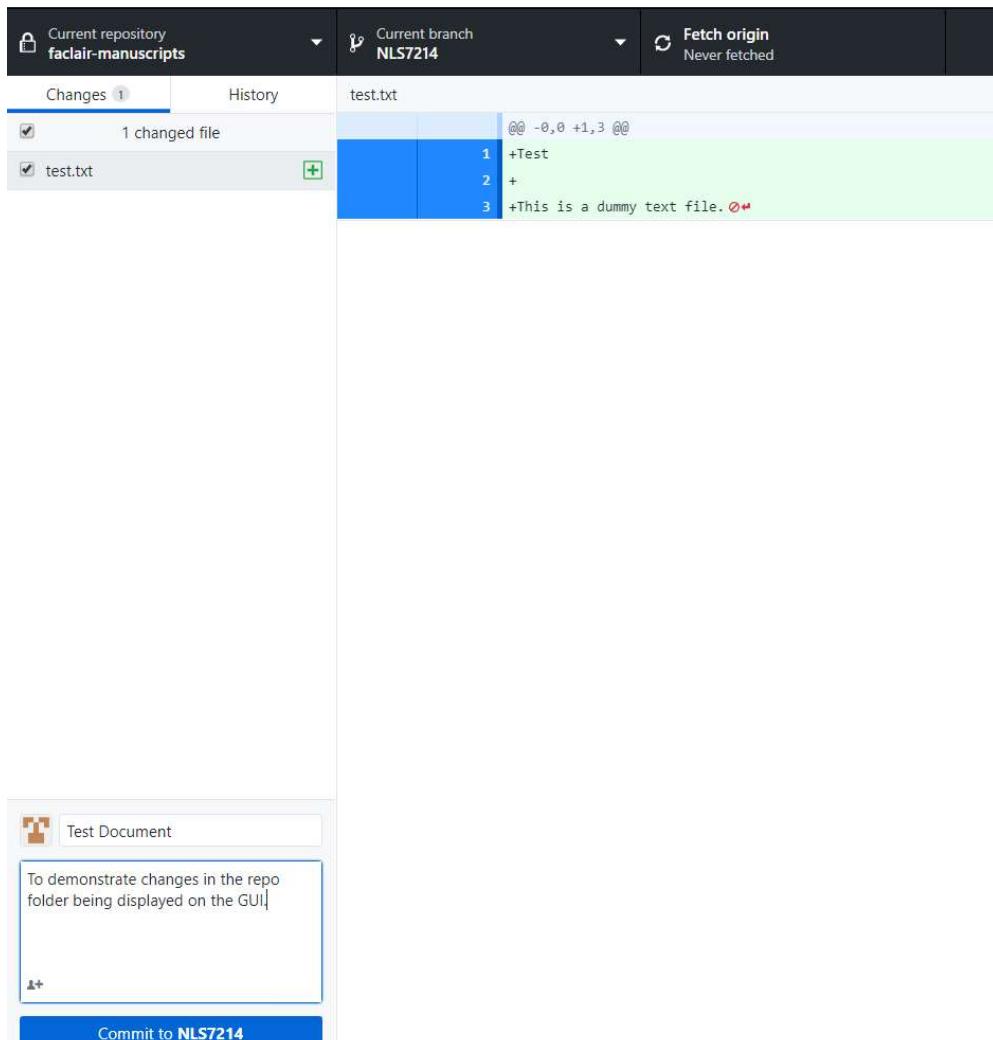
Left sidebar (Navigation pane):

- Dropbox
- OneDrive
- Documents
- Email attachments
- Music
- Pictures
- Public
- Eystein Thanisch
- This PC
- 3D Objects
- Desktop
- Documents
- Downloads
- iCloud Photos
- Music
- Pictures
- Videos
- OS (C:)
- Libraries
- Network
- Control Panel
- Recycle Bin
- Data
- TESTING
- XML

Bottom status bar:

12 items 1 item selected 25 bytes

Figure 31



To “save” your change on *GitHub*, click “Commit to [current branch]” at the bottom-left (adding a note on the changes you are saving), then “Push origin” at the upper right. Your changes are now saved to the online (“remote”) version of this branch (NLS7214; but **not** to master). Each of these saves (“commits”) represents a waypoint. The repo as it was at that waypoint can be reloaded at any time in the future, if required. It is thus adviseable to commit and push to origin regularly, as this creates more options for rolling back time when bug-fixing.

None of the changes you make in the repo folder will be visible outside of your machine until you commit and push to origin. Nonetheless, make sure that any work you do that is meant as a direct contribution to the corpus is done in the repo folder (on the appropriate branch); it will not be tracked or backed up by *GitHub* otherwise. Conversely, it would be safest to keep any work or changes that you do not wish (necessarily) to become part of the corpus and be viewed by the rest of the team out of this folder.

Merging your work with the master branch on *GitHub* and updating your own from master can be a complicated process requiring caution and communication. A proposed operating procedure is set out in CHAPTER 9. The procedure thus far sets you up with a clone of the corpus that is being tracked by *GitHub* and that is sufficient for you to start work.

The Repo Folder

The folder that you clone from *GitHub* contains all the transcription files, XML databases, schemas, stylesheets, and JavaScript files involved in the project (see Figure 7). Most of these files are located in sub-folders with self-explanatory names (e.g. “Stylesheets”).

Otherwise, the two central XML databases (`corpus.xml` and `hwData.xml`) are located in the top-level repo folder. Alongside them are the *oXygen* project file (see below), the `README` file (currently empty), and `corpus.html`. This is the file that contains the output from a transformation (i.e. when a stylesheet is applied to `corpus.xml`). It is overwritten everytime a new tranformation is run, so it will contain whatever the latest output has been at the time the repo was cloned or the latest output you generated yourself by running a transformation.

For this reason, if you are at all likely to want to keep your output to use again, you should copy and paste `corpus.html` into another folder on your machine, as it will be very easy for you to overrite the file in your repo folder by mistake.

Creating an *oXygen* Project

The files in the *GitHub* repo folder can be opened in a variety of text editors, of which *oXygen* is the most useful (and use of it is assumed in this documentation). To get the most out of *oXygen*'s management features, we need to create a project for the FnaG MSS corpus in *oXygen XML Editor*.

To do this,

1. Open *oXygen XML Editor*;
2. From the “Project” menu, select “New Project”;
3. File Explorer will open: it is asking you to save a file, “`newProject.xpr`”;
4. Find your FnaG *GitHub* repo folder and open it in the File Explorer window;
5. Give the file an appropriate name (it might be best to give it the same name as your repo folder), retaining the `.xpr` extension;
6. Save it to your *GitHub* repo folder.

Opening this project in *oXygen* will allow searches across all the files in the repo and allow easy navigation between them. Any of the files in the repo can also be opened individually.

Adding a New Transcription File

For the header and high-level body elements in a transcription file, a standard structure must be used. Hence it makes sense to use a template rather than constructing each transcription file from scratch.

To set up the templates on your version of *oXygen XML Editor* (this only needs to be done once),

1. In the “Options” menu, click “Preferences”;
2. In the “Preferences” window, go to “Editor”, then “Templates”, then “Document Templates”;
3. Click “New” at the bottom of the window to add a new folder for templates;
4. Find your repo and select the “Training Materials” folder;
5. Click “Apply” and close the window.

Next, to create a new transcription file,

1. In the “File” menu, click “New”, or click the icon in the toolbar;
2. In the window that opens, find the “Training Materials” folder;
3. In “Training Materials”, select “transcriptionTemplate.xml”;
4. Place the new file in the “Transcriptions” folder rather than in “Training Materials”;
5. Replace “untitled.xml” at the end of the new file’s filepath with a name in the form “transcriptionn.xml” (in our example, transcription17.xml).

This will create a new TEI/XML file called “transcription17.xml” in the “Transcriptions” folder with the required common elements already in place.

Note that empty elements (e.g. <p/>) in the template will not cause a problem for the processing software. You do not therefore need to fill in every element of the template immediately and can leave elements empty. At worst, this will cause some blank spaces, stray punctuation, or headers without paragraphs in any draft output generated but this is all perfectly manageable.

For plain and annotated copies of the transcription file template, see Appendix 5.

Registering the New Transcription File

Before we begin transcription, we need to add some data:

- To the transcription file header;
- To the corpus file;
- To project records on *Dropbox*.

This is necessary so as:

- To ensure that the file will pass basic XML validation by the processing software;
- To obtain the unique ids needed for marking up the transcription proper;
- To ensure that draft editions generated from the transcription are formatted intelligibly;
- To alert other members of the team to this transcription's existence and the fact that the unique ids selected are now in use.¹⁰

Essential Header Data

First, in the new xml file (in our example, transcription17.xml), various pieces of data need to be added to the header (line references are to Appendix 5a¹¹).

- line 4: In the file's <TEI> element, change "N" in the @xml:id value to the transcription number (17, in our example; so, "T17");
- line 8: In the <title> element, change "N" in the text content to the transcription number (17, in our example; so, <title>Transcription 17</title>);
- line 12: Add another <respStmt> element and children and fill in your own details, with the <resp> transcriber;¹²
- line 20: In <idno>, add the value of @xml:id from <TEI> as text content (i.e. "T17");
- line 29: In <msIdentifier>, change "N" in the value of @sameAs ("MSN") to the transcription number (i.e. "MS17").
- line 30: In <settlement>, add the town/city where the manuscript is located as text content.
- line 31: In <repository>, add the name of the library where the manuscript is located as text content.

¹⁰ GitHub will not provide alerts if the same unique id is used on different branches. GitHub can be used to view work on other branches, as long as it has been committed and pushed, but this is cumbersome and must be done proactively. For this reason, records of assigned work and use of unique ids are kept on *Dropbox*.

¹¹ Elements may change line in the xml transcription file as data is added; hence reference is made to the blank copy.

¹² Unless told otherwise, your three-letter id will be the first letter of your first name, then the first and second letter of your surname (English); e.g. Martina Maher = MMA.

- line 32: In `<idno type="shelfmark">`¹³, add the manuscript's shelfmark as text content.
- line 165: In `<change when="yyyy-mm-dd" who="XXX"/>`,
 - Add today's date as the value of `@when` (format as shown above);
 - Add your id as the value of `@who`;
 - Describe the change you have made as the text content of `<change>` (e.g. "Created transcription file");
- Save the file.

Essential Corpus Data

Next, open `corpus.xml` in the repo folder. This is the central corpus file, holding data of potential relevance to all transcription files. A guide to this file can be found in Appendix 6.

Add the following data (line references are to Appendix 6a¹⁴):

- line 24: Add another `<msDesc>` plus descendants within `<listBibl type="mss">` and fill in the details for the new manuscript (as above, except here `@xml:id` is used instead of `@sameAs`);
- line 169: Add another `<change>` element, filling in the details as above, except, in the text content, note which transcription you have added;
- line 172: Add another `<xi:include href="example.xml" parse="xml">` element; change the value of `@href` from "example.xml" to the filepath to the new transcription file (e.g. "Transcriptions/transcription17.xml");
- Save the file.

Updating the Project Record

Finally, we need to update the universally accessible project records to inform other team members that you are working on this manuscript:

1. Leaving the repo folder, open the shared "FnaG MSS Corpus" *Dropbox* folder;
2. Open the "Project" sub-folder;
3. Open "xmlid_lists.xls";
4. In the "MSS" sheet, add the MS number of the new transcription, a short version of the manuscript's shelfmark, and your initials in the respective columns.

¹³ Note that, in TEI, the same element can be re-used in different contexts. In this case, `<idno>` is specifically the id number for the manuscript, whereas, as child of `<publicationStmt>`, it is the id number for the TEI file.

¹⁴ Changes should be made in the actual file, `corpus.xml`, not in the blank version in Appendix 6, which is to be used for information only.

Review

The new transcription file has now been added to the corpus, as well as identified and registered in terms of which manuscript it represents. Much more metadata needs to be added to the file header and to the corpus file but this can be done at later stages (see CHAPTERS 2 and 5). The next step is to add the frame for the first text for transcription to this transcription file, as well as data associated to the transcription and corpus files (see CHAPTER 2).

Chapter 2

Identifying and Structuring Different Types of Text

“Cid dianad toisreach seo? Ni hanna.”

- *Auraicept na n-Éces*
(ed. G. Calder)

Introduction

This chapter looks at how to get the appropriate template set up for a text within a transcription file. It also details the data relating to that text that needs to be added to the transcription file header, to the corpus file, and to the project records.

How the text is structured, and thus its outer elements within in the transcription file <[body](#)>, will depend on its form. There are three forms, according to the project's current analysis (others can be devised):

- Prose;
- Verse;
- Divided Prose.¹⁵

¹⁵ "Divided prose" is prose that falls naturally into divisions of at most a couple of sentences, with no direct connection between these divisions. Each division is thus most clearly presented as self-contained when the editorial text is generated. Examples currently in the corpus include a collection of medical aphorisms (T5.2; the Gaelic *Colliget*) and *Liber Scintillarum* (T4.4), a thematically grouped collection of extracts from the Bible and the Church Fathers.

Examples

Sub-divisions of a text can be of a different form from the frame next (e.g. a poem embedded in prose). We will be using four examples (including NLS Adv. 72.1.4) in this chapter and subsequently to consider how to mark up texts of each form and texts containing material in multiple forms.

All manuscripts involved are available via [ISOS](#), as are catalogue entries.

Prose

Bruighion Bheg na hAlmhain; NLS Adv. 72.1.36 (T12), fols 96^r–106^r.

Scribe: Eoghan Mac Gilleoin (17th-century).

A marked up sample of prose is presented in APPENDIX 6.

Verse

‘Is maith mo leaba, is olc mo shuain’; NLS Adv. 72.1.36 (T12), fols 114^r1–114^v9.

Scribe: Eoghan Mac Gilleoin (17th-century).

A marked up sample of verse is presented in APPENDIX 7.

Divided Prose

Definitions; NLS Adv. 72.1.4 (T17), fol.24^r.

Scribe: anon. (16th-century).

A marked up sample of divided prose is presented in APPENDIX 8.

Multiple Forms

Cath Fionntrágha; NLS Adv. 72.2.11 (T3).

Scribe: Alasdair mac Mhaighstir Alasdair (18th-century).

Adding a New Text

Certain tasks need to be performed when adding a new text to a transcription file regardless of its form. We will cover these before proceeding to examine what is specific to each form. We will use the “Definitions” on fol. 24^r of NLS Adv. 72.1.4 as our initial example.

Corpus Data

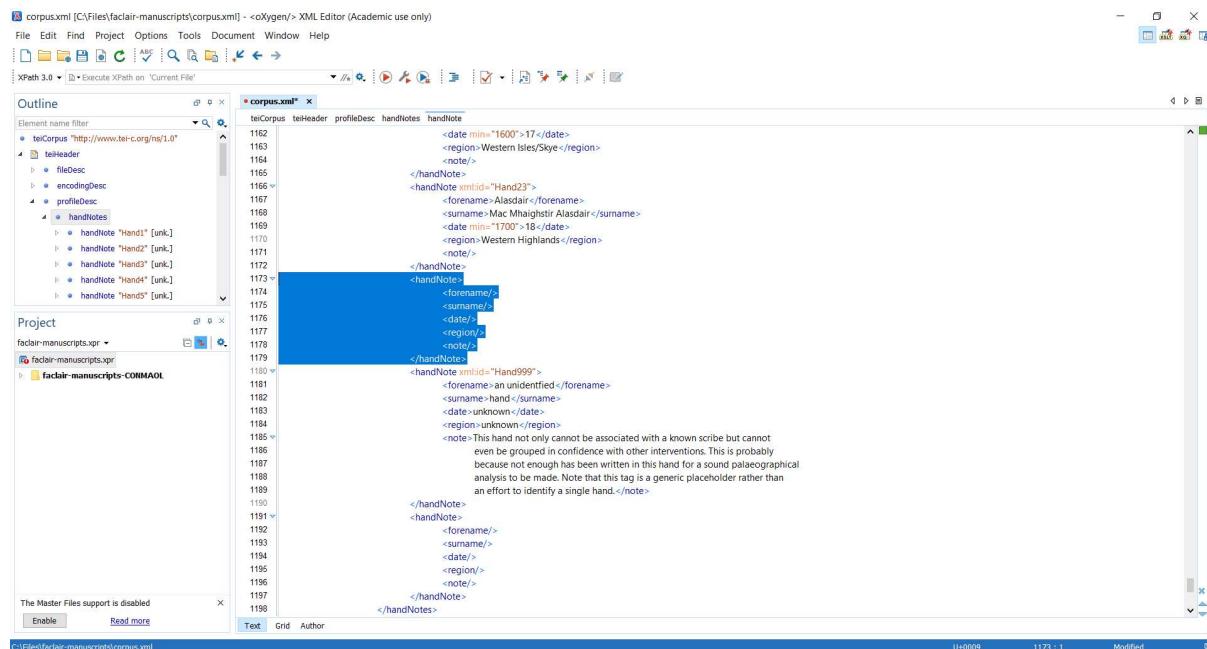
First, identify the main scribe of the text, whether from a catalogue entry or from your own research. In our example, the scribe (according to Black 2011) has not been identified.

In corpus.xml, go to <handNotes> (Appendix 6a l.156). Check whether this hand already has a record, using the catalogue entry or your own research to identify whether it appears in any other manuscripts from which it might have been transcribed. Our example text is in NLS Adv. 72.1.4’s Hand 2, which Black identifies with Hand 5 of NLS Adv. 72.1.27. This manuscript is not yet included in the corpus so it is safe to assume this is the hand’s first appearance.

If the hand were already recorded in corpus.xml, note the value of the @xml:id attribute of the corresponding <handNote> element.

In this case, however, copy the empty <handNote> template element plus all descendants and paste it in within <handNotes>, after the most recent identified hand and before the entry for Hand999.

Figure 32



Assign “Hand” plus the next sequential hand number to it as the value of @xml:id. In our example, this would be “Hand24”. The hand is anonymous and the dates of the hand are also thus uncertain; however, what we are calling Hand24 begins a text at fol. 25^r6, after a note in the hand of Níall Óg mac Néill mhic Bheathadh. As the latter has a floruit in 1541 (Bannerman 1998: 137), we can assume provisionally that Hand24 belongs to the 16th century. The Beaton hands that appear in the manuscript are belong to individuals associated with Kilelane, Islay, but we cannot be sure that this is where the other hands conducted their activity.

Our initial record of this hand might thus look as follows (Figure 33):

Figure 33

```
<handNote xml:id="Hand24">
    <forename>unk.</forename>
    <surname>unk.</surname>
    <date type="floruit" min="1500" max="1600">16</date>
    <region cert="medium">Islay</region>
    <note>
        <p>Hand 2 in NLS Adv. 72.1.4 and Hand 5 in NLS Adv. 72.1.5.</p>
    </note>
</handNote>
```

In time, a full description of this hand's traits can be written up under the `<note>` element, based on encounters with the hand across the project. For now, the briefest cross references are provided.

Transcription File Data

Next, we create an entry for the new text in the transcription file header. Under `<msContents>` (Appendix 5a, I.34), make a copy of the template for `<msItem>` plus all descendants and paste it in, also as a child of `<msContents>`.

Within this `<msItem>` element, add the details about the text that you have available. This will vary but the following are recommended (line references to Appendix 5b, where the elements are annotated with more information on what is required):

- `<msItem n="itemN" xml:id="itemID" resp="HandXXX">` (I.117):
 - @n is a sequential number, indicating this text's position of the array of texts transcribed from this manuscript, in the order in which they appear in the manuscript. This often requires some forethought as to which texts from the manuscript are likely to be transcribed.¹⁶ In this case, however, the Definitions on fol. 24^r of NLS Adv. 72.1.4 comprise the first Gaelic text in the manuscript. So @n can safely = "1";
 - @xml:id is the manuscript's id, plus @n. So @xml:id = "MS17.1", in this case;
 - @resp indicates the main scribe responsible for the text; as per the `<handNote>` we just added, @resp = "Hand24";
- `<locus>` (I.39): fol. 24^r;
- `<incipit>` (I.40): "Re sine";
- `<explicit>` (I.41): "co haicidach".

The initial entry in `<msItem>` should thus resemble the following (Figure 34):

Figure 34

```
<msItem n="1" xml:id="MS17.1" resp="Hand24">
    <locus>fol. 24<hi rend="sup">r</hi></locus>
    <title>Definitions</title>
    <incipit>Re sine ...</incipit>
```

¹⁶ It is not very difficult to change the sequential number assigned to a text after it has been added to the corpus, as the text is not referenced in the XML files outside its own transcription file. It is still a situation worth avoiding where possible, however.

<code><explicit>... co haicidach</explicit></code>
<code><!-- Further elements here (can remain empty) --></code>
<code></msItem></code>

Finally, in the transcription file `<handDesc>` element (Appendix 5a, l. 123), add a `<handNote>` element, with a `@corresp` attribute with the id of the main scribe of the text as a value: so,

Figure 35

<code><handNote corresp="Hand24"></code>
<code> <note></code>
<code> <!-- Further elements here (can remain empty) --></code>
<code> </note></code>
<code></handNote></code>

Project Record Data

As a last step before we proceed to add the text proper to the transcription file, open the “Project” *Dropbox* folder, then open “xmlid_lists.xlsx”. On the “Hands” sheet, if it be new to the corpus, add details of the text’s main hand.

The New Text

The `<div>` Element

Finally, we can add the first element of the text itself. Within the transcription file’s `<body>` element (Appendix 5a l.169), at the position corresponding to the text’s position in the manuscript relative to the other texts included in the transcription file, add a `<div>` element. (Appendix 5a l.170). This is the delimiter of one text from another in the file.

The `<div>` element requires various attributes:

- `@n`: this is the sequential number for the text, as in `<msItem> @n;`
- `@type`: this is the form of the text (“prose”, “verse”, or “divProse”);
- `@resp`: the main hand responsible for the text;
- `@corresp`: = `<msItem> @xml:id;`

The `<div>` element we insert for the “Definitions” on fol. 24r of NLS Adv. 72.1.4 would be as follows:

Figure 36

<code><text></code>
<code> <body></code>
<code> <div n="1" type="divProse" resp="Hand24" corresp="MS17.1"></code>
<code> <!-- Text continues here --></code>
<code> </div></code>
<code> </body></code>
<code></text></code>

Form-specific Elements

So far, only `<div>` @type is dependent on the text's form. The elements added within `<div>`, however, will determine the basic editorial structure of the text and so are dependent on its form.

Prose

Prose text is simply contained inside `<p>` (paragraph) element, which is a child of `<div>`. One `<p>` element is used to enclose the entire text. Page (`<pb>`), column (`<cb>`), and line markers (`<lb>`) are then placed within the `<p>` element and the text proceeds from there.

Bruighion Bheg na hAlmhain, from NLS Adv. 72.1.36, would be framed as follows (Figure 37):

Figure 37

```
<text>
  <body>
    <div n="2" type="prose" resp="Hand9" corresp="MS12.2">
      <p>
        <pb n="96r" xml:id="MS12.96r"/>
        <lb n="1" xml:id="MS12.86r.1"/>
        <!-- Text continues here -->
      </p>
    </div>
  </body>
</text>
```

Verse

Verse is structured within an `<lg>` ("line group") element (in this case, a stanza, defined using @type), then a `<l>` ("line"), corresponding to the analysis of the metre applied by the transcriber.

Due to stylesheet requirements, the location of the `<pb>` and `<lb>` elements is important:

- `<pb>` must be placed outside the `<lg>` element at the start of the poem. If a page break occurs within the poem, `<pb>` can be placed wherever is appropriate, although it should never be placed between `<lg>` and `<l>`.
- `<lb>` should always be placed within the next `<l>` element.

'Is maith mo leaba, is olc mo shuain', from NLS Adv. 72.1.36, would be framed as follows (Figure 38):

Figure 38

```
<text>
  <body>
    <div n="4" type="verse" resp="Hand9" corresp="MS12.4">
      <pb n="114r" xml:id="MS12.114r"/>
      <lg n="1" type="stanza" xml:id="MS12.4.1">
        <l n="a" xml:id="MS12.4.1a">
          <lb n="1" xml:id="MS12.114r.1"/>
          <!-- Metrical line continues here -->
        </l>
        <!-- Stanza continues here -->
      </lg>
      <!-- Text continues here -->
    </div>
```

```
</body>  
</text>
```

Divided Prose

Divided Prose is structured like verse, within <lg> elements (with @type set to “proseDiv”), except that the lines (<l>) are determined not by an identifiable metre but by self-contained units of content. These may or may not also be represented in terms of manuscript page layout.

The “Definitions” from NLS Adv. 72.1.4 would be framed as follows (Figure 39):

Figure 39

```
<text>  
  <body>  
    <div n="1" type="divProse" resp="Hand24" corresp="MS17.1">  
      <pb n="24r" xml:id="MS12.24r"/>  
      <lg n="1" type="proseDiv" xml:id="MS17.1.1">  
        <l xml:id="MS12.4.1a">  
        <lb n="1" xml:id="MS12.24r.1"/>  
        <!-- Text continues here (first definition (Latin)) -->  
        </l>  
        <l xml:id="MS12.4.1b">  
        <!-- Text continues here (first definition (Gaelic)) -->  
        </l>  
      </lg>  
      <lg n="2" type="proseDiv" xml:id="MS17.1.2">  
        <l n="a" xml:id="MS12.4.2a">  
        <!-- Text continues here (second definition (Latin)) -->  
        </l>  
        <l xml:id="MS12.4.2b">  
        <!-- Text continues here (second definition (Gaelic)) -->  
        </l>  
      </lg>  
    </div>  
  </body>  
</text>
```

In the above text, line breaks continue to occur but, as they do not correspond with a new <l> element after the first, they are not represented here.

If the page was structured in columns (it is not), the column would indicated by placing the <cb> element after <pb>.

Precisely how many <lg> elements are used, containing how many <l> elements apiece, depends on the text and editorial decisions made in relation to it. In this case, each definition (<lg>) is given in Latin (<l>), then in Gaelic (<l>). The <lg> enclosing each version means that the defintion can be retrieved and presented in both languages, for the purposes of comparison. However, a set of dicta in Gaelic could potentially be grouped according to theme or manuscript page layout or suchlike.

Multiple Forms

Some texts contain material in some combinations of the three forms covered above. In such circumstances, some special arrangement of elements in both the text body and the text header is required to structure and present the transcription properly.

Arrangement of Body Elements

A `<div>` element can be placed as the child of another `<div>` element to represent a subdivision in the text (the stylesheet will represent the subdivision with a title and a line break in the edition).¹⁷ This could be used to structure a long and complex text that is in the same form throughout. However, it is **essential** to add a new, embedded `<div>` if the form changes.

For example, *Cath Fionntrágha*, in NLS Adv. 72.2.11, is a prose saga containing embedded verse would be structured as follows:

Figure 40

1	<code><text></code>
2	<code><body></code>
3	<code><div n="1" type="prose" resp="Hand23" corresp="MS3.1"></code>
4	<code><p></code>
5	<code><pb n="1" xml:id="MS3.p1"/></code>
6	<code><lb n="1" xml:id="MS3.p1.1"/></code>
7	<code><!-- Prose continues here --></code>
8	<code></p></code>
9	<code><div n="1.1" type="verse" resp="Hand23" corresp="MS3.1.1"></code>
10	<code><lg n="1" type="stanza" xml:id="MS3.1.1.1"></code>
11	<code><l n="a" xml:id="MS3.1.1.1a"></code>
12	<code><!-- Metrical line continues here --></code>
13	<code></l></code>
14	<code><!-- Stanza continues here --></code>
15	<code></lg></code>
16	<code><!-- Verse text continues here --></code>
17	<code></div></code>
18	<code><div corresp="MS3.1" type="prose" resp="Hand23"></code>
19	<code><p></code>
20	<code><!-- Prose continues here --></code>
21	<code></p></code>
22	<code><div n="1.2" type="verse" resp="Hand23" corresp="MS3.1.2"></code>
23	<code><!-- Verse text continues here --></code>
24	<code></div></code>
25	<code></div></code>
26	<code><div corresp="MS3.1" type="prose" resp="Hand23"></code>
27	<code><p></code>
28	<code><!-- Prose continues here --></code>
29	<code></p></code>
30	<code><div n="1.3" type="verse" resp="Hand23" corresp="MS3.1.3"></code>
31	<code><!-- Verse text continues here --></code>
32	<code></div></code>
33	<code></div></code>
34	<code></div></code>
35	<code></body></code>
36	<code></text></code>

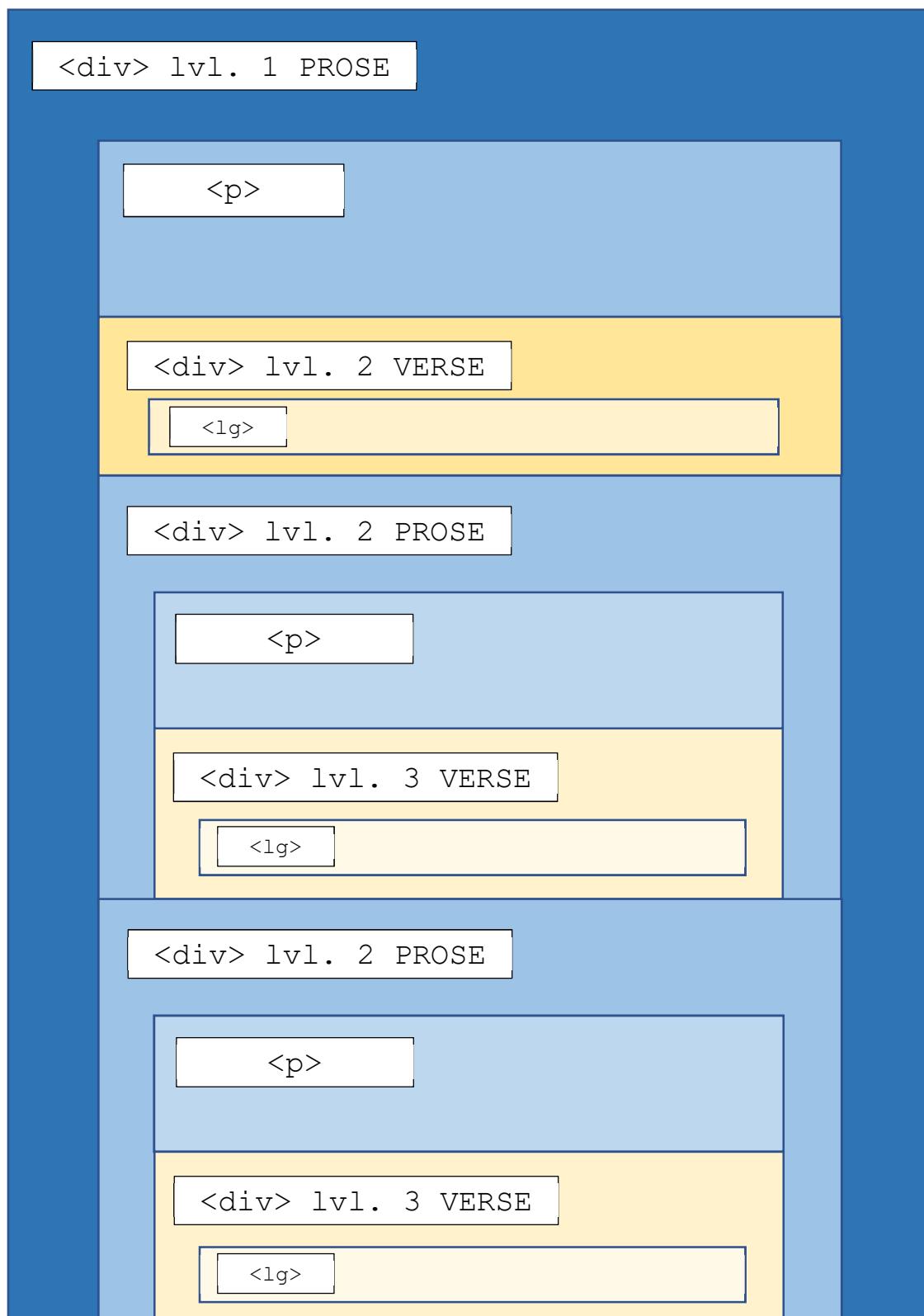
¹⁷ TEI does offer the elements `<div1>`, `<div2>`, `<div3>`, etc. to represent different levels of subdivision. However, the rules governing use of these elements proved too restrictive for the situations presented by manuscript materials so multiple generic `<div>` elements are used instead. In any case, processing software can calculate the level of the `<div>` based on its position in the XML tree.

The required structure of the multiple `<div>` elements is perhaps not intuitive. Assuming the text begins as prose, the initial prose `<div>` (line 3) cannot, within TEI conventions, be used to enclose all top-level (i.e. non-embedded) prose in the text. The initial prose `<div>` (line 3) has a `<p>` child (line 4) that encloses the prose until the first second-level `<div>` (line 9), before which the `<p>` child is closed (line 8). After the first second-level `<div>` (line 9), which contains a poem embedded in the text (e.g. ‘Triallim siar go hEirinn áin’, pp.1.25–2.3), another second-level `<div>` (line 18) is embedded with a `<p>` child (line 17), which contains the resumed prose. Its `@corresp` attribute keeps it associated with the initial top-level prose `<div>` (line 3).

Further verse appears; the next verse `<div>` (line 22) is a child of the second-level prose `<div>` (line 18). However, at the end of the verse, both `<div>` elements are closed and a new second level prose `<div>` (line 26) is added as a child of the top-level `<div>` (line 3). This can then contain the next verse `<div>`. In the tradition of *opera geminata*, the text then proceeds in pairs of prose and verse `<div>` elements, the latter a child of the former.

A simpler diagram may help (Figure 41):

Figure 41



There is nothing particular about prose and verse in this context. This model would apply whichever combination of forms were involved. That is, the first form, however little textual material is actually in it, becomes the overarching form for the text. While this might be troubling in theory, in practice

almost all stylesheet operations select material's closest `<div>` ancestor and thus capture the form at that particular point in the text.

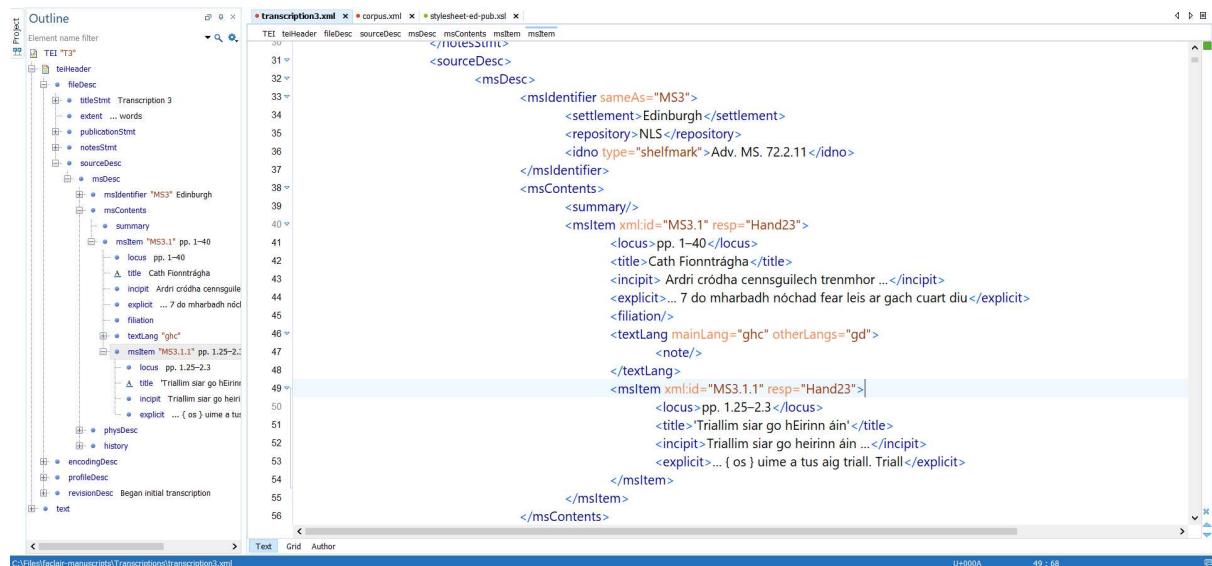
Arrangement of Header Elements

As already stated, each `<div>` child/descendant of the `<body>` element must correspond to an `<msItem>` descendant of `<teiHeader>`. Thus, for each sub-division of a text that you have represented using a `<div>` element **that represents either a deliberate sub-division or a change of form** (rather than a later continuation of the top-level form), you must add an `<msItem>` element plus descendants as a child of the `<msItem>` that represent the overall text. So, for each of the verse `<div>` elements in Figure 40, an `<msItem>` element would need to be added.

If a `<div>` is assigned a new `@n` and `@corresp` value, it should have its own `<msItem>` element in the header; if it reverts to a previous `@n` and `@corresp` value, its `<msItem>` element will already exist.

Figure 42 shows the first poem from *Cath Fionntrágha* represented as a child `<msItem>` element in the header (oXygen).

Figure 42



Revision Log

Once you have added the basic data and structure for the new text, go to the `<revisionDesc>` (Appendix 5a l.164) element in the transcription file header and add a `<change>` element (Appendix 5a l.165) noting the new text that has been added.

Review

- Three textual forms are currently recognised in the corpus: “prose”, “verse”, and “divProse”.
- Texts can consist of combinations of multiple forms.
- A text’s form is indicated by the @type attribute of its enclosing `<div>` element.
- The form determines the text’s overall TEI structure.
- A change of form is represented by embedding a `<div>` element within the text’s enclosing `<div>` element; this can be a complex operation.
- Embedded `<div>` elements within `<div>` must correspond to `<msItem>` elements in the transcription file header unless they stand for text at the top-level in the text (i.e. returning to the top-level `<div>`).

Chapter 3

Basic Transcription

“Thog iad na siùil bhreaca
Bhaidealacha dhònach,
Shìn iad na coilpeinean raga
Teanna rìghne
Ri fiodhannan àrda fada
Nan colg bìth-dhearg.

Cheangladh iad gu gramail snaompach,
Gu neo-clearbach
Troimh shùilean nan cromag iarainn
'S nan cruinn fhailbeag.

Cheartaich iad gach ball den acfhainn
Ealamh dòigheil,
'S shuidh gach fear gu freastal tapaidh
'Bhuill bu chòir dha."

- Alasdair mac Mhaighstir Alasdair, *Birlinn Chlann Raghnaill* (ed. D. Thomson).

Introduction

At long last, we are in a position to consider how to add and mark up actual text transcribed from manuscripts. This chapter covers the general strategy and principles underlying the mark-up scheme devised for the project, then sets out how to mark up straightforward, unproblematic text. Chapter 4, meanwhile, covers the representation in the mark-up of problems, uncertainty, and special manuscript features.

The scheme presented here is, of course, open to refinement and revision, and feedback on it from both a technical and a philological point of view is welcome. It should be borne in mind that the corpus already contains over 18,000 marked up words, so any changes could require time-consuming revision of the existing transcriptions, but this may, of course, be nonetheless worthwhile.

Prerequisites

This chapter assumes that you have already familiarised yourself with and acted the preceding sections of this user documentation. In particular, you will need to have set up a transcription file and updated the corpus file ready for the new text and, as per the instructions in CHAPTERS 1 and 2. You should also be on an appropriate *GitHub* branch.

You may or may not have already completed the transcription you plan on marking up. It is possible to transcribe directly into TEI and, for a transcriber conversant in the project's TEI conventions, this is certainly faster. However, while learning TEI, it may be advisable to have a transcription prepared and ready, then mark it up. Note that you will still need access to the manuscript (unless your transcription's apparatus is extraordinarily detailed), as the TEI mark-up will require details not normally included wholesale even in a scholarly edition (e.g. which abbreviations are used; scribal word division; the reason why a reading is uncertain).

The following online resources should also be accessible (internet connection thus required),

- *ISOS* (<https://www.isos.dias.ie>), or an image of the manuscript from elsewhere, or the manuscript itself;
- *eDIL*: <http://www.dil.ie> ;
- *Dwelly* (via *Faclair Beag*): <http://www.faclair.com/> ;
- *Tionscadal na Nod*: http://www.vanhamel.nl/codecs/Project:Tionscadal_na_Nod ;
- *HDSG MSS Slips* [completion due Autumn 2018; DASG staff to provide details] ;
- The file “xmlid_lists.xlsx” from the project *Dropbox* folder;
- The *AHK* script containing the hotkey library running, if in use (advised);

plus any reference works for grammar and palaeography that you find helpful.

For an example of the coding described in this chapter in action, watch [FnaG MSS- Basic Transcription Demo](#)

General Principles

Before we look at the details of the mark up, however, it is worth considering the general editorial principles that animate the TEI strategy that has been devised for the project.

Focus on the Manuscript Readings

FnaG's lexicographers, when they use the MSS Corpus, will be interested not just in semantics and range of vocabulary and so forth. They will also be investigating matters such as orthography, and what this might reveal about pronunciation and intellectual influences, or deviations from standard morphology or syntax and the insights these might yield into dialect. Therefore, the default position when transcribing is to give the unamended manuscript form. Where this could be misleading, or when our interpretation of the form must needs be shown, there are facilities in TEI for making editorial interventions and emendations, but these must never be silent. It must always be possible for users of the transcriptions to access the manuscript form.

Focus on the Individual Manuscript

Our primary "informants", on this project, are the scribes of the manuscripts and the corpus offers transcriptions of their work, not editions of texts. Cross-references to other manuscript versions of the same text, where helpful, can be made via footnotes or in the introduction but each transcription file must contain material solely from a single manuscript.¹⁸

Accountability

Lexicographers, when using the project's transcriptions, will be needing to assess the credibility of each piece of data they provide and make informed decisions about whether and how it can be used in the dictionary. For each form, therefore, the transcriptions must flag up any cause for uncertainty and provide details. These must be as precise as possible, with allowances made for the need to maintain pace when adding transcriptions and for the lexicographers' ability to conduct their own investigations.

In textual editing, editors are often conceived of as taking on decision-making responsibility and readers are expected to place a certain amount of trust in editors, at least pertaining to less controversial decisions. On the FnaG MSS Project, despite our policy of presenting manuscript readings by default, we are required to make interpretative decisions, when expanding abbreviations or marking up a reading with a dictionary headword, for example. Given the future use of the transcriptions we produce, we need to show the decision that was made and explain if and how it was difficult.

Logocentricity

The fundamental unit of the TEI mark-up on the project is the word (<w>). More precisely, a "word" is defined in this context as the smallest unit for which a dictionary headword would customarily be created. Compound words or verbal complexes, when the components are identifiable with existing dictionary headwords, are marked up as a single word but then split into their component words by embedding <w> elements within an enclosing <w>; the stylesheet, when generating the editorial text, will maintain their status as compounds or verbal complexes with appropriate word division.

¹⁸ TEI does offer facilities for encoding variant readings from multiple witnesses to a text (the *Critical Apparatus* module). This is not currently in use on the project but experimentation with it is certainly a possibility, if close comparison between versions of texts were felt to be useful.

There is currently no provision in place for marking up sentences, phrases, or tokens, although this may be applied in the future. The focus on indivisible lexical items is to allow queries of the resulting database — the corpus's transcriptions, by XML's very nature, are both editions and databases at the same time — to track usage of lexical items to the widest extent in the corpus.

One Transcription, Two Systems of Reference

In texts, or sub-divisions thereof, that are “verse” or “divided prose” in terms of form, it is important that two systems of reference be maintained in parallel. Every reading from the manuscript must be locatable both on the manuscript page and in the editorially-imposed structure (e.g. metrical lines). This enables both an editorial text and a diplomatic edition to be generated from the same TEI/XML mark-up. Details are given later in the chapter and were covered briefly in CHAPTER 2.

This does not apply to “prose” texts, as, there being no other editorial structure to apply to them, they are by default presented with the line division as it is on the manuscript page. There are other policy differences (e.g. word division) that justify the creation of editorial and diplomatic texts in these cases, of course.

Speed

Priority should obviously be given to producing transcriptions and commentary that are as careful, credible, and accountable as possible. This is, however, a time-limited project with ambitious objectives, so opportunities to achieve time efficiencies in TEI-related work have been taken where feasible. These include:

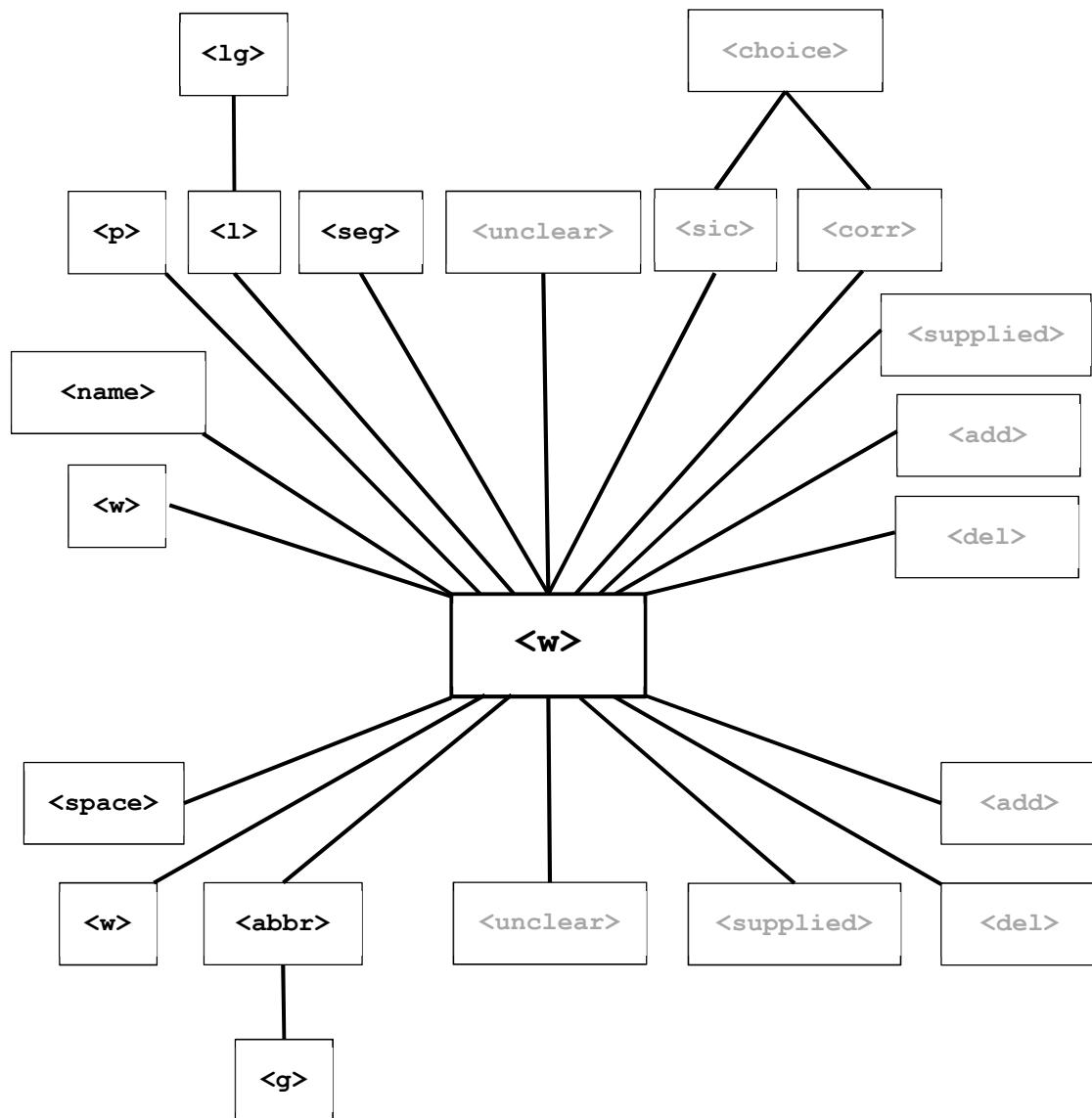
- The creation of a library of hotkeys using the *AHK* software package (see INTRODUCTION);
- The use of the headword database (*hwData.xml*) to centrally store lexical data so that only the minimum needs to be added during mark-up;
- The auto-generation of certain sections of the transcription file, such as the bibliography.

Transcription

The Word

As alluded to above, the representation of a word is at the core of the project’s “logocentric” TEI scheme. The potential parents and children of the `<w>` element are shown in Figure 43 (greyed out elements are covered in CHAPTER 4).

Figure 43



An individual word is marked up as follows,

```
<w lemma="focal" lemmaRef="http://www.dil.ie/22400"
ana="noun">facal</w>
```

The text content, “facal”, is the manuscript form. This should be an exact transcription from manuscript, except in the circumstances discussed in CHAPTER 4. Upper-case letters can be used to indicate capitalisation or display script (e.g. MS12.114^r.1; APPENDIX 8b I.7). For indication of spaces internal to a word, see below.

Three basic attributes are required for each word:

@lemma: the reading's dictionary headword, as identified by the transcriber;

@lemmaRef: a URL for the dictionary headword (@lemma) entry;

@ana: a simple POS (part-of-speech) tag (noun, verb etc.), according to the scheme set out in the "ana" sheet of xmlid_lists.xlsx (*Dropbox*, "Project" folder).

The dictionary headword/URL should be sought first in eDIL, then (if the reading cannot be identified with an entry in eDIL) in Dwelly (via *Faclair Beag*). Every effort should be made to use one of these resources, as various functions are set up to handle URL forms specific to them. If another resource proves consistently useful in this context, then the project team should be informed so that the resource can be considered for formal inclusion and the necessary technical upgrades undertaken.

In addition, to add a reference to this word in the HDSG-A MSS Slips Archive (when ready), use @lemmaSl to add the slip headword and @slipRef to add its URL in the Archive.

The @lemmaRef is very important for subsequent functions in the system. It acts as a unique identifier for the reading's headword, allowing all instances of a given headword throughout a text or file or throughout the corpus to be retrieved accurately, as well as stats to be generated on the corpus's overall contribution and richness. In another important function, it acts as a link to the record for the headword in the headword database (hwData.xml). This is drawn upon to provide a Dwelly headword/URL for headwords also in eDIL, as well as additional morphological information.

For these reasons, the @lemmaRef should be consistent in form (e.g. the "http://" prefix should always be included) and either eDIL or Dwelly (in that order of preference) should be used consistently for a given headword.

Special procedure for using Dwelly (via *Faclair Beag*)

Each Dwelly entry in the digitised version on www.faclair.com has a unique URL, which can be used to link back to that entry. For example, if the entry for *focal* is retrieved, the URL is:

<http://www.faclair.com/ViewDictionaryEntry.aspx?ID=53D7536354783855A07F93735CB692B3>

However, this URL is also unique to this search. A different URL is generated if the page is reloaded *focal* is searched for again:

<http://www.faclair.com/ViewDictionaryEntry.aspx?ID=16ABC9AEE56DE112DC8A4980D3660C82>

Both URLs still point to the entry for *focal*. However, as the @lemmaRef is meant to act as a unique ID for the headword for the purposes of searching the corpus, and not just link to the entry, having different URLs for the same headword is a problem.

It is therefore adviseable, when adding a Dwelly headword/URL, to check the corpus for previous occurrences of that headword. If found, the URL should be copied from those previous occurrences. Methods for searching the corpus are discussed in CHAPTER 7.

In time, it is envisaged that the headword database will contain a URL for every word in the corpus. It can then be searched for the designated Dwelly URL. However, at present, it only contains details for the most frequently occurring words so is unlikely to provide a solution.

While searches and queries can be performed on the corpus with @ana as a criterion, it does not play any role in identifying the headword and can thus be safely varied according to the usage in a

particular instance. That is, the @ana attribute should contain the POS for the manuscript form in context, not for the abstract dictionary headword. So, for example, an adjective would be tagged as a “noun” if it is interpreted as being used in this instance as a substantive even if the dictionary entry makes no reference to a substantive use for that adjective.

There will also be cases where the relationship with the dictionary headword is looser; these will be discussed in CHAPTER 4.

Within the Word

Abbreviations

In most pre-modern manuscripts, abbreviations, suspensions, and contractions are used to a greater or lesser extent. This was brought to a very high state of refinement in medieval Gaelic manuscript culture, meaning that an elaborate and often ingenious repository of abbreviations was potentially available to scribes. In order to provide the most accurate and accountable rendering of a manuscript form, we need to show in transcriptions not only where abbreviations have been expanded but what the abbreviation is and how sure we are that we have expanded it correctly.

As an example, we will consider the first two quatrains of ‘Is maith mo leba, is olc mo shúain’ (MS12.114r.1–4; see APPENDIX 8). In the second word, “maith”, the lenition of the “t” is indicated by a *spiritus asper* (a “lenition tick”). This is marked up in our TEI follows:

```
<w lemma="maith" lemmaRef="http://www.dil.ie/31438"
    ana="adj">mai<abbr cert="high">t<g ref="g4">h</g></abbr></w>
```

Two elements, each with an attribute, are involved: <abbr> (@cert) and <g> (@ref).

<abbr> encloses every character involved in the abbreviation (“th”, in this instance). The **direct** text content of <abbr> (“t”, in this instance) consists of the character(s) written out on the manuscript page and bearing whatever indicates the abbreviation marker (the *spiritus asper*, in this instance).

@cert: this attribute’s value is the level of certainty with which you have expanded the abbreviation; the options are “high”, “medium”, “low”, and “unknown”.¹⁹

<g> (“glyph”) encloses the characters that are not directly visible on the manuscript page in their usual form but are expanded out of the abbreviation (“h”, in this instance).²⁰

@ref: this attribute’s value is an id number corresponding to an @xml:id in a database of abbreviations stored in corpus.xml (see APPENDIX 6a l.148). This database contains details such as the abbreviation’s name, a link to its entry in *Tionscadal na Nod* (if available), and a short description. For a quick reference list, see the “Glyphs” sheet in xmlid_lists.xlsx (*Dropbox*, “Project” folder). “g4”, for example, is the id for *spiritus asper*.

¹⁹ “unknown” is used where the manuscript text is no longer legible but where an older transcription/edition exists of it that shows expansion of abbreviations. The certainty is “unknown” because you are unable to perform an independent assessment of the previous editor’s work.

²⁰ Many abbreviations are formed by placing a letter in superscript (e.g. superscript “a” is read “ar” or “ra”). In such cases, the superscript letter plus expansion should be placed in <g> and the letter above which the superscript letter is placed should be the direct text content of <abbr>.

Some abbreviations require complete transformation when expanded; none of the characters involved in the abbreviation remain in the expanded form. In such instances, <abbr> has no direct text content and the expansion is simply placed within <g>. For example, in MS12.114^r.2, the preposition “ar” is written as a “q” with a stroke through the descender. This is rendered in our TEI as follows:

```
<w lemma="ar" lemmaRef="http://www.dil.ie/3902"
ana="prep"><abbr cert="high"><g ref="g8">ar</g></abbr></w>
```

Each character bearing one or more abbreviations should have its own <abbr> element. While it might be tempting, when faced with a string of abbreviations, to enclose them all the same <abbr> element, later functions and queries will need to be able to identify the certainty level of a specific expansion and queries could potentially be composed to retrieve abbreviations rather than words.

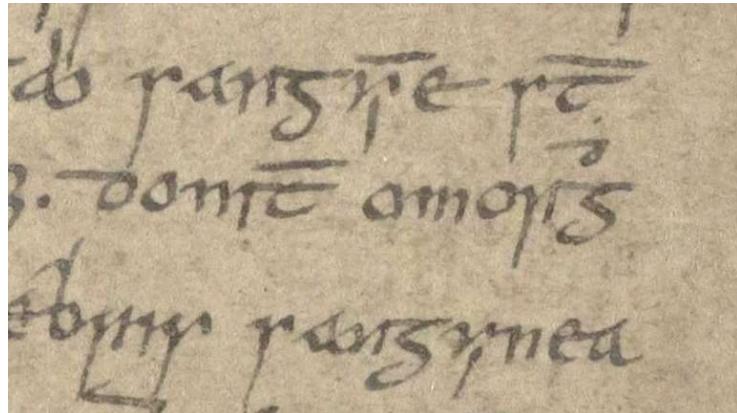
Thus, for example, “chaird” (MS12.114^r.3) is rendered as follows in our TEI:

```
<w lemma="càird" lemmaRef="
http://www.faclair.com/ViewDictionaryEntry.aspx?ID=007AB404DEE
A6B0878BD3BEECF89ABF2" ana="noun"><abbr cert="high">c<g
ref="g4">h</g></abbr><abbr cert="high"><g
ref="g11">air</g></abbr>d</w>
```

If the same letter carries multiple abbreviations, however, the same <abbr> element should be used (it being not possible, or at least not straightforward, to subdivide a letter).

For an example of this, we must leave NLS Adv. 72.1.36 and turn to a form in NLS Adv. 72.1.33 (MS5), p.37.24 (Figure 44): “morgudh”.

Figure 44



The ending, “-udh”, is indicated by a superscript d (standing for “ud”), which itself carries a *spiritus asper* (“h”). This is rendered in our TEI as follows, using only one <abbr> element.

```
<w lemma="morgad" lemmaRef="http://www.dil.ie/32573"
ana="vnoun">mor<abbr cert="medium"><g ref="g24">ud</g><g
ref="g4">h</g></abbr></w>
```

If you are unable to match the abbreviation you are looking at against any of the abbreviations listed in corpus.xml/xmlid_lists.xlsx, you will need to add a new record of it and create a new id. See APPENDIX 6b I.503 for details.

`<abbr>` elements cannot span `<w>` boundaries in TEI, so the element must be closed before the `<w>` element is closed. Occasionally, scribes will use an abbreviation to give the last character(s) of one word and the initial character(s) of the next word.²¹ In these cases, two separate `<abbr>` elements will need to be created in each `<w>` to represent the abbreviations contribution to each word. This means any stats involving counts of abbreviations will be slightly inaccurate.

For the handling of lower certainty abbreviations and representation of other problems with reading, see CHAPTER 4.

Word-Internal Spaces

Scribes will sometimes place deliberate spaces within what we would analyse as a single, indivisible word. A `<space>` element can be placed inside a `<w>` element to represent this. It will be ignored in the editorial text, so the word will be presented there as one word without spacing. The word will be represented with a space in the diplomatic text.

For an example, see “ceadna” / “cead na” at NLS Adv. 72.1.4 fol. 24r7 (APPENDIX 9b ll.117–119 and the different reanderings in APPENDIX 9c).

For more details on `<space>`, see below.

Words in Languages Other than Gaelic

Words in languages other than Gaelic (of whatever period or register) should still be marked up in `<w>` elements. However, the only attribute required is `@xml:lang`, the value being the standard XML tag for the language. See Appendix 9 for examples of Latin text.

The boundary between a loanword in Gaelic from another language and the rendering of a word from another language in orthography influenced by Gaelic can be very difficult to draw. Since the policy of FnaG is to be inclusive of the diverse different forms of Gaelic that have existed historically in Scotland, the weighting should perhaps be towards Gaelic in such situations.

The tags currently in use are:

- la (Latin);
- sco (Scots);
- hbo (Ancient Hebrew);
- und (unknown).

For a full list of tags that can be used, see <https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>

Around the Word

Word Division

Obviously, one encounters spaces more often between words. This project’s definition of a space is relative. That is, the span between letters does not need to pass any particular measurement to qualify as a `<space>`. Instead, the span between letters is marked up with `<space>` if it is noticeably larger than other spaces between other letters. The `<space>` element is omitted if the scribe could have left more of a space than they did, had they wanted to. Thus, if the scribe is struggling to fit his text

²¹ Unfortunately, no examples were to hand at the time of writing.

into a line and generally has to leave less of a span between words, then a smaller span will qualify as a deliberate space and thus for mark-up as <space>.

The treatment of each <space> element is determined by its @type:

“scribal”: This represents a space in the scribe’s text, according to the above criteria. This will be the most commonly used <space> element in most transcriptions;

“editorial”: This defines a space as inserted by a later editor; it is only to be used if the manuscript itself is illegible and an older transcription/edition is being relied upon;

“force”: This will insert a space in the editorial text but **not** in the diplomatic text. This can be useful for representing page layout and indentation;

“em”: This inserts a long space, 3× the length of the standard space. This can be useful for representing page layout and indentation. It will result in a space being inserted in both editorial and diplomatic texts.

It should be noted that word division should continue to be represented at the end of editorial divisions (e.g. at the end of metrical lines), should these (as is often the case) not coincide with manuscript lines. This is so that word division can continue to be represented in the diplomatic text.

See the mark-up around the <l> elements in Appendices 8b and 9b and the results in Appendices 8c and 9c for examples.

Compound Words

To mark up a group of <w> elements as belonging to a compound word, enclose them all in a single outer <w> element. For example, the name “Trénmór” (“threnmhoir”; NLS Adv. 72.1.36, fol.96’3; APPENDIX 7b II.83–88) is represented in our TEI as follows:

```
<w ana="adj, adj">
  <w lemma="tré" lemmaRef="http://www.dil.ie/41798"
    ana="adj"><abbr cert="high">t<g ref="g20">re</g></abbr>n</w>
  <w lemma="mór" lemmaRef="http://www.dil.ie/32548"
    ana="adj"><abbr cert="high">m<g ref="g4">h</g></abbr>oir</w>
</w>
```

For additional special rules regarding the mark-up of names, see below.

The outer <w> element need only have an @ana attribute, with its value being a comma-separated list of the POSs of the words in the compound. The processing software, with reference to the stylesheet, will use the combination of POSs to determine whether and where word division is to be applied. In most cases, at present, it will run compounds together (compare the treatment of “threnmhoir” in the editorial and diplomatic texts in APPENDIX 7c). This also means that a list of compound words in the corpus, with data on their components, could potentially be generated by a query. The component words otherwise retain their identity as separate words.

<space> elements can still be placed between components of compound words as required (e.g. NLS Adv. 72.1.36, fol.96’2; APPENDIX 7b I.50) and the rules described above will be applied.

Currently, constructions defined as “compounds” for these purposes include:

- Nouns preceded by their adjective outside of a copulative construction;

- Combinations of nouns and adjectives run together by the scribe or seemingly meant to be joined together based on, for example, the lenition of the initial of a component after the first;
- Verbs, nouns, and pronouns with a following emphatic or pronominal;
- Verbal complexes, including particles, infix pronouns, suffixed pronouns, emphatics, and pronominals, as well as the verb itself.

It should be noted that the primary relevance of compounds in the context of this project's TEI is word division in the editorial text, with the secondary relevance being to querying scribes' concepts of word boundaries. Compounds should thus be formed with editorial conventions in mind.

Occasionally, a word will have more than one component but this can only be indicated by the value of the @ana attribute, as the word cannot otherwise be split into component words that make sense in themselves, there being too many mutations involved in their conjunction. This policy has been adopted for conjugated prepositions (e.g. APPENDIX 7a fol. 96^r8; APPENDIX 7b l. 246). While some examples, like "fris" could be easily split, in the mark up, into preposition and pronoun/possessive ("fri" + "s"), the results would be misleading in many cases. For example, attempting to mark up "agum" or "leis" would lead to either the preposition "agu"/"lei" or the conjugation "um"/"is" being returned in searches, neither of which would be accurate. Therefore, in such cases, the parts-of-speech involved in the form should simply be put in a comma separated list as the value of @ana, with no enclosed <w> elements required.²² The stylesheets have been adapted to parse the @ana value and identify affected forms as either a preposition or as a pronoun/possessive, as applicable.

Names

Names be enclosed in <name> elements and then rendered with <w> elements as described above. For example, our previous example of "Trénmór" ("threnmhoir"; NLS Adv. 72.1.36, fol.96^r3; APPENDIX 7b ll.82–89) is represented fully in our TEI as follows:

```
<name type="personal">
  <w ana="adj, adj">
    <w lemma="tré" lemmaRef="http://www.dil.ie/41798"
      ana="adj"><abbr cert="high">t<g
      ref="g20">re</g></abbr>n</w>
    <w lemma="mór" lemmaRef="http://www.dil.ie/32548"
      ana="adj"><abbr cert="high">m<g
      ref="g4">h</g></abbr>oir</w>
  </w>
</name>
```

For the simpler example of Fionn, see NLS Adv. 72.1.36, fol.96^r3; APPENDIX 7b ll.66–69.

Names are differentiated by @type:

- "personal": human individuals;
- "place": locations;
- "population": peoples and communities;

²² This policy is open to discussion. If easier retrieval of the components of conjugated prepositions and suchlike is felt to be worth some slightly misleading forms being occasionally returned, then (with a certain amount of work adjusting those examples already in the database) the project can proceed on that basis.

“other”: miscellaneous.

A `<name>` element can be embedded within another `<name>` element, if needed. For example, in the *Book of Deer* (MS1.4r.1), we find the placename “Pett Ma[e]l Dúib”. This is a placename, but it contains a personal name, which in turn contains another personal name. This would be rendered in our TEI as follows:

```
<name type="place">
    <w lemma="pet(t)" lemmaRef="http://www.dil.ie/34300"
        ana="noun">pett</w>
    <name type="personal">
        <w lemma="Mael" lemmaRef="http://www.dil.ie/31244"
            ana="noun">mal</w>
        <name type="personal">
            <w lemma="dub" lemmaRef="http://www.dil.ie/18985"
                ana="noun">dúib</w>
        </name>
    </name>
</name>
```

Names should be marked up as such as a matter of policy. They offer a qualitatively different sort of lexicographical evidence, such as by fossilising forms that are no longer productive and it is thus important to be able to easily distinguish them in queries and searches.

Other Elements

Non-Word Elements

Elements other than words will often be needed to represent a scribe’s rendering of the text fully.

Punctuation

One such case is punctuation. For example, at NLS 72.1.4 fol. 24^r9 (APPENDIX 9b l.120), the scribe appears to use a comma. Punctuation is represented by its nearest equivalent in UTF-8, enclosed in a `<pc>` element. `<pc>` elements can also be placed within a `<w>` element, if the punctuation appears as part of the rendering of a word (e.g. NLS 72.1.4 fol. 24^r2; APPENDIX 9b l.22–23).

If a punctus is used as part of an abbreviation, it is not represented as a separate `<pc>` element. Sometimes, scribes add points around other abbreviations, perhaps to flag up the proximity of an abbreviation (e.g. NLS Adv. 72.1.4 fol. 24^r7; APPENDIX 9b ll.117–119). If this appears to be the sole purpose of a piece of punctuation, it is also not represented, according to the project’s current policy, but is considered an innocuous part of the abbreviation.

Punctuation used to separate metrical lines (e.g. throughout the text in Appendix 8) should be represented, as this is part of the evidence used for analysing the poem’s metre and thus confirming certain readings.

The `<pc>` element is also used to render simple decorations or line-fillers (e.g. NLS Adv. 72.1.4 fol. 24^r10; APPENDIX 9b l.147).

Numbers and Dates

Numbers can be marked up as words, even when rendered in Roman and Arabic numerals by the scribe (e.g. NLS Adv. 72.1.36 fol. 96^r8; APPENDIX 8b l.238), as long as the number is equivalent to a single word (i.e. 1–10 or divisible by 10 up to a certain point). If not, the `<num>` element can be used.

For year dates in the form “1613”, the `<date>` element should be used. More complex dates should be marked up using `<num>` and `<w>` elements, as needed.

New Lines and Other Divisions

The elements used to indicate new pages (`<pb>`), new columns (`<cb>`), and new lines (`<lb>`) have already been noted, and examples of `<pb>` and `<lb>` in action can be found in APPENDICES 7–9 (`<cb>` functions in exactly the same way). These elements can be placed almost anywhere within the mark-up (they cannot be placed within `<g>` but this should never be needed), except their order in relation to each other should be logical. So the divisions on the manuscript page can always be rendered accurately, even if they split `<w>` elements.

Annotation

While as many opportunities as is practical have been created to encode formal data within the mark-up, there will inevitably be complex or noteworthy situations requiring special comment, explanation, or cross-reference. There are currently three ways of adding this.

Element Comments

Any element in the mark-up can have a @comment attribute (which has been specially customised for this project), the value being the transcriber's comment on that element. However, the only element set up to display this sort of comment in the output is <w> (the comment displays in the tooltip on mouseover in the HTML). If being able to make visible comments on other sorts of elements seems to be very valuable, this should be raised with the project team and the necessary upgrades can be made.²³

Footnotes

To add a footnote within the transcription, the following template can be placed almost anywhere in the transcription <body>:

```
<note type="fn" resp="NNN">
    <p><!-- Text of footnote here --></p>
</note>
```

A number will be automatically assigned to the footnote by the processing software and the text of the footnotes placed at the end of each <div>. A reference number hyperlinked to the footnote will be placed at the location of the <note type="fn"> element.

The @resp attribute should be the id of the transcriber adding the note or responsible for its content.

For the mark-up used in commentary and academic prose, see CHAPTER 5.

'altText' Notes

If there is another version of the text being transcribed that could be of interest as the subject of close comparison, this can also be placed in a <note> element as a sibling following the element containing the section of transcription to which the comparison is being made:

```
<note type="altText" resp="NNN">
    <p><!-- Text here --></p>
</note>
```

This text does not need to be marked up in the same way as transcriptions and can be simply typed in within the <p> element. Indeed, the transcription mark-up will not work in this context. Instead, the mark-up designed for commentary should be used here (see CHAPTER 5).

The @resp attribute should be the id of the transcriber adding the note or responsible for its content.

²³ The @comment facility has not been used extensively so there are few examples in the corpus at present.

Organisation of Code

As exemplified in APPENDICES 7b, 8b, and 9b, indentation can and should be used to represent the hierarchy of XML elements within the TEI mark-up. If the place in the hierarchy is lost, *oXygen XML Editor* has a function that automatically adds indentation accurately (Ctrl + Shift + P, by default). A new line should be started for each element **above <w> in the hierarchy** by hitting Return. No new lines should be added in this way within <w> or any of its descendants. *oXygen* will interpret this as adding a space within the text content itself and the processing software will then overwrite any rules regarding word division for that <w> element.

For reasons of space, the contents of <w> elements in APPENDICES 7b, 8b, and 9b sometimes straddle line breaks in this document and automatically indenting code in *oXygen* will often have the same effect. However, while producing the same visual effect, this is **not** the equivalent of pressing Return and will cause no problems.

Review

- The project's mark-up strategy emphasises accountability, access to primary evidence, and individual lexical items.
- A range of resources are required in order to add the required data during mark-up.
- The focus of activity and of the element hierarchy is on the word (<w>).
- Only a minority of the information that can be made available about each word is added within the <w> element; much is drawn from elsewhere in the software by the processing software.
- Facilities are available that allow comment to be made on specific points in the transcription.
- Care is required when organising marked up text, even when simply considering visual clarity at the XML level.

Chapter 4

Advanced Transcription

“Do-dechad-sa,’ ol ind óclach, ‘a tírib ingnad, a tírib gnáth, co fesur uait-siu fót forsa mbeba ocus fót for a ngénir fis 7 anfis.””

- From *Immacaldam Choluim Chille 7 ind Óclaig oc Carraig Eolaig* (ed. Carey).

Introduction

This chapter sets out how to handle (in terms of TEI mark-up) more complex situations that can occur on the manuscript page and how to flag up unknowns and points of uncertainty. These fall into three categories:

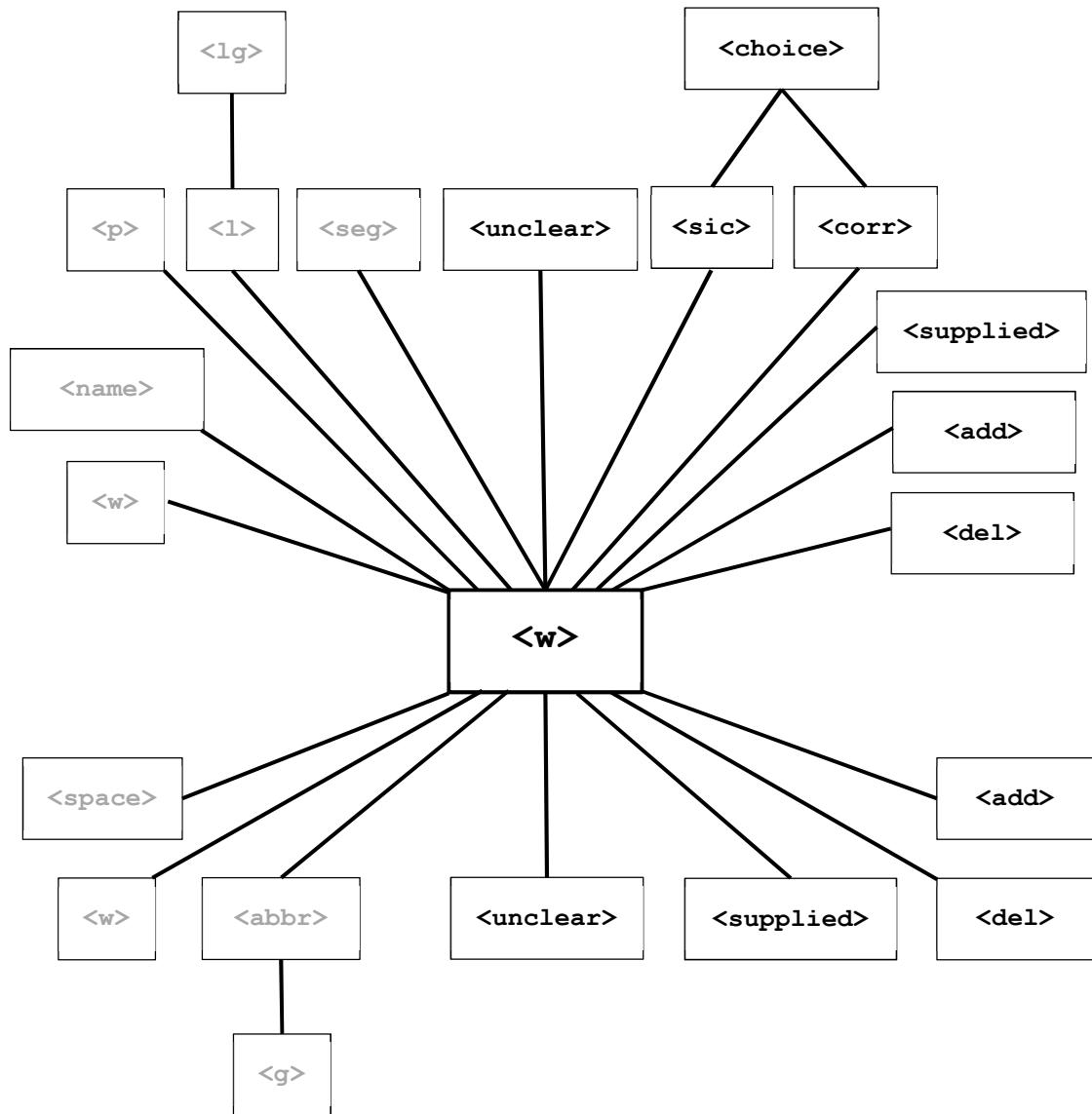
- Special text segments;
- Later scribal interventions;
- Unclear text;
- Editorial emendation.

While this can all be problematic hermeneutically, the general principles and techniques resemble those set out in CHAPTER 3 and there is little involved here that is hugely more difficult conceptually.

Elements Involved

As in CHAPTER 3, the elements involved in marking up more problematic sections cluster around <w>.

Figure 45



Special Text Segments

Sections of text playing particular functions should be marked up as such, to ensure appropriate layout and formatting is applied in the editorial text.

Titles

If a text is provided with a title in the manuscript, then the text of the title should be marked up normally and enclosed in the element `<seg type="title">`. It is envisaged that this will be used only for strings of text set out in some way as a title by the scribe. For an example, see APPENDIX 7b II.6–20. Titles will be presented in bold in the output. ×

Catchwords

Sometimes, what will be the first word of a new folio side or page will be written at the foot of the preceding folio side or page, as a kind of relative foliation. This is a “catchword”. It should be placed marked up as normal then enclosed in `<seg type="catchword">`, which will allow it to be differentiated visually and in searched from the continuous run of the text. If it appears (as it probably will) on a new line, then this should be indicated with an `<lb>` element.

Dúinte

A *dúnad* is the echo of the first line of a poem in the last line, to signal that the poem is ended and complete. Occasionally, *dúinte* can appear within poems to signal a change in topic or at points where the poem originally ended before further material was added.

Scribes will often highlight a *dúnad* by writing it out again, sometimes on a new line. If this has been done, then the repetition of the *dúnad* (**not** the original *dúnad*) should be marked up as normal (c.f. Figure 38) within an `<lg>` element differentiated using the `@type` and then an `<l>`:

Figure 46

```
<div n="1" type="verse" resp="Hand9" corresp="MS12.1">
    <!-- PRECEDING QUATRAINS OF POEM HERE -->
    <lg n="6" type="stanza" xml:id="MS12.1.6">
        <!-- QUATRAIN LINES OF POEM HERE -->
    </lg>
    <lg type="dúnad" xml:id="MS12.1.7">
        <l xml:id="MS12.1.7a"><lb n="14" xml:id="MS12.85v.14">
            <!-- TEXT OF DÚNAD -->
        </l>
    </lg>
</div>
```

Note that the `<lg>` and `<l>` elements used for the *dúnad* still require unique ids as the values of their `@xml:id` attributes, although they should not have numbers (`@n`). A suitable system should be adopted for ensuring that the *dúnad* elements’ ids are unique. In Figure 46, the *dúnad* concludes the poem so its ids are formed from the next number in sequence from the number of the last stanza (as if it were a stanza itself). If the *dúnad* had appeared mid-poem, some alternative suffix would have to be found (e.g. “.dun1”) to form an id that would not recur again in the corpus.

As with catchwords, the *dúnad* will likely be repeated on a new line. If so, this should be indicated using `<lb>`, which should have both an `@n` and an `@xml:id` attribute.

Cinn fo Eitib

To save space, scribes will sometimes place text in gaps at the rightward ends of preceding (or occasionally following) lines. This is usually separated from the line proper by a symbol resembling “//” or suchlike (for which the term is *cenn fo eiti* (“head under wing”)), although it can also be done silently, in which case it is detectable only by considering the text’s sense or by comparison with another manuscript witness.

This is marked up in three steps:

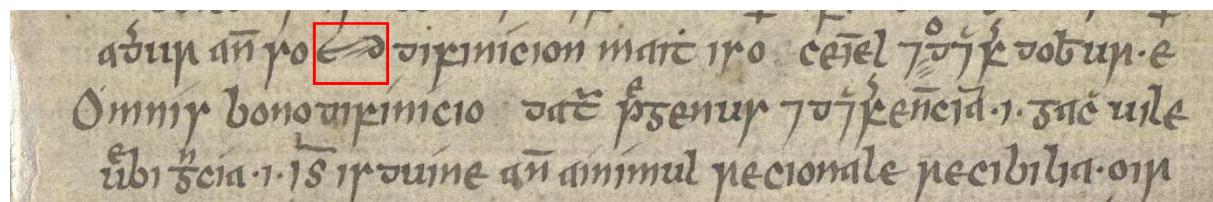
4. An `<lb>` element is inserted at the end of the line where the text ought to be read; instead of an `@xml:id`, it has a `@sameAs` attribute, with the value being the `id (@xml:id)` of the `<lb>` element marking the line where the text was placed by the scribe;
5. The text moved by the scribe is marked up as normal after the `<lb sameAs>` element, with the next true manuscript line begun as normal with an `<lb>` element.
6. Any *cinn fo eitib* included by the scribe should be enclosed in a `<pc>` element, then this `<pc>` element should be enclosed in `<seg type="cfe">`.

For an example, we must turn to NLS Adv. 72.1.2 fol.35r10–12 (Figure 48). The mark-up should thus resemble the following (Figure 47):

Figure 47

<code><div n="3" type="prose" resp="Hand20" corresp="MS14.3"></code>
<code> <!-- PRECEDING LINES OF TEXT --></code>
<code> <lb n="10" xml:id="MS14.35r.10"/><!-- MS FOL. 35r10 TEXT HERE --></code>
<code> <lb n="11" xml:id="MS14.35r.11"/></code>
<code> <!-- MS FOL. 35r11 TEXT HERE --></code>
<code> <lb sameAs="MS14.35r.10"/></code>
<code> <seg type="cfe"></code>
<code> <pc>//</pc></code>
<code> </seg></code>
<code> <!-- TEXT RELOCATED FROM MS FOL. 35r10 HERE --></code>
<code> <lb n="12" xml:id="MS14.35r.12"/></code>
<code> <!-- MS FOL. 35r12 TEXT HERE --></code>
<code></div></code>

Figure 48: *Cenn Fo Eiti* Highlighted



Programming work is still in progress to stabilise the output in such situations. The objective is that the relocated text appear:

- Where it is meant to be read and without any separating symbols in the editorial text;
- Where the scribe located it and with any separating symbols in the diplomatic text.

At present, the relocated text will always display where it is meant to be read, although with an accurate line reference (Figure 49).

Figure 49

10. aderur ann so //
11. Omnis bono difinicio datur per genus et dethferenniam .i. gach uile
10. **difinicion** is o e-ceinel agus \ o / deithfir do berur . e
12. uerbi gracia .i. is ed is duine ann ainimul recionale recibilia . oir

Later Scribal Interventions

Manuscript text can be altered, either by the hand that originally inscribed it or by a later hand, and the alteration can consist of deleting text, adding text, or both. We want to capture as much about scribe's behaviour as possible and mistakes and re-editing can yield particular insights into attitudes to language and style. It is therefore important that sections of manuscript text in which there have been later interventions be carefully marked up so that the original (where recoverable), each intervention, and the final outcome are all made available.

Deleted Text

Sometimes, deleted text is neatly scored out or, in a medieval convention, dots are placed beneath the characters or words that are supposed to be removed (subpunction). In other cases, text is erased or scored out completely, so that what was there originally is irrecoverable.

Identifiable Agent

An example of the former is found at NLS Adv. 72.1.36, fol. 96^r1 (APPENDIX 7b II.22–35). Here, the scribe, Eoghan Mac Gilleoin, wrote out the title of the next tale in the manuscript, *Bruighion Cheisi Choruinn* (see fols 104^v–111^r), not the correct title, *Bruighion Bheag na hAlmhain*. He has then subpuncted the erroneous title and added the correct title to the left.

The element encloses the text deleted by the scribe, which is otherwise marked up as normal. This element has a @resp attribute, with the value being the id for the hand responsible for the deletion (in this case, “Hand9”). The same ink was used in the subpunction as in the subpunction and in the inscription of the new title, which is also in the same hand. Hence we can identify the hand responsible for the deletion.

The text marked up with , in the output, will be struck through (see APPENDIX 7c) but otherwise retrievable through all queries and searches and accessible to the various reading aids built into the digital edition. The tooltip generated by placing the mouse over any of the deleted words will display a notification that the word was deleted and the hand responsible.

Unidentifiable Agent

In cases where text is simply scraped away or a different ink is used to score it out, we will find it much more difficult to identify the hand responsible. If a hand has added new text in place of what was deleted or has somehow taken account of the deletion, we could posit that that hand was responsible. Other evidence relating to the manuscript might supply names of individuals known to have intervened in it who might thus be candidates. However, if no clear identification emerges, the element should be used above, with the @resp attribute given the value “Hand999”. This indicates an unidentified hand.²⁴

Added Text

Added text is split into two categories:

3. Insertions, which are to be read within the main text;
4. Glosses, which expand upon the main text but do not directly fit within it.

The two categories are handled in a very similar way, using the <add> element.

²⁴ Note that this indicates specifically an **unidentified** hand, not an anonymous hand. That is, it is used to designate scribal contributions that cannot even be grouped securely with other scribal contributions, let alone attributed to a known scribe.

Insertions

Insertions are enclosed in an `<add>` element, which has a set of attributes:

- `@type`: the sort of addition (in this case, “insertion”);
- `@place`: where the scribe has placed the text to be inserted, from the following options:
 - “above”;
 - “below”;
 - “margin, left”;
 - “margin, right”;
 - “margin, top”;
 - “margin, bottom”;
 - “inline”.
- `@resp`: the id for the hand responsible for the insertion. This attribute is required even if the same hand is also responsible for the main text.

The intended location of the insertion, which is where it should be marked up within the `<add>` element, may be obvious where the scribe placed it on the page. Alternatively, the scribe may have added callmarks to place it. If its location is uncertain, the material should be included as close as possible to its physical location, outside any elements below `<p>` and `<lg>` level, and the issue explained in a footnote.

Within the `<add>` element, the inserted text should be marked up as normal. In the output, it will be enclosed in punctuation marks that indicate the location of the text for insertion on the manuscript:

- above: “\.../”;
- below: “/...\”;
- margin, left: “>...>”;
- margin, right: “<...<”;
- margin, top: “\\...//”;
- margin, bottom: “//...\\”;
- inline: “|...|”.

The tooltip will display a message indicating the hand responsible for the addition and this information will also be retrieved if an e-slip is generated for the word.

For an example, see APPENDIX 7b II.6–20.

If only characters within a word have been added, then the `<add>` element can be placed within `<w>` and the added characters can be its direct text content. This will not change which hand is recorded as responsible for the word but the addition and another hand (if applicable) will be noted in tooltip note.

Glosses

Glosses are marked up in almost exactly the same way, except that the value of @type is “gloss”. In output, the gloss will likewise have its location on the manuscript page indicate but it will also be preceded by the abbreviation “gl.”, in bold.

If the gloss clearly relates to a particular word or short phrase, then both the `<add>` element containing the gloss and the word/phrase to which it relates can be placed within `<seg type="gloss">`. This will mean that lemma and gloss can be retrieved and related to each other by queries and that tooltip messages will highlight their relationship. If the gloss cannot be attached to a very specific point in the text, however, it can simply be placed in a `<add type="gloss">` element at an appropriate location.

As there are no examples of glosses in the corpus at the time of writing, the following fictional example has been created. It has been imagined that another scribe of NLS Adv. 72.1.4 (APPENDIX 9) has glossed “sine” with “.i. gan” at fol. 24r1.

Figure 50

```
<seg type="gloss">
    <w xml:lang="la">sine</w>
    <add type="gloss" resp="Hand25" place="above">
        <w lemma=".i." lemmaRef="http://www.dil.ie/26892"
            ana="abbr">.i.</w>
        <space type="scribal"/>
        <w lemma="cen" lemmaRef="http://www.dil.ie/8581"
            ana="prep">gan</w>
    </add>
</seg>
```

This produces the following output, in the editorial text:

Figure 51

1.
1. { re sine \gl. .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis } 3. gan resun { is } -gap- 4. { ni is { cuis } } -gap-

The tooltip messages clarify the text and object of the gloss and the hand responsible:

Figure 52

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.

1. { re sine \gl .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis }

Add Comment

Language: Latin:
unk. unk. (24)
- this word is difficult to decipher
- this word is in a language other than Gaelic

Abbreviations: none

[Back to MS contents](#)

A gloss has been added by Maol-Sheachlainn mac lollainn Mhic an Leatha Ruaidh (Hand25): ".i. gan"

ap-

Figure 53

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.

1. { re sine \gl .i. gan / } { racione } {nul} 2. la est
.i. an { raed bis } 3. g

Add Comment

cen: prep
Maol-Sheachlainn mac lollainn Mhic an Leatha Ruaidh (25)
- this word is difficult to decipher

Abbreviations: none

This has been added as a gloss on "sine" by Maol-Sheachlainn mac lollainn Mhic an Leatha Ruaidh (Hand25).

Change of Hand

The hand rendering the main text might change. That is, a hand other the one which began the text may take over for a greater or lesser period. If this occurs, the change in hand is indicated with a closed `<handShift/>` element, placed wherever the change occurs, with the `@new` attribute having as value the id of the hand that is taking over. The text is then marked up as normal. If the first hand takes over again, then this needs to be indicated with another `<handShift>` element. If a new `<div>` (whether a sibling or child of the current `<div>`) is added, then its `@resp` attribute will update the hand again.

Whether to use `<handShift>` or `<add type="insertion" place="inline">` is determined mainly by the nature of the new hand's contribution. If it is brief and seems like maintenance of the main text, it is an insertion. If it is longer and seems integral to the main text, it is a change of hand.

Unclear Text

Manuscripts vary in their physical condition and authors/scribes vary greatly in their adherence to orthographies, dialects, and standards that are recognisable to us. The resulting issues need to be documented within our mark-up for the information of FnaG's lexicographers.

If the resulting issues could render the text unhelpfully misleading for users, editorial emendation should be considered instead (on which, see the following section).

Problematic Text

Unclear Text

Text that cannot be read and interpreted with certainty should be placed in an `<unclear>` element. `<unclear>` can be used to enclose specific characters within a `<w>` element or multiple elements at `<w>` level and above. It should be as focused as possible on the section of the text which is unclear.

Use of `<unclear>` assumes that some sort of approximation can be made of the text enclosed in it. If the text cannot be reproduced, a `<gap>` element should be used instead (see below).

`<unclear>` takes a number of attributes:

- `@reason`: what is making the enclosed text unclear, of the following values:
 - “text_obscure”: characters are visually difficult to identify (e.g. APPENDIX 8b ll.49–56);
 - “damage”: loss of writing surface means that characters are missing;
 - “fold”: the writing may well be legible but the arrangement of the page in the digital image obscures it;
 - “interp_obscure”: there is no visually apparent problem but a word cannot be identified satisfactorily (e.g. APPENDIX 9b ll.41–44);
 - “abbrv”: an uncertain expansion of an abbreviation is hampering interpretation;
 - “char”: a specific character is ambiguous, hampering interpretation;
 - “met”: a line of verse is metrically irregular;
- `@cert`: how certain the text enclosed in `<unclear>` is, of the following values:
 - “high”,²⁵
 - “medium”;
 - “low”;
 - “uncertain”;
- `@resp`: who is responsible for the reading of the text enclosed in `<unclear>`; this might be the transcriber, a reviewer of the transcription, or a bibliographical reference to an edition of the text outside of the project.

²⁵ While technically meaning “high certainty”, `<unclear cert="high">` will still display the text as problematic, although the problems will be characterised in the output as slight.

Placement of <unclear> will depend on @reason. <unclear reason="char"> is designed to enclose a single character, for example, and <unclear reason="met"> only makes sense in the context of poetry. <unclear> elements can be placed within other <unclear> elements and thus build a more detailed account of the problems the transcriber has experienced. For example, at APPENDIX 9b ll.47–57, the form “cuis” (?) has been placed in <unclear reason="text_obscurer">. On account of the difficulties of reading this word, the phrase of which it is part (“ni is cuis”) has been placed in <unclear reason="interp_obscurer">. The problems with making out a key word have thrown into question the interpretation of other words which themselves can be made out visually.

<unclear reason="damage"> is distinct from all the others. All the others indicate a problem with reading the text that could theoretically be resolved by someone with better perception, knowledge, equipment, luck, or suchlike. Meanwhile, <unclear reason="damage"> indicates that text has been lost irrecoverably and that what is enclosed in the element is hypothesised by an editor (see also ILLEGIBLE OR LOST TEXT, below). <unclear reason="damage"> implies reconstruction of text that the scribe once did produce but its use should be reserved for where the writing surface has been completely removed.

In terms of output, the @cert value will generate a colour code for the enclosed text indicating the level of difficulty and messages will display in a word’s tooltip detailing any problems affecting that word:

Figure 54

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.

1. { re sine \gl. .i. gan / } { racione } { nul } 2. la est
.i. an { raed bis } 3. gan resun { is } -gap- 4. { ni is { cuis } } -gap-

Add Comment

[Back to MS contents](#)

ni: noun
unk. unk. (24)
- the interpretation of this word, or its context, is doubtful; there is a particular issue with "cuis"
Moderate problems with this reading
Abbreviations: none

Nature

Figure 55

24r: unk. unk. (Hand24) [Collect e-Slips](#) [Headword Search](#)

1.

1. { re sine \gl. .i. gan / } { racione } { nul } 2. la est
.i. an { raed bis } 3. gan resun { is } -gap- 4. { ni is { cuius } } -gap-

[Add Comment](#)

[Back to MS contents](#)

cúis: noun
unk. unk. (24)
- the interpretation of this word, or its context, is doubtful
- this word is difficult to decipher
Moderate problems with this reading

Abbreviations: none

nature

Unidentifiable Words

If a word cannot be identified with any known dictionary headword, regardless of confidence level (doubtful identifications can be indicated using `<unclear reason="interp obscure">`), then the `@lemma` of the relevant `<w>` should be set to “UNKNOWN” (case sensitive). `@ana` can be given if the POS is obvious.

New Words

If a word does not match a dictionary headword but can be interpreted using a dictionary headword (e.g. an otherwise unattested adjectival form of an attested noun), then the attributes in `<w>` are slightly different:

- `@lemma`: the proposed dictionary headword for the new word, prefixed with “*”;
- `@source`: the existing dictionary headword used to interpret it or on which it seems to be based;
- `@lemmaRef`: the url for `@source`;
- `@ana`: the POS of the word as used in the transcription.

This can also be used for items that are not necessarily new to Gaelic philology but whose interpretation requires explanation. For example, see APPENDIX 7b ll. 125–131. Here, it is hypothesised that “caithe” is a past participle of the verb “caithid”. Normally, a participle could be understood as a form of the verb and the verb simply given as the headword. In this case, the interpretation is more doubtful, so both `@lemma` and `@source` have been given.

Problematic Abbreviations

As already noted in CHAPTER 3 (ABBREVIATIONS), the `<abbr>` element also has a `@cert` level that can be used to indicate uncertainty (or the lack thereof). Adding “medium” or “low” as values will produce colour coding and messages flagging up the problem with the abbreviation, which can then be further examined via the e-slip table.

Figure 56

Form	MS Ref.	MS Context	Text Form	Issues?	Abbrevs
cumsgaigthe	MS17 24r 9	... 9. tosach cumsanaid agus cumsgaigthe maille 10. fein ni co ...	Prose	moderate;	M_stroke Suspension_stroke

Figure 57

Form	MS Ref.	MS Context	Text Form	Issues?	Abbrevs
cumsgaigthe	MS17 24r 9	... 9. tosach cumsanaid agus cumsgaigthe maille 10. fein ni co ...	Prose	moderate;	M_stroke Suspension_stroke

Note that, unlike with <unclear>, a “high” value for @cert in <abbr> denotes an absence of uncertainty, not “slight” uncertainty.

Illegible or Lost Text

Gaps

If a section of text cannot be made out at all and cannot be hypothesised with sufficient confidence, a closed <gap> element should be placed there, with the following attributes:

- @reason: what has caused the gap, of the following values:
 - “text obscure”: the text is illegible, although the writing surface is intact”;
 - “damage”: the writing surface has been lost;
 - “fold”: the writing may well be legible but the arrangement of the page in the digital image obscures it;
 - “text_omitted”: there is no visually apparent problem but disjunction in the text suggests that material has been omitted;²⁶
- @extent: the length of the gap, using whatever unit seems most appropriate (e.g. “3 chars”, “1 ms. line”); “unknown” is also an option, particularly where loss of writing surface is involved;
- @resp: the id for the individual responsible for measuring and assessing the gap.

See APPENDIX 9b II. 45, 58 for examples.

Characters

If some characters can be made out, but not an entire identifiable word, then these characters can be enclosed in a <c> element (no attributes required), with <gap> being used where the word cannot be made out. However, note that <c> elements cannot, in TEI, contain <abbr>/<g> elements. So, if the word fragment includes an abbreviation, it will have to be placed in a <w> element for the

²⁶ If the missing text can be confidently reconstructed, it can be placed in <supplied> (see below) instead. <gap> should be used where no text can be hypothesised.

abbreviation to be expanded. The @lemma can then be set to “UNKNOWN” and <[gap](#)> elements placed to indicate the lost portions of the word.

Editorial Emendations

Transcriptions in the FnaG MSS Corpus should in general reflect the forms used in the manuscript at hand as closely as possible. However, where a form deviates severely from what is expected, to the extent that interpretation is obscured, editorial interventions can be made. This is both to avoid causing users needless confusion and delay and to “show our working”, in terms of how the form came to be identified. At the same time, it is important the the original manuscript reading be always retrievable.

This is done in a number of ways.

Supplying Characters

If the emendation simply involves adding characters to a word, then the characters should be included within the `<w>` element and then enclosed within `<supplied>`, which should have a `@resp` attribute indicating who supplied the characters. For an example, see APPENDIX 8b I.63.

In the editorial text, the added characters will be enclosed within square brackets and a tooltip note will flag up the fact that an editor has intervened in the form. In the diplomatic text, these characters will be omitted silently.

`<supplied>` cannot be used inside `<g>`. If characters need to be supplied inside the expansion of an abbreviation, then the abbreviation should be expanded in amended form and the `@cert` level of `<abbr>` adjusted downwards, should the resulting expansion be unusual.

Supplying Words

Entire words (`<w>`) can also be enclosed by `<supplied>`, again with a `@resp` attribute identifying the editor responsible. `<w>` elements enclosed by `<supplied>` should be otherwise marked up as normal. Whether such `<w>` elements should contain `<abbr>/<g>` elements depends on why the form is being supplied (perhaps so if the omission was accidental, perhaps not if the text is inherently faulty) and whether any clear models exist for how that scribe would have rendered that word. It is ultimately up to the transcriber.

Emended Forms versus Original Forms

If a more complex emendation of a word is required, which involves either deleting characters or rearranging their order, then a `<choice>` element is used (without attributes). The `<choice>` element encloses `<sic>` and `<corr>`. The `<w>` element, with all attributes, containing the original manuscript reading is placed in `<sic>` and the `<w>` element, with all attributes, containing the emended form is placed in `<corr>`, which has a `@resp` attribute indicating the editor responsible. Any `<space>` elements or suchlike should be placed outside of `<choice>`.

The model is therefore as follows:

Figure 58

<code><choice></code>
<code><sic></code>
<code><w><!-- MS reading here --></w></code>
<code></sic></code>
<code><corr></code>
<code><w><!-- Emended form here --></w></code>
<code></corr></code>
<code></choice></code>

The output will display the emended form, followed by “*/alt*”. Both will be enclosed in curly brackets. Tooltip messages displayed by each will indicate the original and emended forms and an e-slip generated from “*/alt*” will display the original reading.

Competing Readings

Functionality was developed to allow the transcriber to encode two alternative readings of short strings of words in parallel to one another, if there really is no way of choosing between them. This would not imply that one was correct/emended but that both were possibilities. The functionality consisted of placing two `<unclear>` elements within a `<choice>` element and then using a system of reference numbers to link the counterpart `<w>` elements in each `<unclear>` element to one another.

However, the results proved very difficult to process coherently and the whole concept could be characterised as an abdication of editorial responsibility. For these reasons, this functionality has been deprecated for now, although examples can still be found in the corpus. Instead, in such situations, it is suggested that one reading be chosen by whatever means, the `<unclear>` element used if needed, and any alternative readings set out in a footnote.

If this functionality or something like it seems worthwhile and worth the effort, then this matter can be re-visited.

Review

- Numerous aspects of page layout, textual apparatus, and scribal behaviour can be marked up in TEI.
- TEI mark-up can also be used to track the hand producing text at any given point.
- Unclear or problematic text can be marked up and assessed in terms of cause and confidence of interpretation.
- There are various ways to amend transcriptions while still retaining the original manuscript reading.

Chapter 5

Transcription Introductions

“Redig dam a De do nim
cindead coir ar chomaimsir
a lis cech leirfhilead lind
fis na reigided romaind.

Indeosad co seim solas
cach re head na amaros
soirb cach daigen co dagar
doirb aiged ar an dalad.”

- [?Giolla Íosa Mór?] Mac Firbhisigh
(*Book of Lecan*, fol. 190^ra33–38).

Introduction

As well as the transcriptions themselves and the metadata embedded in the mark-up, the transcription files also require introductions. These contextualise them in historically, culturally, and socially and highlight their major linguistic pitfalls and points of interest. Like the texts, the introductions are marked up in TEI/XML.

However, the mark-up is slightly different from that used for the transcriptions themselves. This chapter sets out the main techniques involved in producing scholarly writing and attendant apparatus in the project's adaptation of TEI. Note that it does not cover the specific content or structure of different sections of the introductions nor what is to be dealt with where. This is summarised in APPENDIX 5b but is primarily to be decided under the direction of the Manuscripts Consultant.

The facilities set out in this chapter can also be used in footnotes to the transcriptions proper.

Unfortunately, at the time of writing, only two files in the corpus (T2 and T8) currently have viable introductions. Adding more introductions is a high priority for the project.

Formatting Text

When writing introductory material (i.e. within <`teiHeader`>), text can mostly be inputted directly, without the word-by-word, space-by-space mark-up used in the transcriptions. The basic text formatting (bold, italics, etc.) available in most word processors can be reproduced in our TEI. Input within some elements, such as titles, will be formatted automatically but certain tools are at the general disposal of the writer of the introduction. As previously, text is formatted by enclosing it within a particular element. That element is <`hi`> ("highlight"). The formatting is then determined by the value of @rend ("rendering"):

Bold: "bold";

Italics: "italics";

Underline: "underline";

Subscript: "sub";

Superscript: "sup";

Thus, to produce the manuscript reference, "fol. 24^r10", within a paragraph in an introduction:

```
... fol. 24<hi rend="sup">r</hi>10 ...
```

<`hi`> elements can be placed within each other to format text multiple ways simultaneously or to add additional formatting to a particular substring of text. For example,

```
<hi rend="italics"><hi rend="bold">This text is in italics and in bold.</hi></hi>
```

This text is in italics and in bold.

```
<hi rend="italics">This text is mostly in italics, although with occasional use of <hi rend="bold">bold</hi> for emphasis.</hi>
```

This text is mostly in italics, although with occasional use of bold for emphasis.

It is also possible to colour text. To do so,

4. Enclose the text in <`hi` rend="colour:"> ... </`hi`>;
5. Visit https://www.w3schools.com/colors/colors_picker.asp and select the name or code for the desired colour (omit the prefixed "#" from the code);
6. Add the name or code after "colour:" in the @rend value.

Thus,

```
<hi rend="colour:red">This text is red.</hi>
This text is red.
```

TEI/XML — and HTML, into which it will be transformed — offer a lot of options for formatting text (this is the same underlying technology that powers most word processors, after all), so, if more functionality is needed than is set out here, this should be made known to the project team, as probably can be added.

Lists

To form a list within a `<p>` element,

3. Add a `<list>` element;
4. For each item, add an `<item>` element as the child of `<list>`, with the text as its content.

If needed, sub-headings within the list can be added by adding a `<head>` element as a child of `<list>`, with the sub-heading as its direct text content.

As an example, the REVIEW section from the previous chapter would be marked up as shown in Figure 59.²⁷

Figure 59

<code><list></code>
<code><item></code> Three textual forms are currently recognised in the corpus: “prose”, “verse”, and “divProse”. <code></item></code>
<code><item></code> Texts can consist of combinations of multiple forms. <code></item></code>
<code><item></code> A text’s form is indicated by the @type attribute of its enclosing <code>&lt;div&gt;</code> element. <code></item></code>
<code><item></code> The form determines the text’s overall TEI structure. <code></item></code>
<code><item></code> A change of form is represented by embedding a <code>&lt;div&gt;</code> element within the text’s enclosing <code>&lt;div&gt;</code> element; this can be a complex operation. <code></item></code>
<code><item></code> Embedded <code>&lt;div&gt;</code> elements within <code>&lt;div&gt;</code> must correspond to <code>&lt;msItem&gt;</code> elements in the transcription file header unless they stand for text at the top-level in the text (i.e. returning to the top-level <code>&lt;div&gt;</code>). <code></item></code>
<code></list></code>

These steps will produce a bullet-pointed list. To produce a numbered list, add `@n` to each `<item>/<head>` element, with whatever sequential value is required under the chosen system (letters, Arabic numerals, Roman numerals, etc.). For example, the two-step process set out at the beginning of this section would be marked up as in Figure 60.

Figure 60

<code><list></code>
<code><item n="1"></code> Add a <code>&lt;list&gt;</code> element; <code></item></code>
<code><item n="2"></code> For each item, add an <code>&lt;item&gt;</code> element as the child of <code>&lt;list&gt;</code> , with the text as its content. <code></item></code>
<code></list></code>

²⁷ The angle brackets in the list items have been “escaped” (in this instance, rendered as HTML codes) to avoid confusing the processing software, which may otherwise pick them up as part of the XML. For simplicity, the hyperlinks from the original list have not been included here or in Figure 60.

@n can be omitted from any element. For example, it can be omitted from <[head](#)> and the numbering restarted in the new section of the list.

Tables

Adding a table in TEI/XML is somewhat complex, as one needs to map out the table using simply an element hierarchy rather than any sort of visual interface.

6. A new paragraph (<p>) should be created for the table;
7. Add a <table> element as child of <p>, which will enclose the entire table;
8. If the table is to have column headers, add a <row> element as child of <table>, which will enclose the header row of the table. It should have the attribute: @role: "label";
9. For each column in the table, add a <cell> element as a child of <row role="label">; each <cell> element should have the attribute @role: "data" and contain as text content the header for each column (if applicable);
10. For each data row in the table, add a <row> element with @role "data", then, for each column in that row, add a <cell role="data"> element, with the appropriate data as its text content.

A header row (<row role="label">) can be added at any point, with a new set of column headers. The number of columns can also be changed within the table. In a <cell> element, add the attribute @cols, with the value being the number of columns the cell should take up.

Thus, if a table began with five columns, the code in Figure 61 would begin a new section of the table with only two (wider) columns, each with a new heading:

Figure 61

```
<row role="label">
    <cell role="label" cols="1">
        <!-- column heading here -->
    </cell>
    <cell role="label" cols="4">
        <!-- New column heading here -->
    </cell>
</row>
```

Of course, it may well prove simpler to add a new table in a separate <p> element.

As an example, Figure 62 presents a table summarising the texts used as examples in this manual:

Figure 62

```
<p>
    <table>
        <row role="label">
            <cell role="label">Manuscript Shelfmark</cell>
            <cell role="label">Text Title</cell>
            <cell role="label">Form</cell>
            <cell role="label">Corpus Transcription Number</cell>
        </row>
        <row role="data">
            <cell role="data">NLS Adv. 72.1.36</cell>
```

```

<cell role="data"><hi rend="italics">Bruighion Bheag na
hAlmhain</hi></cell>
<cell role="data">Prose</cell>
<cell role="data">12.2</cell>
</row>
<row role="data">
    <cell role="data">NLS Adv. 72.1.36</cell>
    <cell role="data">' Is maith mo leaba, is olc mo
shúain'</cell>
    <cell role="data">Verse</cell>
    <cell role="data">12.4</cell>
</row>
<row role="data">
    <cell role="data">NLS Adv. 72.1.4</cell>
    <cell role="data">Definitions</cell>
    <cell role="data">Divided Prose</cell>
    <cell role="data">17.1</cell>
</row>
<row role="data">
    <cell role="data">NLS Adv. 72.2.11</cell>
    <cell role="data"><hi rend="italics">Cath
Fionntrágha</hi></cell>
    <cell role="data">Multiple Forms</cell>
    <cell role="data">3.1</cell>
</row>
<row role="label">
    <cell role="label" cols="2">Transcriber</cell>
    <cell role="label" cols="2">Text(s)</cell>
</row>
<row role="data">
    <cell role="data" cols="2">EPT</cell>
    <cell role="data" cols="2">12.2, 12.4, 17.1</cell>
</row>
<row role="data">
    <cell role="data" cols="2">MMA</cell>
    <cell role="data" cols="2">3.1</cell>
</row>
</table>
</p>

```

This renders the HTML output as shown in Figure 63.

Figure 63

Manuscript Shelfmark	Text Title	Form	Corpus Transcription Number
NLS Adv. 72.1.36	<i>Bruighion Bheag na hAlmhain</i>	Prose	12.2
NLS Adv. 72.1.36	'Is maith mo leaba, is olc mo shúain'	Verse	12.4

NLS Adv. 72.1.4	Definitions	Divided Prose	17.1
NLS Adv. 72.2.11	<i>Cath Fionntrágha</i>	Multiple Forms	3.1
Transcriber		Text(s)	
EPT		12.2, 12.4, 17.1	
MMA		3.1	

Quotations

Brief quotations can simply be placed in quote marks ("...") within the paragraph, following the usual syntactic conventions (for adding references, see REFERENCES, below). If you are quoting a transcription from the same file in brief, bear in mind the automated facility described in REFERENCES, below, as this may render quotation of context unnecessary.

Longer quotations should be placed in a `<quote>` element (child of `<p>`). This will present the quotation in a separate, indented paragraph. After the `<quote>` element has been closed, the enclosing `<p>` element should also be closed and a new `<p>` element opened if need be. The structure thus follows the *opus geminatum* format that emerges from embedding `<div>` elements in other `<div>` elements (see MULTIPLE FORMS).

Figure 64

```
<p>
    <!-- Prose commentary here -->
    <quote>
        <!-- Long quote here -->
    </quote>
</p>
<p>
    <!-- Next paragraph here -->
</p>
```

If the long quotation is in prose, it can simply be inputted as the direct text content of `<quote>`. If it is verse, then `<l>` elements can be placed as the children of `<quote>` and each line added therein.

Figure 65

```
<p>
    <!-- Prose commentary here -->
    <quote>
        <l><!-- Verse quote: line 1 --></l>
        <l><!-- Verse quote: line 2 --></l>
        <l><!-- Verse quote: line 3 --></l>
        <l><!-- Verse quote: line 4 --></l>
    </quote>
</p>
<p>
    <!-- Next paragraph here -->
</p>
```

References

Form of Citation

Citations in introductions and footnotes to FnaG MSS transcriptions are by author-date. Thus, the first page of the following work,

William Gillies, ‘The Gaelic of Niall MacMhuirich’, *Transactions of the Gaelic Society of Inverness* 65 (2011), 69–95,

would be cited as follows: “Gillies 2011a: 69”.

If, as is the case here, author + date does not uniquely identify the work in question within the corpus (i.e. the author has multiple publications in that year that appear in the FnaG MSS Project bibliography), letters of the alphabet should be added to the dates in the citations to render them unique.

Citations of manuscripts that appear in the corpus should be by their corpus code (e.g. “MS12”). Citations of other manuscripts should be in the shortest intelligible form. Citations of texts from the corpus can be by title or by corpus number (e.g. “Bruighion Bheag na hAlmhain” or “MS12.2”).

Mark-Up

When you cite a manuscript, a piece of primary or secondary literature, or the transcription you are introducing, or a known scribal hand, a reference can be added to it within the XML document as well as by the conventional, human-readable means. This allows further details to be retrieved and displayed at the point at which the citation is made, for the convenience of users.

Enclose the key component of the reference (e.g. “2011a”, in the above example) in a `<ref>` element, with the following attributes:

`@type`: the type of reference being made, of the following values:

“bib”: A piece of primary or secondary literature listed in the project bibliography;

“ms”: A manuscript listed in the project bibliography;

“text”: A text transcribed within the present transcription file (i.e. a `<div>`);

“hand”: A hand registered within the corpus;

“text_ed_line”: A line from a text in the present transcription file, according to an editorially analysed structure (e.g. a line of verse);

“text_dip_line”: A manuscript line from a text in the present transcription file;

`@target`: the `@xml:id` for whatever is being cited.

Care should be taken to ensure that the right `@type` is used for the right citation, particularly in the case of “text_ed_line” versus “text_dip_line”.

For `@target`, the id should be sought:

- In `<listBibl>` in `corpus.xml` for “ms” and “bib”;
- In `<msContents>` in the present file for “text”;
- In `<handNotes>` in `corpus.xml` for “hand”;

- Within the marked-up text for “text_ed_line” (<1>) and “text_dip_line” (<1b>).

The result, from the above example, would be:

Gillies <ref type="bib" target="Gillies2011a">2011a</ref>: 69

As a result, in the HTML output, clicking on the reference will generate inline:

- A long version of the citation in the case of “bib” or “ms”;
- Further details in the case of “text” or “hand”;
- A copy of the metrical or topic-based line in the case of “text_ed_line”;
- A copy of the manuscript line in the case of “text_dip_line”.

This also allows the citations of manuscripts and primary and secondary literature in each file to be tracked and relevant bibliographies to be generated automatically in each instance.

Below, Figure 66 shows a paragraph from the introduction to T2 with some inline citations of secondary literature and manuscripts as they first appear. Figure 67 shows all the citations expanded.

Figure 66

Other Activities

The same scribe is also the sole hand in the first fragment (fol. 1–5) bound into the composite manuscript, Edinburgh, NLS Adv., MS 72.1.31. This fragment preserves at least 13 poems in his hand about or attributed to Colum Cille, although its poor condition frustrates close examination of many of them. The genre as whole has been examined in detail by Herbert (2005). Four of these poems, as well as a text of ‘Aingil De dom dhin’, appear in Oxford, Bodleian Library, MS. Laud Misc 615, which Herbert and O’Sullivan (1973: 175) date to 1532–1545 and associate with the circle of scholars working on Columban material for the Ó Neills of Tyrone, NW Ireland. Furthermore, the page layout of the first fragment in NLS Adv. 72.1.31 — oblong pages with long lines, often one quatrain per line — resembles that of Laud Misc. 615. Indeed, the quatrains of ‘Aingil De dom dhin’ in Egerton 2899 are each also set out one to a line. This might suggest that the present scribe had intellectual contact with scholars in the north of Ireland and drew on similar literary resources.

Add Comment

Figure 67

Other Activities

The same scribe is also the sole hand in the first fragment (fol. 1–5) bound into the composite manuscript, Edinburgh, NLS Adv., MS 72.1.31 [Edinburgh, NLS Adv., MS. 72.1.31]. This fragment preserves at least 13 poems in his hand about or attributed to Colum Cille, although its poor condition frustrates close examination of many of them. The genre as whole has been examined in detail by Herbert (2005) [Herbert (2005), Maire, ‘Becoming an Exile: Colum Cille in Middle-Irish Poetry’, in Joseph F. Nagy (ed.) and L. E. Jones (ed.), *Heroic Poets and Poetic Heroes in Celtic Tradition: A Festschrift for Patrick K. Ford*, CSANA Yearbook 3–4, Dublin: Four Courts Press, 131–40]. Four of these poems, as well as a text of ‘Aingil De dom dhin’, appear in Oxford, Bodleian Library, MS. Laud Misc 615, which Herbert and O’Sullivan (1973) [Herbert (1973), Maire, and Anne O’Sullivan, ‘The Provenance of Laud Misc. 615’, *Celtica* 10, 174–92] — oblong pages with long lines, often one quatrain per line — resembles that of Laud Misc. 615. Indeed, the quatrains of ‘Aingil De dom dhin’ in Egerton 2899 [London, British Library, MS. Egerton 2899] are each also set out one to a line. This might suggest that the present scribe had intellectual contact with scholars in the north of Ireland and drew on similar literary resources.

Add Comment

Figure 68 shows a paragraph, also from T2, on some of Hand7’s traits. Figure 69 shows the citations of the text, some of which are to the editorial text and some of which are to the diplomatic text, in their expanded form.

Figure 68

Other traits

- As noted, the scribe not infrequently uses pronunciation spellings (e.g. “na b[p]earn” (q20b)). On other occasions, however, the forms that he renders imply a misunderstanding of the text’s sense. For example, he appears to be referring to an eighth (otherwise unknown) archangel when he writes “scaphel blath | a bharr” (fol. iii¹5–6; “Scaphael, fine his helmet”); other manuscript witnesses have “scathfhial” (“sheltering veil”) at this point. In another example, he suspiciously repeats “uriel” (fol. iii¹10) in the verse on Archangel Uriel at a point where an adjective would be expected (= “ur[fh]iel”, air- + fial?), although there is confusion regarding this line throughout the manuscript witnesses (O’Nolan 1912: 255).

Add Comment

Figure 69

Other traits

- As noted, the scribe infrequently uses pronunciation spellings (e.g. "na b[p]jeann" (q20b

[parthalon na bpeann](#)

)). On other occasions, however, the forms that he renders imply a misunderstanding of the text's sense. For example, he appears to be referring to an eighth (otherwise unknown) archangel when he writes "scaphel blaith | a bharr" (fol.iii^f5

[airCor dibh michelmora gloir nigloir andraphael doraidh rumscaphelblaith](#)

-6

[a bharrCor dibhuriel ámrénacgadromfeithcor dibh sariel smuthbanniam de aradhreich](#)

; "Scaphael, fine his helmet"); other manuscript witnesses have "scathfhal" ("sheltering veil") at this point. In another example, he suspiciously repeats "uriel" (fol.iii^f10

[Uriel fial na feart](#)[dibh uriel](#) na dlochtsachedain famchorpnaeg again anto

) in the verse on Archangel Uriel at a point where an adjective would be expected (= "ur[fb]iel", *air-* + *fial*?), although there is confusion regarding this line throughout the manuscript witnesses (O'Nolan 1912: 255).

[Add Comment](#)

The text content of <[ref](#)> is not relevant from a processing point of view; the key data is contained in its attributes. Thus, any sort of reference to the resource or entity in question can be enclosed in <[ref](#)>, not just a formal citation. Indeed, <[ref](#)> might be useful for providing clarification under such circumstances.

Automated Features

It should be noted that the transcription file header, once processed using stylesheet-ed-pub.xsl, will be slightly different from the raw TEI/XML:

- The sections will be in a slightly different order;
- Some generic material will have been imported from corpus.xml;
- Some additional data will have been generated automatically; e.g.:
 - A word count will be calculated in <[fileDesc](#)>;
 - A list of hands active in the text, and their respective roles, will be added for each <[msItem](#)>;
 - A *Works Cited* list, generated from <[ref](#)> elements in the file, will be added as the last child of <[body](#)>.

It is adviseable to make each section of the transcription file header as self-contained, in terms of content, as possible, while avoiding gross repetition of material. This is not only because the header will be re-structured in the output but because different sections of the header might end up being viewed in isolation from each other, depending on how the corpus is eventually made available.

That being said, it is also adviseable to acquaint oneself with the structure of the transcription file header output before planning the content for the input. More introductions to files in the corpus will be made available in the near future and these can be studied to this end.

Review

- Each transcription file must have an introduction (<[teiHeader](#)>) covering the literary, historical, and linguistic context of the file's contents.
- The content of the introductions is set out in APPENDIX 5b and in instructions from the Manuscript Consultant.
- Various facilities are available to replicate scholarly apparatus in the transcription file introductions.
- Some of these will generate interactive reading aids in the HTML output.
- The HTML output will re-structure the introduction, as authored in the original TEI/XML file.

Appendices

Appendix 1

ID	Form	Lemma	POS	Scribe	Date	Ref.	Context
Definite Article							
don	<u>do</u>	prep, art	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	<i>ms. line 1</i>	Soraidh slan don aoidche [a]reir :
don	<u>do</u>	prep, art	Eoghan Mac Gilleoin (9)	17	MS12.85v.13		soraidh slan don naoidche [a]reir
san	<u>i</u>	prep, art	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)		agus tu san chul ud thail
in	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)		ar nach ceillin in rosc rún
an	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	<i>ms. line 5</i>	Truagh an chuibhreach do ni an chiall :
an	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	<i>ms. line 5</i>	Truagh an chuibhreach do ni an chiall :
na	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.5		re silladh suirghech na súl
an	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line 6</i>	ni ferda an tost do ni an beil :
an	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line 6</i>	ni ferda an tost do ni an beil :
na	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	<i>ms. line 8</i>	Cho leig egra lucht na mbrég :
an	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	<i>ms. line 10</i>	Cuinamh dhuine an aoidhche anocht :
an	<u>in</u>	art	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	<i>ms. line 11</i>	na leig an mhaidin astech :

Nouns						
mhuire	Muire	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	<i>ms. line 12</i> Och a mhuire a bhuime tseng :
aoidche	adaig	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	<i>ms. line 1</i> Soraidh slan don aoidche [a]reir :
aoidhche	adaig	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	<i>ms. line 10</i> Cuinamh dhuine an aoidhche anoch t :
naoidche	adaig	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	soraidh slan don naoidche [a]reir
ansa	ansae	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line 2</i> is da mbeith ansa dom cuir a ccroich :
brath	bráth	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
bhuime	buime	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	<i>ms. line 12</i> Och a mhuire a bhuime tseng :
beil	bél	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	<i>ms. line 4</i> giodh nach bfuliid beil ar bheil .
bheil	bél	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	<i>ms. line 4</i> giodh nach bfuliid beil ar bheil .
beil	bél	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line 6</i> ni ferda an tost do ni an beil :
bheil	bél	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
cean	cenn	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	os tu is cean as gach cleir
cleir	clár	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	os tu is cean as gach cleir

ccroich	croch	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line 2</i> is da mbeith ansa dom cuir a ccroich :
chuibhr each	cuimrec h	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	<i>ms. line 5</i> Truagh an chuibhreach do ni an chial : :
chiall	cíall	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	<i>ms. line 5</i> Truagh an chuibhreach do ni an chiall : :
gcuil	cúl	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	fada gear ag dul ar gcuil
chul	cúl	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	agus tu san chul ud thail
dhuine	duine	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	<i>ms. line 10</i> Cuinamh dhuine an aoidhche anocht :
dias	días	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata días istighsi anocht :
dias	días	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.7	Ata dias istighsi anocht
ebla	ecla	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	<i>ms. line 8</i> Cho leig ebla lucht na mbrég :
lucht	lucht	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	<i>ms. line 8</i> Cho leig ebla lucht na mbrég :
lá	lá	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	eiridh is cuir amach lá
lamh	lám	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	<i>ms. line 13</i> tarthuidh agus gabh mo lamh :
mhaidin	matan	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	<i>ms. line 11</i> na leig an mhaidin astech :

mbrég	noun	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	<i>ms. line 8</i> Cho leig egla lucht na mbrég :
rosc	rosc	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	ar nach ceillin in rosc rún
rosc	rosc	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún
rosc	rosc	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
rún	rún	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún
sgeil	scél	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún
smid	smid	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
Soraidh	soraid	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	<i>ms. line 1</i> Soraidh slan don aoidche [a]reir :
soraid	soraid	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.14	<i>ms. line 14</i> soraid slan
súl	sul	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	re silladh suirghech na súl
súl	súil	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	is géir géir a silladh súl
shuil	súil	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>ms. line 9</i> achd tuigsi ani adéir mo shuil :
tigh	tech	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata dias istighsi anocht :

tigh	tech	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.7	Ata dias istighsi anocht
tost	tost	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line 6</i> ni ferda an tost do ni an beil :
tús	tús	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	is truagh nach é anocht a tús

Pronouns

dom	do	prep, pron	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line</i> zis da mbeith ansa dom cuir a ccroich :
ani	ní	pron	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>ms. line 9</i> achd tuigsi ani adéir mo shuil :
si	si	pron	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata dias istighsi anocht :
si	si	pron	Eoghan Mac Gilleoin (9)	17	MS12.85v.7	Ata dias istighsi anocht
sin	sin	dpron	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
so	so	dpron	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
tu	tu	pron	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	agus tu san chul ud thaill
tu	tú	pron	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	os tu is cean as gach cleir
é	é	pron	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	is truagh nach é anocht a tú
ud	út	dpron	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	agus tu san chul ud thaill

Possessive Pronouns						
a	<u>a</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	is truagh nach é anocht a tús
a	<u>a</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	is géir géir a silladh súl
a	<u>a</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún
a	<u>a</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún
a	<u>a</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
mo	<u>mo</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
mo	<u>mo</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>ms. line 9</i> achd tuigsi ani adéir mo shuil :
mo	<u>mo</u>	poss	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	<i>ms. line 13</i> tarthuidh agus gabh mo lamh :
Adjectives						
gach	<u>cach</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	os tu is cean as gach cleir
ferda	<u>ferda</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line 6</i> ni ferda an tost do ni an beil :
fada	<u>fota</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	fada gear ag dul <i>ar</i> gcuil
gear	<u>gér</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	fada gear ag dul <i>ar</i> gcuil

géir	<u>gér</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	is géir géir a silladh súl
géir	<u>gér</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	is géir géir a silladh súl
mall	<u>mall</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
rún	<u>rún</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	ar nach ceillin in rosc rún
tseng	<u>seng</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	<i>ms. line</i> 12 Och a mhuire a bhuime tseng :
slan	<u>slán</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	<i>ms. line</i> 1 Soraidh slan don aoidche [a]reir :
slan	<u>slán</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	soraidh slan don naoidche [a]reir
slan	<u>slán</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.14	<i>ms. line</i> 14 soraidh slan
soraidh	<u>soraid</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	soraidh slan don naoidche [a]reir
suirgh ech	<u>suirgech</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	re silladh suirgech na súl
truagh	<u>trúag</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	is truagh nach é anocht a túis
Truagh	<u>trúag</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	<i>ms. line</i> 5 Truagh an chuibreach do ni an chial : :
truag	<u>trúag</u>	adj	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
Numbers						

Attá						
mbeith	attá	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line</i> 2 is da mbeith ansa dom cuir a ccroic h :
Ata	attá	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata dias istighsi anocht :
bfaillid	attá	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	<i>ms. line</i> 4 giordh nach bfaillid beil ar bheil .
Ata	attá	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.7	Ata dias istighsi anocht
Is						
is	is	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	is truagh nach é anocht a túis
is	is	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	is géir géir a silladh súl
ni	is	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line</i> 6 ni ferda an tost do ni an beil :
is	is	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	os tu is cean as gach cleir
Verbs						
adéir	as-béir	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>ms. line</i> 9 achd tuigsi ani adéir mo shuil :
ceillin	ceilid	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	ar nach ceillin in rosc rún
Cuina mh	cuimnid	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	<i>ms. line</i> 10 Cuinamh dhuine an aoidhche an ocht :
cuir	cuirid	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line</i> 2 is da mbeith ansa dom cuir a ccroich :

cuir	cuirid	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	eiridh is cuir amach lá
giodh	cía	conj, verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	<i>ms. line</i> 4 giodh nach bfuilid beil <i>ar bheil</i> .
ni	do-gní	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	<i>ms. line</i> 5 Truagh an chuibhreach do ni an chiall :
ni	do-gní	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line</i> 6 ni ferda an tost do ni an beil :
ni	do-gní	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún
gabh	gaibid	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	<i>ms. line</i> 13 tarthuidh agus gabh mo lamh :
leig	léicid, léig	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	<i>ms. line</i> 8 Cho leig egla lucht na mbrég :
leig	léicid, léig	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	<i>ms. line</i> 11 na leig an mhaidin astech :
tarthuidh	tárrthaigid	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	<i>ms. line</i> 13 tarthuidh agus gabh mo lamh :
eiridh	éirgid	verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	eiridh is cuir amach lá

Verbal Nouns

dul	dul	vnoun	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	fada gear ag dul ar gcuil
silladh	sillid	vnoun	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	is géir géir a silladh súl

silladh	sillid	vnoun	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	re silladh suirghech na súl
tuigsi	tuicse	vnoun	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>ms. line</i> gachd tuigsi ani adéir mo shuil :

Participles

Adverbs

amach	immach	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	eiridh is cuir amach lá
anocht	innocht	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	is truagh nach é anocht a túis
anocht	innocht	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata dias istighsi anocht :
anocht	innocht	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.7	Ata-dias-istighsi-anocht
anocht	innocht	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	<i>ms. line</i> <i>10</i> Cuinamh dhuine an aoidhche anocht :
[a]reir	irráir	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	<i>ms. line</i> <i>1</i> Soraidh slan don aoidche [a]reir :
[a]reir	irráir	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	soraidh slan don naoidche [a]reir
astech	istech	adv	Eoghan Mac	17	MS12.85v.11 (q.5)	<i>ms. line 11</i> na leig an m'haidin astech :

			Gilleoin (9)			
na	<u>ná</u>	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.11 (q.5)	<i>ms. line 11 na</i> leig an mhaidin astech :
Cho	<u>nícon,</u> <u>cha</u>	adv, part	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	<i>ms. line</i> <i>g Cho</i> leig egla lucht na mbrég :
thaill	<u>tall</u>	adv	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>agus tu san chul ud thaill</i>

Conjunctions

ar	<u>ar</u>	conj	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	<i>ar</i> nach ceillin in rosc rún
giodeh	<u>cía</u>	conj, verb	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	<i>ms. line 4 giodeh</i> nach bfuilid beil ar bheil .
nach	<u>nach</u>	conj	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>is truagh nach</i> é anocht a túis
is	<u>ocus</u>	conj	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line 2 is</i> da mbeith ansa dom cuir a ccroich :
agus	<u>ocus</u>	conj	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>agus</i> tu san chul ud thaill
is	<u>ocus</u>	conj	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	<i>is</i> truag gan sin mar so go brath
is	<u>ocus</u>	conj	Eoghan Mac	17	MS12.85v.11 (q.5)	eiridh <i>is</i> cuir amach lá

			Gilleoin (9)			
agus	<u>ocus</u>	conj	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	<i>ms. line 13</i> tarthuidh agus gabh mo lamh :
Prepositions						
achd	<u>acht</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	<i>ms. line 9</i> achd tuigsi ani adéir mo shuil :
ar	<u>ar</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	fada gear ag dul ar gcuil
ar	<u>ar</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	<i>ms. line 4</i> giadh nach bfuilid beil ar bheil .
ar	<u>ar</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún
ar	<u>ar</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
gan	<u>cen</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
go	<u>co</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
don	<u>do</u>	prep, art	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	<i>ms. line 1</i> Soraidh slan don aoidche [a]reir :
dom	<u>do</u>	prep, pron	Eoghan Mac	17	MS12.85v.2 (q.1)	<i>ms. line 2</i> is da mbeith ansa dom cuir a ccroich :

			Gilleoin (9)			
don	<u>do</u>	prep, art	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	soraidh slan don naoidche [a]reir
da	<u>dá n-</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line 2</i> is da mbeith ansa dom cuir a ccroich :
re	<u>fri</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	re silladh suirghech na súl
a	<u>i</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.2 (q.1)	<i>ms. line 2</i> is da mbeith ansa dom cuir a ccroich :
is	<u>i</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata dias istighsi anocht :
is	<u>i</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.7	Ata dias istighsi anocht
san	<u>i</u>	prep, art	Eoghan Mac Gilleoin (9)	17	MS12.85v.9 (q.4)	agus tu san chul ud thaill
mar	<u>immar</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
ag	<u>oc</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	fada gear ag dul ar gcuil
as	<u>ás</u>	prep	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	os tu is cean as gach cleir

Particles									
a	a	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	<i>ms. line 12</i> Och a m̄huire a bhuime tseng :			
a	a	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	<i>ms. line 12</i> Och a m̄huire a bhuime tseng :			
do	do	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.5	<i>ms. line 5</i> Truagh an chuibhreach do ni an chiall :			
do	do	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	<i>ms. line 6</i> ni ferda an tost do ni an beil :			
do	do	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.6	sgeil do ni a rosc ar a rún			
nach	nach	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.3 (q.2)	ar nach ceillin in rosc rún			
nach	nach	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.4 (q.2)	<i>ms. line 4</i> giordh nach bfuilid beil ar bheil .			
Cho	nícon, cha	adv, part	Eoghan Mac Gilleoin (9)	17	MS12.85v.8 (q.4)	<i>ms. line 8</i> Cho leig egla lucht na mbrég :			
os	os	part	Eoghan Mac Gilleoin (9)	17	MS12.85v.12	os tu is cean as gach cleir			
Prefixes									
Unidentified									

Appendix 2

Form	Lemma	Part of Speech	Scribe	Date	Reference	Context
oidhchi	adaig	noun	Eoghan Mac Pháil (8)	17	MS6.8r.8	{at}a d fiacuibh <i>ar</i> an oglach donnchad[h] d ain[fh]iaca a gcuid oidhchi ocus d fiacuibh <i>ar</i> donnch[adh]
oidche	adaig	noun	Niall MacMhuirich (10)	17/18	MS9.3r.12	a sódail 1615 . an bliaduin oidche feill
noidhche	adaig	noun	Niall MacMhuirich (10)	17/18	MS9.5v.9	mairmel le hamhus noidhche <i>ar</i>
aoidche	adaig	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.1 (q.1)	<i>ms. line 1</i> Soraidh slan don aoidche [a]reir :
aoidhche	adaig	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.10 (q.5)	<i>ms. line 10</i> Cuinamh dhuine an aoidhche anocht :
naoidche	adaig	noun	Eoghan Mac Gilleoin (9)	17	MS12.85v.13	soraidh slan don naoidche [a]reir
oidche	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.104va.26	oidche agus thegmaidh hi fein agus esca
aidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.105va.17	is comtrom la agus aidhchi ann Et ader
aidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.105va.25	aidhchi Et thecmad hi agus esga an
aidheamail	*aidchemail	adj	Cairbre Ó Ceanmain (19)	16	MS13.105vb.2	uiscemail aidheamail Et is iad
aidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.105vb.15	an .9. huair d eis medoin aidhchi
aidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.106ra.9	uair d eis meadoin aidhchi
nodhci	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107rb.2	do uenir sa lo agus do luna sa nodhci
oidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107rb.6	oidhchi do mercur Et adeir
aidhcheamail	*oidchemail	adj	Cairbre Ó Ceanmain (19)	16	MS13.107rb.19	uisceamail bannda aidhcheamail agus is e is

noidhche	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107rb.23	noidhche ms. p. 107v col. a
oidhch	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.3	iubiter sa oidhch
aidh[ch]email	* oidchemail	adj	Cairbre Ó Ceanmain (19)	16	MS13.107va.4	Uirgo feramail aidh[ch]email // co hinmedonach \\ agus is e is
oidh _{ms. line 8} Chi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.7	lae do uenir agus is teghais oidh_{ms. line 8} Chi do luna
n _{ms. line 11} oidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.10	sa lo and agus mercuir sa n_{ms. line 11} oidhchi
noidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.13	sa noidhchi
oidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.13	// ann agus agus ioib san oidhchi
noidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.16	sa lo ann agus luna sa noidhchi Et is
noidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.21	sa noidhchi Et do ni { lurga } i[s] sia
noidhchi	adaig	noun	Cairbre Ó Ceanmain (19)	16	MS13.107va.24	sa lo and agas mars sa noidhchi

Appendix 3

ID	Lemma	Stem	Gender	Date	Reference	Context
sga = Old Gaelic mga = Middle Gaelic hsg = Hiberno-Scottish Gaelic (i.e. Classical Gaelic) gd = Scottish Gaelic ga = Irish Gaelic.						
1	adaig	iā	f	17	MS12.85v.1 (q.1)	<i>ms. line 1</i> Soraidh slan don aoidche [a]reir :
2	adaig	iā	f	17	MS12.85v.10 (q.5)	<i>ms. line 10</i> Cuinamh dhuine an aoidche anocht :
3	adaig	iā	f	17	MS12.85v.13	soraidh slan don naoidche [a]reir
4	bél	o	m	17	MS12.85v.4 (q.2)	<i>ms. line 4</i> giodh nach bfuilid beil ar bheil .
5	bél	o	m	17	MS12.85v.4 (q.2)	<i>ms. line 4</i> giodh nach bfuilid beil ar bheil .
6	bél	o	m	17	MS12.85v.6	<i>ms. line 6</i> ni ferda an tost do ni an beil :
7	bél	o	m	17	MS12.85v.8 (q.4)	smid ar mo bheil a rosc mall
8	bráth	u (sga); o (sga).	m	17	MS12.85v.10 (q.5)	is truag gan sin mar so go brath
9	cenn	o	n (sga); m (mga); m (gd).	17	MS12.85v.12	os tu is cean as gach cleir
10	croch	ā	f	17	MS12.85v.2 (q.1)	<i>ms. line 2</i> is da mbeith ansa dom cuir a croich :
11	cúl	o	m	17	MS12.85v.1 (q.1)	fada gear ag dul ar gcuil
12	cúl	o	m	17	MS12.85v.9 (q.4)	agus tu san chul ud thaill
13	días	ā	f	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata días istighsi anocht :
14	días	ā	f	17	MS12.85v.7	Ata días istighsi anocht
15	duine	io	m	17	MS12.85v.10 (q.5)	<i>ms. line 10</i> Cuinamh dhuine an aoidhche anocht :
16	dul		m	17	MS12.85v.1 (q.1)	fada gear ag dul ar gcuil
17	ecla	iā	f	17	MS12.85v.8 (q.4)	<i>ms. line 8</i> Cho leig ecla lucht na mbréig :
18	lá	io	n (sga); m (mga); m (gd).	17	MS12.85v.11 (q.5)	eiridh is cuir amach lá

19	<u>lám</u>	ā	f	17	MS12.85v.13	<i>ms. line 13</i> tarthuidh agus gabh mo lamh :
20	<u>lucht</u>	u	m	17	MS12.85v.8 (q.4)	<i>ms. line 8</i> Cho leig egl a lucht na mbréag :
21	<u>scél</u>	o	n (sga); m (gd); f (gd).	17	MS12.85v.6	sgeil do ni a rosc ar a rún
22	<u>soraid</u>	(); i () .	f	17	MS12.85v.1 (q.1)	<i>ms. line 1</i> Soraidh slan don aoidche [a]reir :
23	<u>soraid</u>	(); i () .	f	17	MS12.85v.14	<i>ms. line 14</i> soraid slan
24	<u>súil</u>	i	f	17	MS12.85v.4 (q.2)	is géir géir a silladh súl
25	<u>súil</u>	i	f	17	MS12.85v.9 (q.4)	<i>ms. line 9</i> achd tuigsi ani adéir mo shuil :
26	<u>tech</u>	s	n (sga); m (mga); m (gd).	17	MS12.85v.3 (q.2)	<i>ms. line 3</i> Ata dias istighsi anocht :
27	<u>tech</u>	s	n (sga); m (mga); m (gd).	17	MS12.85v.7	Ata dias istighsi anocht
28	<u>tús</u>	u (sga); o (mga).	m	17	MS12.85v.2 (q.1)	is truagh nach é anocht a tús

Appendix 4

Words in Corpus:	18036	
Individual Lexical Items in Corpus:	3591	
Mean occurrence rate of an individual lexical item:	5.02	
Names in Corpus:	1541	
Unidentified Words in Corpus:	98	
Proportion of corpus "unclear":	5.48%	
Prose:	77.7%	
Verse:	19.66%	
Material by Beatons:	5.81%	
Material by Mac Mhuirichs:	30.81%	
Words rendered using the most abbreviations (top 5):	<ol style="list-style-type: none"> 1. <i>fiabrusaibh</i> (6) 2. <i>hinnarrbha\ nadh /</i> (6) 3. <i>cumhsanadh</i> (6) 4. <i>chomhghraidh</i> (5) 5. <i>chomhdhail</i> (5) 	
Scribe	Date	Word Count
[unk.] [unk.] (Hand6)	10 th cent.	28
[unk.] [unk.] (Hand1)	12 th cent.	448
[unk.] [unk.] (Hand2)	12 th cent.	112
[unk.] [unk.] (Hand3)	12 th cent.	87
[unk.] [unk.] (Hand4)	12 th cent.	128
[unk.] [unk.] (Hand5)	12 th cent.	0
Dubhghall Albanach mac mhic Cathail (Hand18)	15 th cent.	1859
[unk.] MacEwen (?) (Hand7)	15/16 th cent.	880
[unk.] [unk.] (Hand12)	16 th cent.	1112
Lachlan Mór MacLean (Hand13)	16 th cent.	72
Cairbre Ó Ceanmain (Hand19)	16 th cent.	2384

Eoin mac Domhnaill Ó Conchubhair (Hand20)	16 th cent.	1120
Domhnall Gorm ?Mac Domhnaill (Hand21)	16/17 th cent.	24
Eoghan Mac Pháil (Hand8)	17 th cent.	777
Eoghan Mac Gilleoin (Hand9)	17 th cent.	578
James Beaton (Hand14)	17 th cent.	113
[unk.] [unk.] (Hand15)	17 th cent.	4
Rev. John Beaton (Hand16)	17 th cent.	525
Neil Beaton (Hand17)	17 th cent.	424
Cathal Mac Mhuirich (Hand22)	17 th cent.	394
Niall MacMhuirich (Hand10)	17/18 th cent.	4851
Domnall MacMhuirich (Hand11)	18 th cent.	322
Alasdair Mac Mhaighstir Alasdair (Hand23)	18 th cent.	1867
an unidentified hand (Hand999)	unknown	2

Appendix 5

Appendix 5a: Transcription File Template

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <?xml-model href="../fnag_mss2.rng" type="application/xml"
3  schematypens="http://relaxng.org/ns/structure/1.0"?>
4  <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="TN">
5      <teiHeader>
6          <fileDesc>
7              <titleStmt>
8                  <title>Transcription N</title>
9                  <respStmt corresp="WGI">
10                     <name>William Gillies</name>
11                     <resp>Reviewer</resp>
12                 </respStmt>
13             </titleStmt>
14             <extent/>
15             <publicationStmt>
16                 <authority sameAs="authority"/>
17                 <availability status="restricted" sameAs="availability">
18                     <licence/>
19                 </availability>
20                 <idno/>
21             </publicationStmt>
22             <notesStmt>
23                 <note>
24                     <p>
25                 </note>
26             </notesStmt>
27             <sourceDesc>
28                 <msDesc>
29                     <msIdentifier sameAs="MSN">
30                         <settlement/>
31                         <repository/>
32                         <idno type="shelfmark"/>
33                     </msIdentifier>
34                     <msContents>
35                         <summary>
36                             <p>
37                         </summary>
38                         <msItem n="itemN" xml:id="itemID" resp="HandXXX">
39                             <locus/>
40                             <title/>
41                             <incipit/>
42                             <explicit/>
43                             <note>
44                                 <p comment="Content"/>
45                                 <p comment="Style"/>
46                                 <p comment="Social Context and Intended
47                                     Readers"/>
48                                 <p comment="Studies and Editions"/>
49                             </note>
50                             <filiation>
51                                 <p/>
52                             </filiation>
53                             <textLang mainLang="XXX" otherLangs="XXX">
54                                 <note comment="Orthography and Scribal
55                                     Practices">
56                                     <p/>
57                                     <p>
58                                         <list>
59                                             <head>Spelling System</head>
60                                             <item/>
61                                         </list>
62                                         <list>
63                                             <head>Scribal Practices</head>
64                                             <item/>
```

```

65                                </list>
66                            </p>
67                        </note>
68                        <note comment="Language">
69                            <p comment="Nominal System">
70                                <list>
71                                    <item/>
72                                </list>
73                            </p>
74                            <p comment="Verbal System">
75                                <list>
76                                    <item/>
77                                </list>
78                            </p>
79                            <p comment="Pronouns and Pronominals">
80                                <list>
81                                    <item/>
82                                </list>
83                            </p>
84                            <p comment="Syntax">
85                                <list>
86                                    <item/>
87                                </list>
88                            </p>
89                            <p comment="Lexis">
90                                <list>
91                                    <item/>
92                                </list>
93                            </p>
94                            <p comment="Conclusions">
95                                <list>
96                                    <item/>
97                                </list>
98                            </p>
99                        </note>
100                       <note comment="Style">
101                           <p/>
102                       </note>
103                       <textLang>
104                         </msItem>
105                     </msContents>
106                     <physDesc>
107                         <objectDesc>
108                             <supportDesc>
109                               <support>
110                                 <p/>
111                               </support>
112                               <foliation>
113                                 <p/>
114                               </foliation>
115                               <condition>
116                                 <p/>
117                               </condition>
118                         </supportDesc>
119                         <layoutDesc>
120                           <p/>
121                         </layoutDesc>
122                     </objectDesc>
123                     <handDesc>
124                         <handNote corresp="HandXXX">
125                             <note>
126                               <p comment="The Scribe"/>
127                               <p comment="Other Activities"/>
128                             </note>
129                         </handNote>
130                     </handDesc>
131                 </physDesc>

```

```

132             <history>
133                 <origin>
134                     <p/>
135                 </origin>
136                 <provenance>
137                     <p/>
138                 </provenance>
139                 <acquisition>
140                     <p/>
141                 </acquisition>
142             </history>
143         </msDesc>
144     </sourceDesc>
145 </fileDesc>
146 <encodingDesc>
147     <editorialDecl>
148         <p/>
149     </editorialDecl>
150     <metDecl>
151         <p/>
152     </metDecl>
153     <refsDecl>
154         <p/>
155     </refsDecl>
156 </encodingDesc>
157 <profileDesc>
158     <textClass>
159         <keywords>
160             <term/>
161         </keywords>
162     </textClass>
163 </profileDesc>
164 <revisionDesc>
165     <change when="yyyy-mm-dd" who="XXX"/>
166 </revisionDesc>
167 </teiHeader>
168 <text>
169     <body>
170         <div n="N" type="ZZZ" resp="HandXXX" corresp="NNN"/>
171     </body>
172 </text>
173 </TEI>

```

Appendix 5b: Annotated Transcription File Template

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <!-- Required for all TEI documents. -->
3  <?xml-model href="../fnag_mss2.rng" type="application/xml"
4  schematypens="http://relaxng.org/ns/structure/1.0"?>
5  <!-- Relative filepath to the schema ("fnag_mss2.rng"); it is
6  required. Update if another schema is used or if this file is
7  moved. -->
8  <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="TN">
9  <!-- Root element for this transcription file: @xmlns required
10 in all cases (identifies this as a TEI document); @xml:id = T
11 + [transcription number] (e.g. T4). This is the unique id
12 number for this file. -->
13 <teiHeader>
14 <!-- Contains the header (i.e. introduction) to this
15 transcription file. The current standard stylesheet
16 (stylesheet-ed-pub.xsl) presents the material in a different
17 order from that in which it is presented, inserts some
18 headings, formats the text of the introduction, and retrieves
19 some data automatically. The guidelines that follow take this
20 into account. -->
21 <fileDesc>
22 <!-- This broad-ranging element contains the bulk of the
23 introductory material in the header. -->
24 <titleStmt>
25 <!-- Contains key information about how the file was created.
26 -->
27 <title>Transcription N</title>
28 <!-- The title of the file (displayed to reader); N =
29 transcription number (as in <TEI>; e.g. Transcription 4). -->
30 <respStmt corresp="WGI">
31 <!-- Information on an individual who has contributed to this
32 transcription file; @corresp should contain a unique three
33 letter code designating that person, which will point to a
34 central directory of contributors in the corpus file. -->
35 <name>William Gillies</name>
36 <!-- The contributor's full name (displayed to reader). -->
37 <resp>Reviewer</resp>
38 <!-- Brief (one word) description of their contribution. -->
39 </respStmt>
40 <!-- If multiple individuals have been involved, further
41 <respStmt> elements can be added here. -->
42 </titleStmt>
43 <extent/>
44 <!-- *No action needed* Number of words in the
45 transcription(s) contained in this file; calculated
46 automatically and displayed to reader. -->
47 <publicationStmt>
48 <!-- Infromation on copyright etc. -->
49 <authority sameAs="authority"/>
50 <!-- *No action needed* This element will automatically import
51 details of the FnaG project from the corpus file. -->
52 <availability status="restricted" sameAs="availability">
53 <licence/>
54 <!-- By default, this element imports a standard statement
```

```

55      from the corpus file restricting this file to internal
56      FnaG/DASG use ("For internal use by the Faclair na
57      Gàidhlig/DASG projects."). If the status of this file is
58      different: (1) remove "@status='restricted'" from the parent
59      <availability> element; (2) add a <p> element as a child of
60      <licence>; (3) Detail the file's availability within this <p>
61      element. Removing the "restricted" status does not necessarily
62      mean the file is not still restricted, as this status is not
63      visible to the reader anyway. -->
64      </availability>
65      <idno/>
66      <!-- The unique id for this transcription file, as above (e.g.
67      "T4"); visible to reader. -->
68      </publicationStmt>
69      <notesStmt>
70      <note>
71      <p>
72      <!-- A short summary of the trasncription file: the
73      manuscript's history, what has been transcribed from it, why
74      it might be of interest. Like an Abstract. -->
75      <!-- Further <p> elements can be added to create further
76      paragraphs; however, this summary should be kept as brief as
77      possible. -->
78      </note>
79      </notesStmt>
80      <sourceDesc>
81      <!-- Contains information about the manuscript and its
82      contents. -->
83      <msDesc>
84      <!-- Co-terminous with <sourceDesc>; the sole "source" is an
85      "ms". -->
86      <msIdentifier sameAs="MSn">
87      <!-- Contains basic information designating which manuscript
88      is being transcribed in this file. @sameAs = the unique id for
89      this manuscript (MS + N, with N being the same number as in TN
90      above). This points to a central directory of manuscripts
91      stored in the corpus file. If the manuscript being transcribed
92      is new to the project, an entry for it will need to be created
93      in this directory. -->
94      <settlement/>
95      <!-- The town/city/region where the manuscript is located,
96      according to its standard designation (e.g. "Edinburgh" in
97      Edinburgh, NLS Adv. 72.1.1). A following comma will be
98      inserted automatically during processing. -->
99      <repository/>
100     <!-- The library where this manuscript is located (e.g. "NLS
101     Adv." in Edinburgh, NLS Adv., 72.1.1). A following comma will
102     be inserted automatically during processing. -->
103     <idno type="shelfmark"/>
104     <!-- The shelfmark within that library. -->
105     </msIdentifier>
106     <msContents>
107     <!-- Contains a catalogue of texts transcribed from this
108     manuscript and analyses of them. -->
109     <summary>
110     <p/>

```

```

111 <!-- A summary of the contents of the manuscript and why they
112 are included in the corpus. -->
113 <!-- Further paragraphs can be added by adding further <p>
114 elements. However, again, this summary should be kept brief. --
115 -->
116 </summary>
117 <msItem n="itemN" xml:id="itemID" resp="HandXXX">
118 <!-- Contains information/analysis of an individual text
119 transcribed from the manuscript. It corresponds to a <div>
120 element in the <body> (below). -->
121 <!-- @n = a serial number for this text within those
122 transcribed from the manuscript in manuscript order (e.g. if
123 the text is the first to be transcribed in order of appearance
124 within the manuscript, then "n="1"). If this item is a child
125 of <msItem> (i.e. a subdivision in a text), then the number
126 should be preceded by parent <msItem> @n attribute plus "."
127 (e.g. for the second subdivision of the first transcribed text
128 in the manuscript, "n="1.2"). -->
129 <!-- @xml:id = a unique id for this text. It is composed of
130 the MS unique id and @n, separated by points (e.g. if the MS
131 is MS4 and this is text @n, then "xml:id = "MS4.1"; if this
132 is MS4 and this is the second sub-division of the first
133 transcribed text, then "xml:id = "MS4.1.2"). -->
134 <!-- @resp points to an id identifying the main scribe; the
135 hand responsible for most of the text. This in turn points to
136 a central database of hands kept in the corpus file,
137 containing all known details relating to each hand. The id is
138 in the form "Hand" plus a number. (e.g. "resp="Hand18"). If
139 the manuscript is new to the corpus, the corpus file should be
140 checked for any pre-existing record of the scribe. If present,
141 the pre-existing id should be used. If not, an entry should be
142 created for the hand in the corpus file, along with a unique
143 id.-->
144 <!-- If this text contains sub-divisions (child <div>
145 elements), then corresponding child <msItem> elements can be
146 inserted here as children of an enclosing <msItem> element.
147 This might be done to represent poems in a prosimetric text,
148 for example. Any of the child elements of <msItem> set out
149 here can be added for either the parent or child <msItem>
150 element (i.e. for either the text as a whole or its
151 subdivisions). However, any can also be omitted if this seems
152 expedient. For example, it may not be necessary to conduct a
153 language study (<textLang>) of both the text overall and of
154 its individual components. If a child element of <msItem> is
155 not needed, then remove it completely.-->
156 <locus/>
157 <!-- Page/folio reference for this text, entered as a single
158 string (e.g. "fols 7v16-8r31". -->
159 <title/>
160 <!-- The title of the text (possibly a title or incipit from
161 the manuscript, possibly of the transcriber's devising. -->
162 <incipit/>
163 <!-- The first few words of the text, as it appears in this
164 manuscript. A metrical line, if poetry. -->
165 <explicit/>
166 <!-- The last few words of the text, as it appears in this

```

```

167 manuscript. A metrical line, if poetry. -->
168 <note>
169 <!-- Each of the child <p> elements define a sub-section for
170 comments on the designated topic. The @comment attribute
171 (counter-intuitively) contains the section's sub-heading.
172 Further paragraphs can be added in each section by adding
173 further <p> elements. Adding a <p> element without a @comment
174 attribute will appear without a sub-heading and thus under
175 the most recent sub-heading. <p comment="X"> will create a new
176 paragraph with the sub-heading "X". @comment attributes added
177 to <p> elements that are children of <note> will be of the same
178 level. -->
179 <p comment="Content"/>
180 <!-- Describe the content/subject-matter/narrative/argument of
181 this text. The @comment attribute (counter-intuitively)
182 contains the section's sub-heading. -->
183 <p comment="Style"/>
184 <!-- Broadly describe the literary style of this text (more
185 details will be added below). The @comment attribute (counter-
186 intuitively) contains the section's sub-heading. -->
187 <p comment="Social Context and Intended Readers"/>
188 <!-- Describe the intended readership/audience of this text
189 and the impact it is meant to have. The @comment attribute
190 (counter-intuitively) contains the section's sub-heading. -->
191 <p comment="Studies and Editions"/>
192 <!-- Note any previous work done on this text. The @comment
193 attribute (counter-intuitively) contains the section's sub-
194 heading. -->
195 </note>
196 <filiation>
197 <p/>
198 <!-- Commentary on the textual history and relationships of
199 this item. -->
200 <!-- Further paragraphs can be added by adding further <p>
201 elements. -->
202 </filiation>
203 <textLang mainLang="XXX" otherLangs="XXX"><!-- Contains
204 analysis of the item's language and orthography. The @mainLang
205 attribute is used to designate the overall language of the
206 text; it must contain a registered XML language tag (see
207 https://www.iana.org/assignments/language-subtag-registry). If a mixture of languages
208 are present, a comma-separated list of additional language
209 texts can be placed in @otherLangs. If @otherLangs is not
210 needed, remove it completely. -->
211 <note comment="Orthography and Scribal Practices">
212 <p/>
213 <!-- Introductory text to the language/orthography section,
214 summarising findings or flagging up any methodological issues.
215 -->
216 <!-- Further paragraphs can be added by adding further <p>
217 elements. -->
218 <p/>
219 <!-- <p> with <list> children is a set of lists of distinctive
220 features of spelling and other scribal practices in this text.
221 -->
222

```

```

223      <list>          <!-- Contains a bullet-
224      pointed list. Each bullet-point is created by an <item> child
225      element. -->
226      <head>Spelling System</head>
227      <!-- A heading for the bullet-point list (<list>). -->
228      <item/>
229      <!-- A bullet-point; text for the item is contained in the
230      <item> element. -->
231      <!-- Further bullet points can be added to the list by adding
232      further <item> elements. -->
233      </list>
234      <list>
235      <!-- Contains a bullet-pointed list. Each bullet-point is
236      created by an <item> child element. -->
237      <head>Scribal Practices</head>
238      <!-- A heading for the bullet-point list (<list>). -->
239      <item/>
240      <!-- A bullet-point; text for the item is contained in the
241      <item> element. -->
242      <!-- Further bullet points can be added to the list by adding
243      further <item> elements. -->
244      </list>
245      </p>
246      <!-- The above structure can be made more detailed. Instead,
247      insert a <p> element with the @comment "Spelling System", then
248      add a <list> for different aspects of the spelling system
249      (indicated in the <head> child; e.g. "Vowels", "Consonants"
250      etc.). Another <p> element will then need to be added for
251      Scribal Practices.-->
252      <!-- Paragraphs of continuous prose can also be added as <p>
253      children of <note> -->
254      </note>
255      <note comment="Language">
256      <!-- The same guidelines set out in relation to "<note
257      comment="Orthography and Scribal Practices">" apply here too.
258      -->
259      <p comment="Nominal System">
260      <list>
261      <item/>
262      </list>
263      </p>
264      <p comment="Verbal System">
265      <list>
266      <item/>
267      </list>
268      </p>
269      <p comment="Pronouns and Pronominals">
270      <list>
271      <item/>
272      </list>
273      </p>
274      <p comment="Syntax">
275      <list>
276      <item/>
277      </list>
278      </p>

```

```

279      <p comment="Lexis">
280      <list>
281      <item/>
282      </list>
283      </p>
284      <p comment="Conclusions">
285      <list>
286      <item/>
287      </list>
288      </p>
289      <!-- Paragraphs of continuous prose can also be added as <p>
290      children of <note> -->
291      </note>
292      <note comment="Style">
293      <p/>
294      <!-- More detailed commentary on the text's style is included
295      here, making reference to the preceding language analysis. -->
296      <!-- Further paragraphs can be added by adding further <p>
297      elements. -->
298      </note>
299      </textLang>
300      </msItem>
301      </msContents>
302      <physDesc>
303      <!-- Contains elements detailing the physical characteristics
304      of the manuscript. NB: This concerns now concerns the whole
305      manuscript, not a particular text. However, if large swathes
306      of the manuscript contain material of no relevance to FnaG
307      (e.g. not in Gaelic), then the particular physical
308      characteristics of those parts of the manuscript can be
309      described very briefly.-->
310      <objectDesc>
311      <!-- Contains elements detailing the materials and dimensions
312      of the manuscript. -->
313      <supportDesc>
314      <!-- Contains elements giving details on the manuscript's
315      pages. -->
316      <support>
317      <p>
318      <!-- Describe the page/folio (material, shape, dimensions
319      etc.) -->
320      <!-- Further paragraphs can be added by adding further <p>
321      elements. -->
322      </support>
323      <foliation>
324      <!-- Further paragraphs can be added by adding further <p>
325      elements. -->
326      </p>
327      <!-- Further paragraphs can be added by adding further <p>
328      elements. -->
329      </foliation>
330      <condition>
331      <!-- Describe the condition of the manuscript and flag up any
332      general issues. -->
333      <p/>
334      <!-- Further paragraphs can be added by adding further <p>

```

```

335     elements. -->
336     </condition>
337     </supportDesc>
338     <layoutDesc>
339         <!-- Describe the layout(s) used on the manuscript pages. -->
340         <p>
341             <!-- Further paragraphs can be added by adding further <p>
342             elements. -->
343         </layoutDesc>
344         </objectDesc>
345         <handDesc>
346             <!-- Contains elements detailing the hands involved in the
347             manuscript. NB: This will actually appear at the beginning of
348             the introduction (after the Summary) but TEI requires the
349             material to be placed here in the code. -->
350             <handNote corresp="HandXXX">
351                 <!-- Contains notes on a particular hand; @corresp points to
352                 the unique id for that hand (cf. <msItem> @resp attribute,
353                 above) -->
354             <note>
355                 <p comment="The Scribe"/>
356                     <!-- Observations on the hand's activity in this manuscript
357                     and on the scribe's career and biography, if the hand has been
358                     identified (if not, theories and possibilities can be detailed
359                     here too). The hand's orthographic characteristics will be
360                     covered in more detail in the linguistic analysis (reminder:
361                     this will *follow* this section after processing) but features
362                     of particular note can summarised here too. -->
363                     <!-- Further paragraphs can be added by adding further <p>
364                     elements. @comment, as above, contains the sub-heading. -->
365                 <p comment="Other Activities"/>
366                     <!-- Details of other manuscripts in which the hand also
367                     appears. -->
368                     <!-- Further paragraphs can be added by adding further <p>
369                     elements. @comment, as above, contains the sub-heading. -->
370             </note>
371         </handNote>
372             <!-- If multiple hands are involved, then further <handNote>
373             (+ children) elements can be added here. -->
374         </handDesc>
375         </physDesc>
376         <history> <!-- Contains elements detailing the history of the
377             manuscript. NB: After processing, this will appear after
378             <layoutDesc>. -->
379         <origin>
380             <p>
381                 <!-- The manuscript's date and place of creation. -->
382                 <!-- Further paragraphs can be added by adding further <p>
383                 elements. @comment, as above, contains the sub-heading. -->
384             </origin>
385             <provenance>
386                 <p/>
387                     <!-- Slightly abusing the term "Provenance", this section is
388                     about the manuscript's su -->
389                     <!-- Further paragraphs can be added by adding further <p>
390                     elements. @comment, as above, contains the sub-heading. -->

```

```

391      </provenance>
392      <acquisition>
393      <p/>
394      <!-- How the manuscript entered its present state of
395      ownership/preservation -->
396      <!-- Further paragraphs can be added by adding further <p>
397      elements. @comment, as above, contains the sub-heading. -->
398      </acquisition>
399      </history>
400      </msDesc>
401      </sourceDesc>
402      </fileDesc>
403      <encodingDesc>
404      <!-- Contains elements giving different sorts of information
405      on how the transcriptions are encoded and presented. -->
406      <editorialDecl>
407      <p/>
408      <!-- Notes on any special conventions that were adopted when
409      making the transcription. If not relevant, leave empty. -->
410      <!-- Further paragraphs can be added by adding further <p>
411      elements. @comment, as above, contains the sub-heading. -->
412      <!-- If different conventions were adopted in different texts
413      within the transcription, sub-headings can be used to separate
414      notes each text.-->
415      </editorialDecl>
416      <metDecl>
417      <p/>
418      <!-- If poetry, notes on metrics. If not relevant, leave
419      empty. -->
420      <!-- Further paragraphs can be added by adding further <p>
421      elements. @comment, as above, contains the sub-heading. -->
422      <!-- If multiple poems with different metres are involved,
423      sub-headings can be used to separate notes each poem.-->
424      </metDecl>
425      <refsDecl>
426      <p/>
427      <!-- Notes on any particular conventions to be adopted when
428      referencing texts in this transcription file. If not relevant,
429      leave empty-->
430      <!-- Further paragraphs can be added by adding further <p>
431      elements. @comment, as above, contains the sub-heading. -->
432      <!-- If different conventions were adopted in different texts
433      within the transcription, sub-headings can be used to separate
434      notes each text.-->
435      </refsDecl>
436      </encodingDesc>
437      <profileDesc>
438      <!-- Contains data classifying this transcription file
439      generically, in the context of the corpus. -->
440      <textClass>
441      <!-- Contains data classifying this transcription file
442      generically, in the context of the corpus. -->
443      <keywords>
444      <!-- Contains a list of keywords; possible search terms that
445      would usefully retrieve this transcription file. -->
446      <term/>
```

```

447      <!-- A keyword -->
448      <!-- Further keywords can be added by adding <term> elements
449      here. -->
450      </keywords>
451      </textClass>
452      </profileDesc>
453      <revisionDesc>
454      <!-- Keeps a record of major changes made to the transcription
455      file. -->
456      <change when="yyyy-mm-dd" who="XXX"/>
457      <!-- Details of a major revision. @when is the date of the
458      change; @who is the three letter ID for the main person
459      responsible. -->
460      </revisionDesc>
461      </teiHeader>
462      <text>
463      <!-- This element must enclose <body> for <body> to be
464      included in <teiCorpus> (the root element of the corpus file,
465      in which this file will appear via <xi:include>. -->
466      <body>
467      <!-- This element encloses the main text of the transcription,
468      as opposed to the header (<teiHeader>). -->
469      <div n="N" type="ZZZ" resp="HandXXX" corresp="NNN">
470      <!-- Each text (corresponding to an <msItem> in <teiHeader>)
471      is placed within a <div>. A <div> can be placed within another
472      <div>. @n is the serial number for the text (as in <msItem>);
473      @type can be "verse", "prose", or "divProse"; @resp contains
474      the unique id for the main scribe of this text; @corresp
475      contains the unique id for the text (<msItem> @xml:id) -->
476      <!-- From this point onwards, how the elements are structured
477      depends on the type of text. This is covered in the User
478      Documentation. -->
479      </div>
480      <!-- Subsequent texts can be added by adding further <div>
481      elements here. -->
482      </body>
483      <!-- A bibliography will be automatically generated here from
484      citations made elsewhere in the file (see User Documentation).
485      -->
486      </text>
487      </TEI>

```

Appendix 6

Appendix 6a: Corpus File Map

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <?xml-model href="file:/C:/Files/faclair-manuscripts/fnag_mss2.rng"
3  type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
4  <teiCorpus xmlns="http://www.tei-c.org/ns/1.0"
5  xmlns:xi="http://www.w3.org/2001/XInclude">
6      <teiHeader>
7          <fileDesc>
8              <titleStmt>
9                  <title>Faclair na Gàidhlig: Medieval and Early Modern Manuscript
10                 Corpus
11             </title>
12             <respStmt>
13                 <name/>
14                 <resp/>
15             </respStmt>
16         </titleStmt>
17         <publicationStmt>
18             <authority/>
19         </publicationStmt>
20         <sourceDesc>
21             <listBibl type="mss">
22                 <msDesc>
23                     <msIdentifier>
24                         <settlement/>
25                         <repository/>
26                         <idno type="shelfmark"/>
27                     </msIdentifier>
28                 </msDesc>
29             </listBibl>
30             <listBibl type="lit">
31                 <biblStruct type="book" xml:id="WWW">
32                     <monogr>
33                         <title/>
34                         <respStmt>
35                             <name>
36                                 <forename/>
37                                 <surname/>
38                             </name>
39                             <resp/>
40                         </respStmt>
41                         <imprint>
42                             <pubPlace/>
43                             <publisher/>
44                             <date/>
45                         </imprint>
46                     </monogr>
47                 </biblStruct>
48                 <biblStruct type="article" xml:id="XXX">
49                     <analytic>
50                         <title level="a"/>
51                     </analytic>
52                     <monogr>
53                         <title level="j"/>
54                         <respStmt>
55                             <name>
56                                 <forename/>
57                                 <surname/>
58                             </name>
59                         <resp/>
```

```

60          </respStmt>
61          <imprint>
62              <biblScope unit="volume"/>
63              <biblScope unit="page"/>
64              <date/>
65          </imprint>
66      </monogr>
67  </biblStruct>
68  <biblStruct type="bookSection"  xml:id="YYY">
69      <analytic>
70          <title level="a"/>
71          <respStmt>
72              <resp/>
73              <name>
74                  <forename/>
75                  <surname/>
76              </name>
77          </respStmt>
78      </analytic>
79      <monogr>
80          <title level="m"/>
81          <respStmt>
82              <resp/>
83              <name>
84                  <forename/>
85                  <surname/>
86              </name>
87          </respStmt>
88          <imprint>
89              <pubPlace/>
90              <biblScope unit="page"/>
91              <publisher/>
92              <date/>
93          </imprint>
94      </monogr>
95  </biblStruct>
96  <biblStruct type="webpage"  xml:id="ZZZ">
97      <monogr>
98          <title level="m"/>
99          <respStmt>
100             <resp/>
101             <name>
102                 <forename/>
103                 <surname/>
104             </name>
105         </respStmt>
106         <imprint>
107             <date/>
108             <note type="accessed"/>
109             <note type="url"/>
110         </imprint>
111     </monogr>
112  </biblStruct>
113  </listBibl>
114 </sourceDesc>
115 </fileDesc>
116 <encodingDesc>
117     <projectDesc>
118         <p/>
119     </projectDesc>
120     <samplingDecl>
121         <p/>
122     </samplingDecl>

```

```

123      <editorialDecl>
124          <correction>
125              <p/>
126          </correction>
127          <normalization>
128              <p/>
129          </normalization>
130          <punctuation>
131              <p/>
132          </punctuation>
133          <segmentation>
134              <p/>
135          </segmentation>
136      </editorialDecl>
137      <classDecl>
138          <taxonomy xml:id="POS">
139              <gloss xml:id="aaa"/>
140          </taxonomy>
141          <taxonomy xml:id="probs">
142              <gloss xml:id="bbb"/>
143          </taxonomy>
144          <taxonomy xml:id="lang">
145              <gloss xml:id="ccc"/>
146          </taxonomy>
147      </classDecl>
148      <charDecl>
149          <glyph>
150              <glyphName xml:id="gn" corresp="www.vanhamel.nl/codecs/XXX"/>
151              <note/>
152          </glyph>
153      </charDecl>
154  </encodingDesc>
155  <profileDesc>
156      <handNotes>
157          <handNote xml:id="DDD">
158              <forename/>
159              <surname/>
160              <date from="yyyy" to="yyyy"/>
161              <region cert="x"/>
162              <note>
163                  <p/>
164              </note>
165          </handNote>
166      </handNotes>
167  </profileDesc>
168  <revisionDesc>
169      <change when="yyyy-mm-dd" who="XXX"/>
170  </revisionDesc>
171 </teiHeader>
172 <xi:include href="example.xml" parse="xml"/>
173 <xi:include href="hwData.xml" parse="xml"/>
174 </teiCorpus>

```

Appendix 6b: Annotated Corpus File Map<?xml version="1.0" encoding="UTF-8"?>

```

175 <!-- Required for all TEI documents. -->
176 <?xml-model href="file:/C:/Files/faclair-manuscripts/fnag_mss2.rng"
177 type="application/xml"
178 schematypens="http://relaxng.org/ns/structure/1.0"?>
179 <!-- Relative filepath to the schema ("fnag_mss2.rng"); it is
180 required. Update if another schema is used or if this file is moved. --
181 -->
182 <teiCorpus xmlns="http://www.tei-c.org/ns/1.0"
183 xmlns:xi="http://www.w3.org/2001/XInclude">
184 <!-- Root element for the corpus file: @xmlns required in all cases. --
185 -->
186 <teiHeader>
187 <!-- Contains the header for the entire corpus; i.e., data that could
188 potentially be of relevance to any of the transcription files. Points
189 in the corpus file where transcribers might need to add data are
190 indicated with the words "USER INPUT". -->
191 <!-- Any data already in the corpus file quite possibly relates to
192 multiple texts or transcription files in the corpus. It should not be
193 changed or removed without very thorough consultation and checking. --
194 -->
195 <fileDesc>
196 <!-- Contains data relating to the corpus; in all honesty, in the
197 context of this project, the names of these high-level corpus header
198 elements are not of great significance. TEI conventions demand that
199 certain data appears under certain headers but it is not always
200 obvious which ones. -->
201 <titleStmt>
202 <!-- Contains information about the creation of the corpus. -->
203 <title>Faclair na Gaidhlig: Medieval and Early Modern Manuscript
204 Corpus</title>
205 <!-- The corpus title. -->
206 <respStmt xml:id="UUU">
207 <!-- Describes the contribution of an individual to the corpus;
208 @xml:id contains their unique three-letter id. This is pointed to by
209 @resp and @corresp attributes elsewhere. -->
210 <name/>
211 <!-- The contributor's name -->
212 <resp/>
213 <!-- Their contribution -->
214 </respStmt>
215 <!-- Further contributors can be acknowledged by adding further
216 <respStmt> elements. -->
217 </titleStmt>
218 <publicationStmt>
219 <!-- Contains information on the publication and availability of the
220 corpus -->
221 <authority/>
222 <!-- FnaG project details -->
223 <availability>
224 <!-- Information on the availability of transcription files and data --

```

```

225  ->
226  <p/>
227  <!-- The default statement on the availability of texts; it restricts
228  them to internal FnaG/DASG use. -->
229  </availability>
230  </publicationStmt>
231  <sourceDesc>
232  <!-- The corpus bibliography. -->
233  <listBibl type="mss">
234  <!-- A list of manuscripts relevant to the corpus; this should include
235  both manuscripts from which transcriptions have been made
236  and manuscripts that have merely been cited. -->
237  <msDesc xml:id="VVV">
238  <!-- Contains information on a manuscript; if the manuscript
239  corresponds to a transcription file, then @xml:id = "MS" + the
240  transcription file's number. It should be the same as the value of
241  <msIdentifier> @sameAs in the corresponding transcription file. If the
242  manuscript does not correspond to a transcription file, then it should
243  be given a unique (within the context of the project) id formed of an
244  abbreviated reference to its repository plus its standard shelfmark,
245  with further codes added until the id is likely to be unique (e.g.
246  Dublin, Royal Irish Academy MS 23 P 12 (536) could safely take the id
247  "RIA23P12"). -->
248  <msIdentifier>
249  <!-- Identifying information on the manuscript. -->
250  <settlement/>
251  <!-- The town/city/region where the manuscript is located, according
252  to its standard designation (e.g. "Edinburgh" in Edinburgh, NLS Adv.
253  72.1.1). A following comma will be inserted automatically during
254  processing. -->
255  <repository/>
256  <!-- The library where this manuscript is located (e.g. "NLS Adv." in
257  Edinburgh, NLS Adv., 72.1.1). A following comma will be inserted
258  automatically during processing. -->
259  <idno type="shelfmark"/>
260  <!-- The shelfmark within that library. -->
261  </msIdentifier>
262  </msDesc>
263  <!-- USER INPUT: To add a new manuscript that has been, add another
264  <msDesc> element + children here and fill in the details. -->
265  </listBibl>
266  <listBibl type="lit">
267  <!-- A list of primary and secondary literature relevant to the
268  corpus. -->
269  <biblStruct type="book" xml:id="WWW">
270  <!-- A monograph in the bibliography; @xml:id should contain a unique
271  id for this bibliography item. It can be based on [first] author's
272  surname + publication date, plus an alphabetic serial code if the
273  initially resulting code is not unique -->
274  <monogr>
275  <title/>
276  <!-- The title of the monograph -->

```

```

277 <respStmt>
278 <!-- An individual responsible for the monograph. -->
279 <name>
280 <!-- The individual's name -->
281 <forename/>
282 <!-- The individual's forename -->
283 <surname/>
284 <!-- The individual's surname -->
285 </name>
286 <resp/>
287 <!-- The nature of their contribution: "author", "editor",
288 "translator", or "contributor". -->
289 </respStmt>
290 <!-- If there are multiple authors, or a translator or suchlike, add
291 further<respStmt> elements + children here -->
292 <imprint>
293 <!-- Publication details for the monograph -->
294 <pubPlace/>
295 <!-- Settlement where the monograph was published -->
296 <publisher/>
297 <!-- Publisher of the monograph -->
298 <date/>
299 <!-- Date of the publication of the monograph -->
300 </imprint>
301 </monogr>
302 </biblStruct>
303 <biblStruct type="article" xml:id="XXX">
304 <!-- A journal article in the bibliography; @xml:id should contain a
305 unique id for this bibliography item. It can be based on [first]
306 author's surname + publication date, plus an alphabetic serial code if
307 the initially resulting code is not unique -->
308 <analytic>
309 <!-- Details of the journal article -->
310 <title level="a"/>
311 <!-- Title of the journal article -->
312 <respStmt>
313 <!-- An individual responsible for the journal article. -->
314 <name>
315 <forename/>
316 <!-- The individual's forename -->
317 <surname/>
318 <!-- The individual's surname -->
319 </name>
320 <resp/>
321 <!-- The nature of their contribution: "author", "editor",
322 "translator", or "contributor". -->
323 </respStmt>
324 <!-- If there are multiple contributors, add further<respStmt>
325 elements + children here -->
326 </analytic>
327 <monogr>
328 <!-- Details of the journals -->

```

```

329 <title level="j"/>
330 <!-- Title of the journal -->
331 <imprint>
332 <!-- Publication details of the journal article -->
333 <biblScope unit="volume"/>
334 <!-- The journal volume number -->
335 <biblScope unit="page"/>
336 <!-- Page span of the journal article -->
337 <date/>
338 <!-- publication date of the journal -->
339 </imprint>
340 </monogr>
341 </biblStruct>
342 <biblStruct type="bookSection" xml:id="YYY">
343 <!-- A section from a multi-authored book in the bibliography; @xml:id
344 should contain a unique id for this bibliography item. It can be based
345 on[first] author's surname + publication date, plus an alphabetic
346 serial code if the initially resulting code is not unique -->
347 <analytic>
348 <!-- Details of the book section -->
349 <title level="a"/>
350 <!-- The title of the book section -->
351 <respStmt>
352 <!-- An individual responsible for the book section. -->
353 <name>
354 <forename/>
355 <!-- The individual's forename -->
356 <surname/>
357 <!-- The individual's surname -->
358 </name>
359 <resp/>
360 <!-- The nature of their contribution: "author", "editor",
361 "translator", or "contributor". -->
362 </respStmt>
363 <!-- If there are multiple contributors, add further<respStmt>
364 elements + children here -->
365 </analytic>
366 <monogr>
367 <title level="m"/>
368 <!-- The title of multi-authored book -->
369 <respStmt>
370 <!-- An individual responsible for the overall book (e.g. editor) -->
371 <name>
372 <forename/>
373 <!-- The individual's forename -->
374 <surname/>
375 <!-- The individual's surname -->
376 </name>
377 <resp/>
378 <!-- The nature of their contribution: "author", "editor",
379 "translator", or "contributor". -->
380 </respStmt>

```

```

381 <!-- If there are multiple contributors, add further<respStmt>
382 elements + children here -->
383 <imprint>
384 <!-- Publication details for the book section -->
385 <pubPlace/>
386 <!-- Settlement where the multi-authored book was published -->
387 <biblScope unit="page"/>
388 <!-- Page span of the book section -->
389 <publisher/>
390 <!-- Publisher of the multi-authored book -->
391 <date/>
392 <!-- Date of the publication of the multi-authored book -->
393 </imprint>
394 </monogr>
395 </biblStruct>
396 <biblStruct type="webpage" xml:id="ZZZ">
397 <!-- A webpage in the bibliography; @xml:id should contain a unique id
398 for this bibliography item. It can be based on [first] author's
399 surname + publication date, plus an alphabetic serial code if the
400 initially resulting code is not unique -->
401 <monogr>
402 <title level="m"/>
403 <!-- Title of the webpage -->
404 <respStmt>
405 <!-- An individual responsible for the website (e.g. editor) -->
406 <resp/>
407 <name>
408 <forename/>
409 <!-- The individual's forename -->
410 <surname/>
411 <!-- The individual's surname -->
412 </name>
413 </respStmt>
414 <!-- If there are multiple contributors, add further<respStmt>
415 elements + children here -->
416 <imprint>
417 <!-- Publication details for the website -->
418 <date/>
419 <!-- Date of the website's creation or latest relevant major update --
420 >
421 <note type="accessed"/>
422 <!-- Date on which the website was last accessed -->
423 <note type="url"/>
424 <!-- URL for the website -->
425 </imprint>
426 </monogr>
427 </biblStruct>
428 <!-- USER INPUT: To add another bibliography item, either (a) add the
429 relevant <biblStruct> element + children and fill in the details or
430 (b) export, copy, and paste one or more <biblStruct> elements from a
431 Zotero bibliography. -->
```

```

432 <!-- The order in which items are added to the bibliography within
433 corpus.xml is not important. The stylesheet will generate and sort
434 bibliographies automatically. Thus, new entries can simply be added at
435 the end. -->
436 </listBibl>
437 </sourceDesc>
438 </fileDesc>
439 <encodingDesc>
440 <!-- Information on the conventions adopted when compiling the corpus
441 and making the transcriptions -->
442 <projectDesc>
443 <!-- An overview of the FnaG mss project -->
444 <p/>
445 </projectDesc>
446 <samplingDecl>
447 <!-- The principles according to which material is selected for
448 transcription -->
449 <p/>
450 </samplingDecl>
451 <editorialDecl>
452 <!-- Contains elements describing the editorial principles applied in
453 the course of transcription on the project -->
454 <correction>
455 <!-- Description of the circumstances under which the manuscript text
456 might be amended in the transcription and how this is shown. -->
457 <p/>
458 </correction>
459 <normalization>
460 <!-- Description of the circumstances under which spelling in the
461 manuscript text might be normalised in the transcription and how this
462 is shown. -->
463 <p/>
464 </normalization>
465 <punctuation>
466 <!-- Description of any circumstances under which punctuation is added
467 to the transcription -->
468 <p/>
469 </punctuation>
470 <segmentation>
471 <!-- Policies regarding page, column, line, and word division -->
472 <p/>
473 </segmentation>
474 </editorialDecl>
475 <classDecl>
476 <!-- Keys to codes used in the mark up of transcriptions -->
477 <taxonomy xml:id="POS">
478 <!-- Codes used to mark up part of speech -->
479 <gloss xml:id="aaa"/>
480 <!-- A part of speech (e.g. noun); the code is contained in @xml:id --
481 >
482 <!-- USER INPUT: If another part of speech needs to be added, add
another <gloss> element here. -->

```

```

484 </taxonomy>
485 <taxonomy xml:id="probs">
486 <!-- Codes used to mark up problems with transcribing text -->
487 <gloss xml:id="bbb"/>
488 <!-- A problem (e.g. text_obscur); the code is contained in @xml:id --
489 -->
490 <!-- USER INPUT: If another problem record needs to be added, add
491 another <gloss> element here. -->
492 </taxonomy>
493 <taxonomy xml:id="lang">
494 <!-- Language codes (see https://www.iana.org/assignments/language-
495 subtag-registry/language-subtag-registry) -->
496 <gloss xml:id="ccc"/>
497 <!-- A problem (e.g. text_obscur); the code is contained in @xml:id --
498 -->
499 <!-- USER INPUT: If another language code needs to be added, add
500 another <gloss> element here. -->
501 </taxonomy>
502 </classDecl>
503 <charDecl>
504 <!-- A key to codes used to represent special characters in the
505 manuscript (e.g. abbreviations, ligatures) -->
506 <glyph xml:id="gn" corresp="www.vanhamel.nl/codecs/XXX">
507 <!-- A special character; @xml:id contains the code used to indicate
508 this character in <g> elements in the transcriptions; @corresp
509 contains the URL for an entry (if present) on this special character
510 in the Tiodscadal na Nod database.-->
511 <glyphName/>
512 <!-- A name for the special character; use the name from Tiodscadal na
513 Nod if possible -->
514 <note/>
515 <!-- A description of the special character and what it can stand for
516 -->
517 </glyph>
518 <!-- USER INPUT: If another special character needs to be added, add
519 another <glyph> + children here. -->
520 </charDecl>
521 </encodingDesc>
522 <profileDesc>
523 <!-- Contains information on the hands/scribes involved in the corpus'
524 manuscripts -->
525 <handNotes>
526 <!-- Hands appearing in the corpus -->
527 <handNote xml:id="DDD">
528 <!-- A hand appearing in one or more manuscripts in the corpus;
529 @xml:id is a unique id for each hand, = "Hand" + a serial number.
530 There is also a special id, Hand999, for hands that cannot be
531 identified (NB: this does not simply mean that the scribe is
532 anonymous; it means the sample is too brief or illegible for it to be
533 grouped with any other hand; anonymous hands should have their own
534 record/id) -->
535 <forename/>
```

```

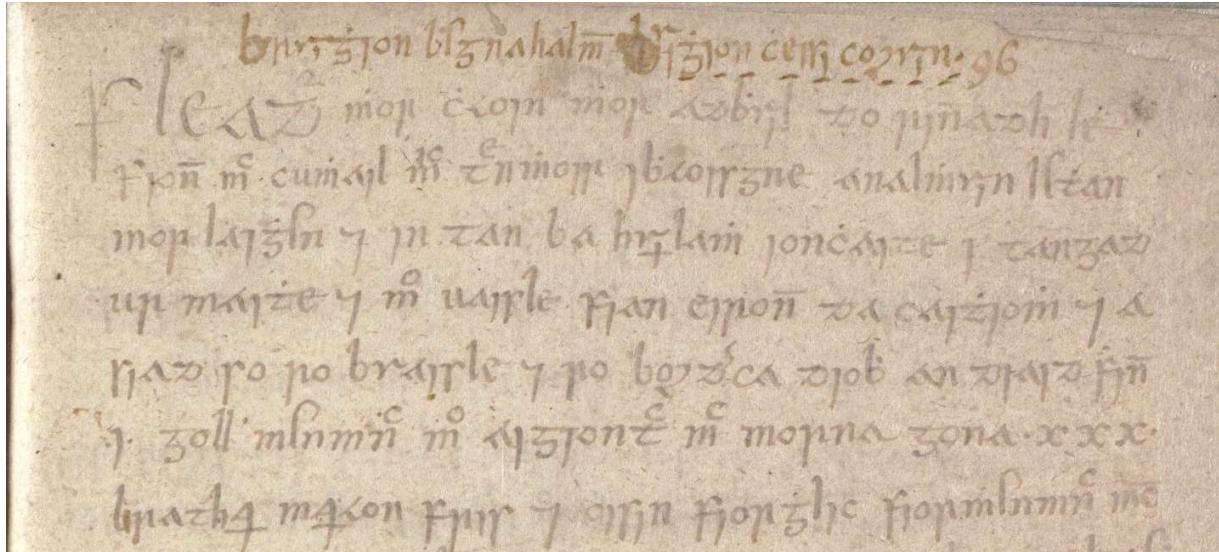
536 <!-- The scribe's first name [and sobriquet] (if known; if not known,
537 "unk.") -->
538 <surname/>
539 <!-- The scribe's surname/patronymic (if known; if not known, "unk.")
540 -->
541 <date from="yyyy" to="yyyy" min="yyyy" type="zzz" cert="xxx"/> <!--
542 The scribe's dating; the <date> element should contain a single number
543 indicating the century in which the scribe was active (e.g. "15"; NB:
544 his whole career, not just when the material being transcribed was
545 produced). If his period of professional activity spanned multiple
546 centuries, each should be given separated by a slash (e.g. "15/16"). --
547 -->
548 <!-- The @from and @to attributes contain specific dates: ideally the
549 scribe's birth-date and death-date but possibly a terminus ad quem and
550 a terminus post quem. For example, for Alasdair mac Mhaighstir
551 Alasdair: @from = 1698 and @to = 1770. @min and @max can be used for
552 to indicate the outer limits of a span in which the scribe's dates,
553 which may not be known, must definitely fall. For example, for a 15th-
554 century scribe: @min = 1400 and @max = 1500. In cases where the
555 potential span consists of two centuries, a special rule must be
556 adopted. For a 15th/16th-century scribe: @min = 1401, @max = 1599. This
557 so that, when the hands are sorted into chronological order, a hand
558 potentially in the previous century is placed behind a hand definitely
559 in the next century. -->
560 <!-- @type indicates whether this date is a "floruit" or a "lifespan".
561 @cert indicates the level of certainty with which these dates have
562 been given ("high", "medium", "low", "unknown"). -->
563 <region cert="xxx"/>
564 <!-- The geographical region(s) in which their scribal activity seems
565 to have taken place; no system is currently in place for describing
566 regions. This will be reviewed once more hands are in the corpus. -->
567 <note>
568 <p/>
569 <!-- The scribe's biography, interests, and palaeographic/orthographic
570 traits are summarised here -->
571 <!-- Further paragraphs can be added by adding further <p> elements --
572 -->
573 </note>
574 </handNote>
575 </handNotes>
576 </profileDesc>
577 <revisionDesc>
578 <!-- Keeps a record of major changes made to the corpus as a whole or
579 to the corpus file. -->
580 <change when="yyyy-mm-dd" who="XXX"/>
581 <!-- Details of a major revision. @when is the date of the change;
582 @who is the three letter ID for the main person responsible. -->
583 </revisionDesc>
584 </teiHeader>
585 <xi:include href="transcriptionTemplate.xml" parse="xml"/>
586 <!-- <xi:include> points to a transcription file; a stylesheet applied
587 to the corpus file will also be applied to any files pointed to by an

```

```
588 <xi:include> element at this point. This allows data from the corpus  
589 file to be used when transforming the transcription files. @href  
590 contains a relative filepath to the transcription file (in this  
591 example, "transcriptionTemplate.xml" would have to be in the same  
592 folder as the corpus file). @parse indicates what sort of file is  
593 being included (currently, only XML (i.e. TEI) files are included). --  
594 >  
595 <xi:include href="hwData.xml" parse="xml"/>  
596 <!-- hwData.xml is included in all instances, as this provides  
597 additional dictionary references and morphological data. -->  
598 </teiCorpus>
```

Appendix 7: Prose Sample

Appendix 7a: NLS Adv. 72.1.36, fol.96r1–8 (beg. of *Bruighion Bheag na hAlmhain*)



Appendix 7b: TEI Markup of NLS Adv. 72.1.36, fol.96r1–8

```
1 <div n="2" type="prose" resp="Hand9" corresp="MS12.2">
2   <p>
3     <pb n="96r" xml:id="MS12.96r"/>
4     <lb n="1" xml:id="MS12.96r.1"/>
5     <space type="em"/>
6     <add place="inline" type="insertion" resp="Hand9"><seg type="title">
7       <w lemma="bruinden" lemmaRef="http://www.dil.ie/7094" ana="noun">brui<abbr
8         cert="high">g<g ref="g4">h</g></abbr>ion</w>
9       <space type="scribal"/>
10      <w lemma="bec" lemmaRef="http://www.dil.ie/5512" ana="adj">
11        <abbr cert="high">b<g ref="g4">h</g></abbr>eg</w>
12       <space type="scribal"/>
13       <w lemma="in" lemmaRef="http://www.dil.ie/28194" ana="art">na</w>
14       <space type="scribal"/>
15       <name type="place">
16         <w lemma="Almu" lemmaRef="http://www.dil.ie/3005" ana="noun">hal<abbr
17           cert="high">m<g ref="g3">ain</g></abbr></w>
18       </name>
19       <space type="scribal"/>
20     </seg></add>
21     <seg type="title">
22       <del resp="Hand9">
23         <w lemma="bruinden" lemmaRef="http://www.dil.ie/7094"
24           ana="noun"><abbr cert="high">b<g
25             ref="g16">ru</g></abbr>i<abbr cert="high">g<g
26               ref="g4">h</g></abbr>ion</w>
27         <space type="scribal"/>
28         <name type="place">
29           <w lemma="UNKNOWN"><abbr cert="high">c<g ref="g4"
30             >h</g></abbr>eisi</w>
31           <space type="scribal"/>
32           <w ana="noun">coruin</w>
33         </name>
34         <pc>.</pc>
35       </del>
```

```

36      </seg>
37      <lb n="2" xml:id="MS12.96r.2"/>
38      <w lemma="fled" lemmaRef="http://www.dil.ie/22295"
39      ana="noun">FLEA<abbr cert="high">D<g ref="g4">H</g></abbr></w>
40      <space type="scribal"/>
41      <w lemma="mó�" lemmaRef="http://www.dil.ie/32548" ana="adj"><abbr
42      cert="high">m<g ref="g4">h</g></abbr>or</w>
43      <space type="scribal"/>
44      <w lemma="cain" lemmaRef="http://www.dil.ie/7787" ana="adj"><abbr
45      cert="high">c<g ref="g4">h</g></abbr>aoin</w>
46      <space type="scribal"/>
47      <w ana="adj, adj">
48          <w lemma="mó�" lemmaRef="http://www.dil.ie/32548"
49          ana="adj"><abbr cert="high">m<g ref="g4">h</g></abbr>or</w>
50          <space type="scribal"/>
51          <w lemma="adbal" lemmaRef="http://www.dil.ie/336"
52          ana="adj">ad<abbr cert="high">b<g ref="g4">h</g></abbr>uil</w>
53      </w>
54      <space type="scribal"/>
55      <w lemma="do, do-gní" lemmaRef="http://www.dil.ie/17098,
56      http://www.dil.ie/17752" ana="part, verb">
57          <w lemma="do" lemmaRef="http://www.dil.ie/17098" ana="part">do</w>
58          <space type="scribal"/>
59          <w lemma="do-gní" lemmaRef="http://www.dil.ie/17752" ana="verb">ri<abbr
60          cert="high">n<g ref="g5">n</g></abbr>adh</w>
61      </w>
62      <space type="scribal"/>
63      <w lemma="la" lemmaRef="http://www.dil.ie/29233"
64      ana="prep">le</w>
65      <lb n="3" xml:id="MS12.96r.3"/>
66      <name type="personal">
67          <w lemma="finn" lemmaRef="http://www.dil.ie/22134"
68          ana="noun">fio<abbr cert="high">n<g ref="g5">n</g></abbr></w>
69      </name>
70      <space type="scribal"/>
71      <w lemma="mac" lemmaRef="http://www.dil.ie/31166" ana="noun"><abbr
72      cert="high">m<g ref="g12">ac</g></abbr></w>
73      <space type="scribal"/>
74      <name type="personal">
75          <w lemma="cumall" lemmaRef="http://www.dil.ie/13810"
76          ana="noun">cu<abbr cert="high">m<g ref="g4">h</g></abbr>ail</w>
77      </name>
78      <space type="scribal"/>
79      <w lemma="mac" lemmaRef="http://www.dil.ie/31166" ana="noun"><abbr
80      cert="high">m<g ref="g12">ic</g></abbr></w>
81      <space type="scribal"/>
82      <name type="personal">
83          <w ana="adj, adj">
84              <w lemma="trén" lemmaRef="http://www.dil.ie/41798"
85              ana="adj"><abbr cert="high">t<g ref="g20">re</g></abbr>n</w>
86              <w lemma="mó�" lemmaRef="http://www.dil.ie/32548" ana="adj"><abbr
87              cert="high">m<g ref="g4">h</g></abbr>oir</w>
88          </w>
89      </name>
90      <space type="scribal"/>
91      <w lemma="úa" lemmaRef="http://www.dil.ie/42573" ana="noun">i</w>
92      <name type="personal">
93          <w lemma="baiscne" lemmaRef="http://www.dil.ie/5250"
94          ana="noun"><abbr cert="high">b<g ref="g4">h</g></abbr>aoisgne</w>
95      </name>
96      <space type="scribal"/>
97      <w lemma="i" lemmaRef="http://www.dil.ie/26891" ana="prep">a</w>

```

```

98      <name type="place">
99          <w lemma="Almu" lemmaRef="http://www.dil.ie/3005"
100             ana="noun">nal<abbr cert="high">m<g ref="g4">h</g></abbr>uin</w>
101      </name>
102      <space type="scribal"/>
103      <w lemma="lethan" lemmaRef="http://www.dil.ie/30046" ana="adj">le<abbr
104          cert="high">t<g ref="g4">h</g></abbr>an</w>
105          <lb n="4" xml:id="MS12.96r.4"/>
106          <w lemma="mó�" lemmaRef="http://www.dil.ie/32548" ana="adj">mor</w>
107          <space type="scribal"/>
108          <name type="population">
109              <w lemma="láigen" lemmaRef="http://www.dil.ie/29385" ana="noun">laigen</w>
110          </name>
111          <space type="scribal"/>
112          <w lemma="ocus" lemmaRef="http://www.dil.ie/33484" ana="conj"><abbr
113              cert="high"><g ref="g2">agus</g></abbr></w>
114          <space type="scribal"/>
115          <w lemma="in" lemmaRef="http://www.dil.ie/28194" ana="art">in</w>
116          <space type="scribal"/>
117          <w lemma="tan" lemmaRef="http://www.dil.ie/40004" ana="noun">tan</w>
118          <space type="scribal"/>
119          <w lemma="attá" lemmaRef="http://www.dil.ie/4927" ana="verb">ba</w>
120          <space type="scribal"/>
121          <w lemma="airlam" lemmaRef="http://www.dil.ie/2167" ana="adj">hu<abbr
122              cert="high"><g ref="g46">ir</g></abbr>la<abbr
123              cert="high">m<g ref="g4">h</g></abbr></w>
124          <space type="scribal"/>
125          <w ana="pref, ptcp">
126              <w lemma="in-" lemmaRef="http://www.dil.ie/28196"
127                  ana="pref">ion</w>
128              <w lemma="*caithe" source="caithid"
129                  lemmaRef="http://www.dil.ie/7996" ana="ptcp"><abbr cert="high">c<g
130                  ref="g4">h</g></abbr>aite</w>
131          </w>
132          <space type="scribal"/>
133          <w lemma="í" lemmaRef="http://www.dil.ie/26895" ana="pron">i</w>
134          <space type="scribal"/>
135          <w lemma="do-icc" lemmaRef="http://www.dil.ie/17788"
136              ana="verb">tangad<lb n="5" xml:id="MS12.96r.6"/>ur</w>
137          <space type="scribal"/>
138          <w lemma="maith" lemmaRef="http://www.dil.ie/31438"
139              ana="noun">mai<abbr cert="high">t<g ref="g4">h</g></abbr>e</w>
140          <space type="scribal"/>
141          <w lemma="ocus" lemmaRef="http://www.dil.ie/33484" ana="conj"><abbr
142              cert="high"><g ref="g2">agus</g></abbr></w>
143          <space type="scribal"/>
144          <w ana="adj, adj">
145              <w lemma="mó�" lemmaRef="http://www.dil.ie/32548"
146                  ana="adj"><abbr cert="high">m<g ref="g15">or</g></abbr></w>
147                  <space type="scribal"/>
148                  <w lemma="úasal" lemmaRef="http://www.dil.ie/42796" ana="noun">uaisle</w>
149          </w>
150          <space type="scribal"/>
151          <w lemma="fian" lemmaRef="http://www.dil.ie/21945" ana="noun">fian</w>
152          <space type="scribal"/>
153          <name type="place">
154              <w lemma="Ériu" lemmaRef="http://www.dil.ie/20288"
155                  ana="noun">eirio<abbr cert="high">n<g
156                  ref="g5">n</g></abbr></w>
157          </name>
158          <space type="scribal"/>
159          <w lemma="do" lemmaRef="http://www.dil.ie/17096" ana="prep, poss">da</w>
160          <space type="scribal"/>

```

```

161 <w lemma="caithem" lemmaRef="http://www.dil.ie/7993" ana="vnoun">cai<abbr
162 cert="high">t<g ref="g4">h</g></abbr>io<abbr cert="high">m<g
163 ref="g4">h</g></abbr></w>
164 <space type="scribal"/>
165 <w lemma="ocus" lemmaRef="http://www.dil.ie/33484" ana="conj"><abbr
166 cert="high"><g ref="g2">agus</g></abbr></w>
167 <space type="scribal"/>
168 <w lemma="is" lemmaRef="http://www.dil.ie/29104" ana="verb">a<lb n="6"
169 xml:id="MS12.96r.7"/>s</w>
170 <w lemma="iat" lemmaRef="http://www.dil.ie/27113" ana="pron">iad</w>
171 <space type="scribal"/>
172 <w lemma="so" lemmaRef="http://www.dil.ie/38191" ana="dpron">so</w>
173 <space type="scribal"/>
174 <w lemma="ro, is" lemmaRef="http://www.dil.ie/35364, http://www.dil.ie/29104"
175 ana="part, verb">
176     <w lemma="ro" lemmaRef="http://www.dil.ie/35364"
177     ana="part">ro</w>
178     <space type="scribal"/>
179     <w lemma="is" lemmaRef="http://www.dil.ie/29104" ana="verb">b</w>
180 </w>
181 <w lemma="úasal" lemmaRef="http://www.dil.ie/42796" ana="noun">uaisle</w>
182 <space type="scribal"/>
183 <w lemma="ocus" lemmaRef="http://www.dil.ie/33484" ana="conj"><abbr
184 cert="high"><g ref="g2">agus</g></abbr></w>
185 <space type="scribal"/>
186 <w lemma="ro, is" lemmaRef="http://www.dil.ie/35364, http://www.dil.ie/29104"
187 ana="part, verb">
188     <w lemma="ro" lemmaRef="http://www.dil.ie/35364"
189     ana="part">ro</w>
190     <space type="scribal"/>
191     <w lemma="is" lemmaRef="http://www.dil.ie/29104"
192     ana="verb">b</w>
193 </w>
194 <w lemma="airdirc" lemmaRef="http://www.dil.ie/1842" ana="adj">or<abbr
195 cert="high">d<g ref="g21">er</g></abbr>ca</w>
196 <space type="scribal"/>
197 <w lemma="de" lemmaRef="http://www.dil.ie/14787" ana="prep, pron">dio<abbr
198 cert="high">b<g ref="g4">h</g></abbr></w>
199 <space type="scribal"/>
200 <w lemma="i" lemmaRef="http://www.dil.ie/26891" ana="prep">an</w>
201 <space type="scribal"/>
202 <w lemma="dead" lemmaRef="http://www.dil.ie/14812" ana="noun">diaid</w>
203 <space type="scribal"/>
204 <name type="personal">
205     <w lemma="finn" lemmaRef="http://www.dil.ie/22134"
206     ana="noun"><abbr cert="high"><unclear reason="text obscure"
207     cert="medium" resp="EPT">f</unclear><g
208     ref="g4">h</g></abbr>i<abbr cert="high">n<g
209     ref="g5">n</g></abbr></w>
210 </name>
211 <lb n="7" xml:id="MS12.96r.8"/>
212 <w lemma=".i." lemmaRef="http://www.dil.ie/26892" ana="abbr">.i.</w>
213 <space type="scribal"/>
214 <name type="personal">
215     <w lemma="goll" lemmaRef="http://www.dil.ie/26369"
216     ana="noun">goll</w>
217 </name>
218 <space type="scribal"/>
219 <w lemma="menmnach" lemmaRef="http://www.dil.ie/31956" ana="adj">menm<abbr
220 cert="high">n<g ref="g12">ach</g></abbr></w>
221 <space type="scribal"/>
222 <w lemma="móir" lemmaRef="http://www.dil.ie/32548" ana="adj"><abbr
223 cert="high">m<g ref="g15">or</g></abbr></w>

```

```

224      <space type="scribal"/>
225      <w lemma="aicentach" lemmaRef="http://www.dil.ie/782" ana="adj">aigion<abbr
226      cert="high">t<g ref="g12">ach</g></abbr></w>
227      <space type="scribal"/>
228      <w lemma="mac" lemmaRef="http://www.dil.ie/31166" ana="noun"><abbr
229      cert="high">m<g ref="g12">ac</g></abbr></w>
230      <space type="scribal"/>
231      <name type="personal">
232          <w ana="noun">morna</w>
233      </name>
234      <space type="scribal"/>
235      <w lemma="co" lemmaRef="http://www.dil.ie/9787" ana="prep,
236      poss">gona</w>
237      <space type="scribal"/>
238      <w lemma="trícha" lemmaRef="http://www.dil.ie/41917" ana="num">.xxx.</w>
239      <lb n="8" xml:id="MS12.96r.9"/>
240      <w lemma="bráthair" lemmaRef="http://www.dil.ie/6585" ana="noun">brath<abbr
241      cert="high"><g ref="g8">ar</g></abbr></w>
242      <space type="scribal"/>
243      <w lemma="immaróen" lemmaRef="http://www.dil.ie/27763" ana="adv">m<abbr
244      cert="high"><g ref="g8">ar</g></abbr>aon</w>
245      <space type="scribal"/>
246      <w lemma="fri" lemmaRef="http://www.dil.ie/24329" ana="prep, pron">fris</w>
247      </p>
248  </div>

```

Appendix 7c: Output from TEI Markup of NLS Adv. 72.1.36, fol.96^r1–8

Editorial Text

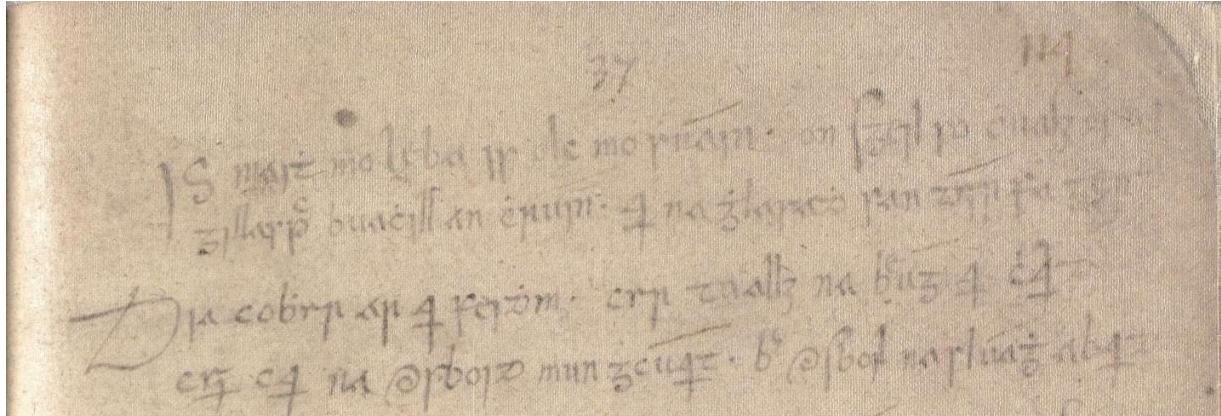
1. |bruighion bheg na halmain| bruighion ~~cheisi~~ coruin .
2. FLEADH mhór chaoin mhóradbhúil do rinnadh le
3. fionn mac cumhail mic trenmhoir i bhaoisgne a nalmhuin lethán
4. mor laigen agus in tan ba huirlamh ionchaite í tangad
5. ur maithe agus moruasle fian eirionn da caithiomh agus a
6. siad so rob uaisle agus rob orderca diobh an diaid {f}hinn
7. .i. goll menmnach mor aigontach mac morna gona .xxx.
8. brathar maraon fris

Diplomatic Text

1. |bruighion bheg na halmain| bruighion ~~cheisi~~ coruin.
2. FLEADH mhór chaoin mhór adbhuil do rinnadh le
3. fionn mac cumhail mic trenmhoir ibhaoisgne analmhuin lethán
4. mor laigen agus in tan ba huirlamh ionchaite í tangad
5. ur maithe agus mor uaisle fian eirionn da caithiomh agus a
6. siad so ro buaisle agus ro borderca diobh an diaid {f}hinn
7. .i. goll menmnach mor aigontach mac morna gona .xxx.
8. brathar maraon fris

Appendix 8: Verse Sample

Appendix 8a NLS Adv. 72.1.36, fol.114^r1–4
(beg. of 'Is maith mo leba, is olc mo shúain')



Appendix 8b: TEI Markup of NLS Adv. 72.1.36, fol.114^r1–4

```
1 <div n="4" type="verse" resp="Hand9" corresp="MS12.4">
2   <pb n="114r" xml:id="MS12.114r"/>
3   <lg n="1" type="stanza" xml:id="MS12.4.1">
4     <l n="a" xml:id="MS12.4.1a">
5       <lb n="1" xml:id="MS12.114r.1"/>
6       <w lemma="is" lemmaRef="http://www.dil.ie/29104"
7         ana="verb">Is</w>
8       <space type="scribal"/>
9       <w lemma="maith" lemmaRef="http://www.dil.ie/31438"
10      ana="adj">mai<abbr cert="high">t<g ref="g4">h</g></abbr></w>
11      <space type="scribal"/>
12      <w lemma="mo" lemmaRef="http://www.dil.ie/32400"
13        ana="poss">mo</w>
14      <space type="scribal"/>
15      <w lemma="lepaid" lemmaRef="http://www.dil.ie/29968"
16        ana="noun">leaba</w>
17      <space type="scribal"/>
18      <w lemma="is" lemmaRef="http://www.dil.ie/29104"
19        ana="verb">is</w>
20      <space type="scribal"/>
21      <w lemma="olc" lemmaRef="http://www.dil.ie/33783"
22        ana="adj">olc</w>
23      <space type="scribal"/>
24      <w lemma="mo" lemmaRef="http://www.dil.ie/32400"
25        ana="poss">mo</w>
26      <space type="scribal"/>
27      <w lemma="súan" lemmaRef="http://www.dil.ie/39098"
28        ana="noun"><abbr cert="high">s<g
29          ref="g4">h</g></abbr>úain</w>
30        <space type="scribal"/>
31        <pc>.</pc>
32        <space type="scribal"/>
```

```

33      </l>
34      <l n="b" xml:id="MS12.4.1b">
35          <w lemma="ó" lemmaRef="http://www.dil.ie/33364" ana="prep,
36              art">on</w>
37          <space type="scribal"/>
38          <w lemma="scél" lemmaRef="http://www.dil.ie/36385"
39              ana="noun">sgeil</w>
40          <space type="scribal"/>
41          <w lemma="so" lemmaRef="http://www.dil.ie/38191"
42              ana="dpron">so</w>
43          <space type="scribal"/>
44          <w lemma="ro-cluinethar" lemmaRef="http://www.dil.ie/35401"
45              ana="verb"><abbr cert="high">c<g
46                  ref="g4">h</g></abbr>ua<abbr cert="high">l<g
47                  ref="g7">us</g></abbr></w>
48          <space type="scribal"/>
49          <unclear reason="text_obscur" cert="medium"
50              resp="Campbell1885">
51              <w lemma="ós" lemmaRef="http://www.dil.ie/34047"
52                  ana="prep">os</w>
53              <w lemma="aird" lemmaRef="http://www.dil.ie/1792"
54                  ana="noun"><abbr cert="high"><g
55                      ref="g11">air</g></abbr>d</w>
56          </unclear>
57      </l>
58      <l n="c" xml:id="MS12.4.1c"><lb n="2" xml:id="MS12.114r.2"/>
59          <name type="personal">
60              <w lemma="gilla" lemmaRef="http://www.dil.ie/25843"
61                  ana="noun">gilla</w>
62              <w lemma="epscop" lemmaRef="http://www.dil.ie/20188"
63                  ana="noun"><supplied resp="EPT">e</supplied>spic</w>
64          </name>
65          <space type="scribal"/>
66          <w lemma="búachaill" lemmaRef="http://www.dil.ie/7194"
67              ana="noun">bua<abbr cert="high">c<g
68                  ref="g4">h</g></abbr>ill</w>
69          <space type="scribal"/>
70          <w lemma="in" lemmaRef="http://www.dil.ie/28194"
71              ana="art">an</w>
72          <space type="scribal"/>
73          <w lemma="cruinne" lemmaRef="http://www.dil.ie/13227"
74              ana="noun"><abbr cert="high">c<g
75                  ref="g4">h</g></abbr>rúi<abbr cert="high">n<g
76                  ref="g5">n</g></abbr></w>
77          <pc>. </pc>
78          <space type="scribal"/>
79      </l>
80      <l n="d" xml:id="MS12.4.1d">
81          <w lemma="ar" lemmaRef="http://www.dil.ie/3902" ana="prep,
82              poss"><abbr cert="high"><g ref="g8">ar</g></abbr><space

```

```

83      type="scribal"/>na</w>
84      <space type="scribal"/>
85      <w lemma="glasadh"
86          lemmaRef="http://www.faclair.com/ViewDictionaryEntry.aspx?ID
87          =2C31ED4BD82050707BD6D22FC40B0CF0" ana="vnoun"><abbr
88          cert="high">g<g
89          ref="g4">h</g></abbr>lasa<abbr cert="high">d<g
90          ref="g4">h</g></abbr></w>
91      <space type="scribal"/>
92      <w lemma="i" lemmaRef="http://www.dil.ie/26891" ana="prep,
93          art">san</w>
94      <space type="scribal"/>
95      <w lemma="túr" lemmaRef="http://www.dil.ie/42481"
96          ana="noun">túir</w>
97      <space type="scribal"/>
98      <w lemma="fo" lemmaRef="http://www.dil.ie/22333"
99          ana="prep">fa</w>
100     <space type="scribal"/>
101     <unclear reason="text_obscur" cert="medium"
102         resp="Campbell1885">
103         <w lemma="geàrd"
104             lemmaRef="http://www.faclair.com/ViewDictionaryEntry.aspx
105             ?ID=303D2FE708D1622B66B32BDB22BD55A" ana="noun"><abbr
106             cert="high">g<g ref="g4">h</g></abbr>eard</w>
107         </unclear>
108     </l>
109 </lg>
110 <lg n="2" type="stanza" xml:id="MS12.4.2">
111     <l n="a" xml:id="MS12.114r.2a"><lb n="3" xml:id="MS12.114r.3"/>
112         <unclear reason="interp_obscur" cert="medium" resp="EPT">
113             <w lemma="día n-" lemmaRef="http://www.dil.ie/15846"
114                 ana="conj">Dia</w>
115         </unclear>
116         <space type="scribal"/>
117         <w lemma="cubar" lemmaRef="http://www.dil.ie/13389"
118             ana="noun">co<abbr cert="high">b<g
119             ref="g4">h</g></abbr>ur</w>
120         <space type="scribal"/>
121         <w lemma="ar" lemmaRef="http://www.dil.ie/3902"
122             ana="prep">ar</w>
123         <space type="scribal"/>
124         <w lemma="ar" lemmaRef="http://www.dil.ie/3906"
125             ana="poss"><abbr cert="high"><g ref="g8">ar</g></abbr></w>
126         <space type="scribal"/>
127         <w lemma="feidm" lemmaRef="http://www.dil.ie/21472"
128             ana="noun">fei<abbr cert="high">d<g
129             ref="g4">h</g></abbr>m</w>
130             <pc>.</pc>
131             <space type="scribal"/>
132     </l>
133     <l n="b" xml:id="MS12.114r.2b">
134         <w lemma="cuirid" lemmaRef="http://www.dil.ie/13660"

```

```

135    ana="verb">cur</w>
136    <space type="scribal"/>
137    <w lemma="UNKNOWN">tual<abbr cert="high">l<g
138      ref="g7">us</g></abbr></w>
139    <space type="scribal"/>
140    <w lemma="in" lemmaRef="http://www.dil.ie/28194"
141      ana="art">na</w>
142    <space type="scribal"/>
143    <w lemma="bréc" lemmaRef="http://www.dil.ie/6611"
144      ana="noun"><abbr cert="high">b<g
145        ref="g20">ré</g></abbr>ug</w>
146    <space type="scribal"/>
147    <w lemma="ar" lemmaRef="http://www.dil.ie/3902"
148      ana="prep"><abbr cert="high"><g ref="g8">ar</g></abbr></w>
149    <space type="scribal"/>
150    <w lemma="càird"
151      lemmaRef="http://www.faclair.com/ViewDictionaryEntry.aspx?ID
152      =007AB404DEEA6B0878BD3BEECF89ABF2" ana="noun"><abbr
153      cert="high">c<g ref="g4">h</g></abbr><abbr cert="high"><g
154        ref="g11">air</g></abbr>d</w>
155  </1>
156  <1 n="c" xml:id="MS12.114.2c"><lb n="4" xml:id="MS12.114r.4"/>
157    <w lemma="cuirid" lemmaRef="http://www.dil.ie/13660"
158      ana="verb">cu<abbr cert="high"><g
159        ref="g46">ir</g></abbr></w>
160    <space type="scribal"/>
161    <w lemma="cor" lemmaRef="http://www.dil.ie/12406"
162      ana="vnoun">c<abbr cert="high"><g ref="g8">ar</g></abbr></w>
163    <space type="scribal"/>
164    <w lemma="in" lemmaRef="http://www.dil.ie/28194"
165      ana="art">na</w>
166    <space type="scribal"/>
167    <w lemma="conspóid" lemmaRef="http://www.dil.ie/12324"
168      ana="noun"><abbr cert="high"><g
169        ref="g28">con</g></abbr>sboid</w>
170    <space type="scribal"/>
171    <w lemma="imma" lemmaRef="http://www.dil.ie/27635"
172      ana="prep, poss">mun</w>
173    <space type="scribal"/>
174    <w lemma="cúairt" lemmaRef="http://www.dil.ie/13327"
175      ana="noun">gcú<abbr cert="high"><g
176        ref="g8">ar</g></abbr>t</w>
177    <pc>.</pc>
178    <space type="scribal"/>
179  </1>
180  <1 n="d" xml:id="MS12.114.2d">
181    <w lemma="beirid" lemmaRef="http://www.dil.ie/5583"
182      ana="verb"><abbr cert="high">b<g
183        ref="g21">eir</g></abbr></w>
184    <space type="scribal"/>
185    <unclear reason="interp obscure" cert="medium" resp="EPT">
186      <w lemma="conspóid" lemmaRef="http://www.dil.ie/12324"

```

```

187      ana="noun"><abbr cert="high"><g
188      ref="g28">con</g></abbr>sboil</w>
189    </unclear>
190    <space type="scribal"/>
191    <w lemma="in" lemmaRef="http://www.dil.ie/28194"
192      ana="art">na</w>
193    <space type="scribal"/>
194    <w lemma="slóg" lemmaRef="http://www.dil.ie/37981"
195      ana="noun">slúa<abbr cert="high">g<g
196      ref="g4">h</g></abbr></w>
197    <space type="scribal"/>
198    <unclear reason="interp_obscurer" cert="medium" resp="EPT">
199      <w lemma="a" lemmaRef="http://www.dil.ie/2"
200        ana="part">a</w>
201      <w lemma="bard" lemmaRef="http://www.dil.ie/5404"
202        ana="noun">b<abbr cert="medium">g
203        ref="g8">air</g></abbr>d</w>
204      </unclear>
205    </1>
206  </lg>
207 </div>

```

Appendix 8c: Output from TEI Markup of NLS Adv. 72.1.36, fol.114^r1–4

Editorial Text

1.

- 1. Is maith mo leaba is olc mo shúain .
on sgeil so chualus { os *aird* }
- 2. gilla [e]spic buachill an chrúinn .
arna ghlásadh san túir fa { *gheard* }

2.

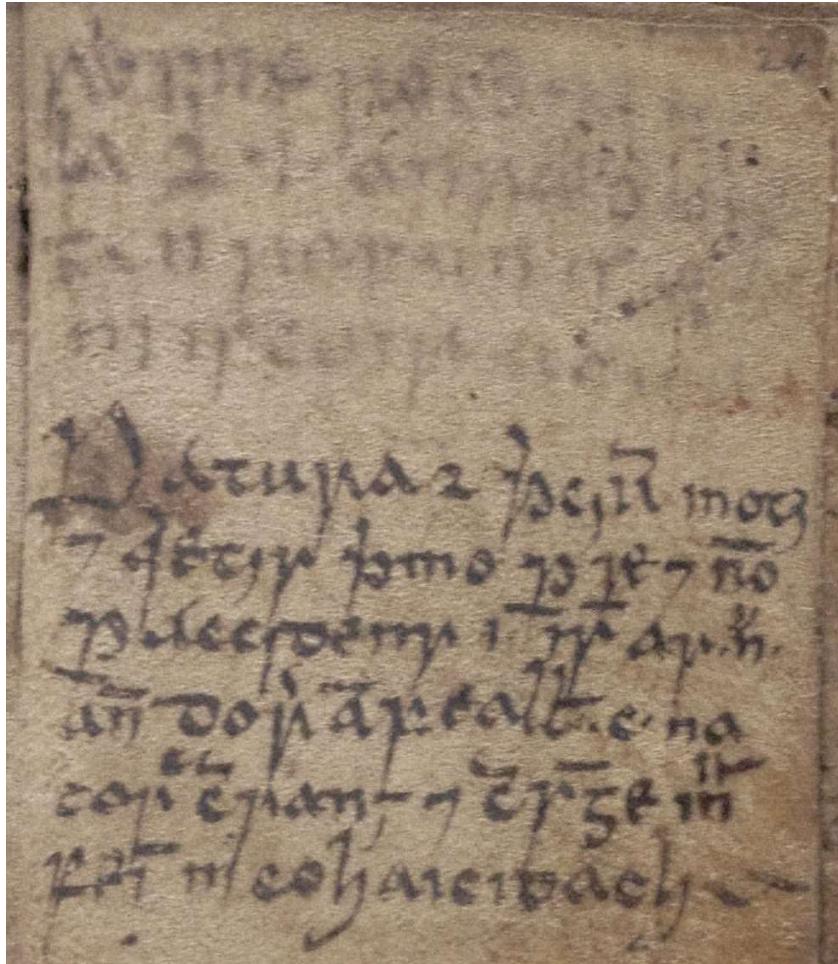
- 3. { *Dia* } cobhur ar *ar* feidhm .
cur *tuallus* na bréug *ar* *cháird*
- 4. cuir *car* na *consboid* mun gcúart .
beir { *consboil* } na slúagh { a *baird* }

Diplomatic Text

- 1. Is maith mo leaba is olc mo shúain . on sgeil so chualus {os*aird*}
- 2. gillaspic buachill an chrúinn. *ar* na ghlásadh san túir fa {*gheard*}
- 3. {*Dia*} cobhur ar *ar* feidhm. cur *tuallus* na bréug *ar* *cháird*
- 4. cuir *car* na *consboid* mun gcúart. beir {*consboil*} na slúagh {*abaird*}

Appendix 9: DivProse Sample

Appendix 9a NLS Adv. 72.1.4, fol.24^r1–10 (Definitions)



Appendix 9b: TEI Markup of NLS Adv. 72.1.4, fol.24^r1–10

```
1 <div n="1" type="divProse" resp="Hand24" corresp="MS17.1">
2   <pb n="24r" xml:id="MS17.24r"/>
3   <lg n="1" type="proseDiv" xml:id="MS17.1.1">
4     <l xml:id="MS17.1.1a">
5       <lb n="1" xml:id="MS17.24r.1"/>
6       <unclear reason="text_obscur" cert="high" resp="Black2011">
7         <w xml:lang="la">re</w>
8         <w xml:lang="la">sine</w>
9       </unclear>
10      <space type="scribal"/>
11      <unclear reason="text_obscur" cert="medium" resp="EPT">
12        <w xml:lang="la">racione</w>
13      </unclear>
14      <w xml:lang="la"><unclear reason="text_obscur" cert="low"
15        resp="EPT">nul</unclear>
16      <lb n="2" xml:id="MS17.24r.2"/>la</w>
17      <space type="scribal"/>
```

```

18 <w xml:lang="la"><abbr cert="high"><g ref="g18">est</g></abbr></w>
19 <space type="scribal"/>
20 </l>
21 <l xml:id="MS17.1.1b">
22   <w lemma=".i." lemmaRef="http://www.dil.ie/26892"
23     ana="abbr">.i.</w>
24   <space type="scribal"/>
25   <w lemma="in" lemmaRef="http://www.dil.ie/28194" ana="art">an</w>
26   <space type="scribal"/>
27   <unclear reason="text_obscur" cert="high" resp="Black2011">
28     <w lemma="rét" lemmaRef="http://www.dil.ie/35158"
29       ana="noun">raed</w>
30     <space type="scribal"/>
31     <w lemma="attá" lemmaRef="http://www.dil.ie/4927"
32       ana="verb">bis</w>
33   </unclear>
34   <lb n="3" xml:id="MS17.24r.3"/>
35     <w lemma="cen" lemmaRef="http://www.dil.ie/8581"
36       ana="prep">gan</w>
37     <space type="scribal"/>
38     <w lemma="résún" lemmaRef="http://www.dil.ie/35156"
39       ana="noun">resun</w>
40     <space type="scribal"/>
41     <unclear reason="interp_obscur" cert="medium" resp="EPT">
42       <w lemma="is" lemmaRef="http://www.dil.ie/29104"
43         ana="verb">is</w>
44   </unclear>
45   <gap reason="text_obscur" extent="3 chars" resp="EPT"/>
46   <lb n="4" xml:id="MS17.24r.4"/>
47   <unclear reason="interp_obscur" cert="medium" resp="EPT">
48     <w lemma="ní" lemmaRef="http://www.dil.ie/33146"
49       ana="noun">ni</w>
50     <space type="scribal"/>
51     <w lemma="is" lemmaRef="http://www.dil.ie/29104"
52       ana="verb">is</w>
53     <unclear reason="text_obscur" cert="medium" resp="Black2011">
54       <w lemma="cúis" lemmaRef="http://www.dil.ie/13695"
55         ana="noun">cuis</w>
56     </unclear>
57   </unclear>
58   <gap reason="text_obscur" extent="5 chars" resp="EPT"/>
59 </l>
60 </lg>
61 <lg n="2" type="proseDiv" xml:id="MS17.1.2">
62   <l xml:id="MS17.1.2a">
63     <lb n="5" xml:id="MS17.24r.5"/>
64     <w xml:lang="la">Natura</w>
65     <w xml:lang="la"><abbr cert="high"><g ref="g18">est</g></abbr></w>
66     <space type="scribal"/>
67     <w xml:lang="la"><abbr cert="high">p<g
68       ref="g25">ri</g></abbr><supplied resp="EPT">n</supplied>c<supplied
69       resp="EPT">ip</supplied>i<abbr cert="high">u<g
70       ref="g1">m</g></abbr></w>
71     <space type="scribal"/>
72     <w xml:lang="la">mo<abbr cert="high">t<g
73       ref="g7">us</g></abbr></w>
74   <lb n="6" xml:id="MS17.24r.6"/>

```

```

75 <w xml:lang="la"><abbr cert="high"><g ref="g2">et</g></abbr></w>
76 <space type="scribal"/>
77 <w xml:lang="la"><abbr cert="high">q<g
78 ref="g25">ui</g></abbr>etis</w>
79 <space type="scribal"/>
80 <w xml:lang="la"><abbr cert="high">p<g
81 ref="g25">ri</g></abbr>mo</w>
82 <space type="scribal"/>
83 <w xml:lang="la"><abbr cert="high">p<g ref="g36">er</g></abbr></w>
84 <space type="scribal"/>
85 <w xml:lang="la">se</w>
86 <w xml:lang="la"><abbr cert="high"><g ref="g2">et</g></abbr></w>
87 <w xml:lang="la"><abbr cert="high">no<g ref="g5">n</g></abbr></w>
88 <lb n="7" xml:id="MS17.24r.7"/>
89 <w xml:lang="la"><abbr cert="high">p<g ref="g36">er</g></abbr></w>
90 <space type="scribal"/>
91 <w xml:lang="la">accidens</w>
92 <space type="scribal"/>
93 </l>
94 <l xml:id="MS17.1.2b">
95   <w lemma=".i." lemmaRef="http://www.dil.ie/26892"
96     ana="abbr">.i.</w>
97   <space type="scribal"/>
98   <w lemma="is" lemmaRef="http://www.dil.ie/29104" ana="verb">is</w>
99   <w lemma="ed" lemmaRef="http://www.dil.ie/19657" ana="pron"><abbr
100    cert="high"><g ref="g3">ead</g></abbr></w>
101   <space type="scribal"/>
102   <w lemma="is" lemmaRef="http://www.dil.ie/29104" ana="verb">as</w>
103   <w lemma="é" lemmaRef="http://www.dil.ie/19325" ana="pron"><abbr
104    cert="high">h<g ref="g57">e</g></abbr></w>
105   <lb n="8" xml:id="MS17.24r.8"/>
106   <w lemma="i" lemmaRef="http://www.dil.ie/26891" ana="prep">a<abbr
107    cert="high">n<g ref="g5">n</g></abbr></w>
108   <space type="scribal"/>
109   <w lemma="do" lemmaRef="http://www.dil.ie/17096" ana="prep">do</w>
110   <w lemma="riar" lemmaRef="http://www.dil.ie/35223"
111     ana="noun"><abbr cert="medium">r<g ref="g21">eir</g></abbr></w>
112   <space type="scribal"/>
113   <w lemma="in" lemmaRef="http://www.dil.ie/28194" ana="art"><abbr
114    cert="high">a<g ref="g5">n</g></abbr></w>
115   <w lemma="fellsam" lemmaRef="http://www.dil.ie/21558"
116     ana="noun">feal<abbr cert="high">l<g ref="g3">sam</g></abbr></w>
117   <w lemma="cétnae" lemmaRef="http://www.dil.ie/8944"
118     ana="num"><abbr cert="medium">c<g ref="g9">ead</g></abbr><space
119      type="scribal"/>na</w>
120   <lb n="9" xml:id="MS17.24r.9"/>
121   <w lemma="toisech" lemmaRef="http://www.dil.ie/41302"
122     ana="noun">to<abbr cert="high">s<g ref="g12">ach</g></abbr></w>
123   <space type="scribal"/>
124   <w lemma="cumsanad" lemmaRef="http://www.dil.ie/13885"
125     ana="vnoun"><abbr cert="high">c<g ref="g1">um</g></abbr>sa<abbr
126      cert="medium">n<g ref="g3">aid</g></abbr></w>
127   <pc>, </pc>
128   <space type="scribal"/>
129   <w lemma="ocus" lemmaRef="http://www.dil.ie/33484"
130     ana="conj"><abbr cert="high"><g ref="g2">agus</g></abbr></w>
131   <space type="scribal"/>

```

```

132 <w lemma="cumscaigthe" lemmaRef="http://www.dil.ie/13891"
133 ana="noun"><abbr cert="high">c<g ref="g1">um</g></abbr>s<abbr
134 cert="medium">g<g ref="g3">aigth</g></abbr>e</w>
135 <space type="scribal"/>
136 <w lemma="immallé" lemmaRef="http://www.dil.ie/27701" ana="adv"
137 ><abbr cert="high">m<g ref="g58">aille</g></abbr></w>
138 <lb n="10" xml:id="MS17.24r.10"/>
139 <w lemma="fadéin" lemmaRef="http://www.dil.ie/20990"
140 ana="pron">fe<abbr cert="high">i<g ref="g5">n</g></abbr></w>
141 <space type="scribal"/>
142 <w lemma="is" lemmaRef="http://www.dil.ie/29104" ana="verb">ni</w>
143 <space type="scribal"/>
144 <w lemma="co" lemmaRef="http://www.dil.ie/9786" ana="prep">co</w>
145 <w lemma="aicídach" lemmaRef="http://www.dil.ie/796"
146 ana="adj">haicidach</w>
147 <pc>\</pc>
148 </1>
149 </lg>
150 </div>

```

Appendix 9c: Output from TEI Markup of NLS Adv. 72.1.4, fol.24^r1–10

Editorial Text

1.

1. { re sine } { racione } {nul} 2. la est
 .i. an { raed bis } 3. gan resun { is } -gap- 4. { ni is { cuis } } -gap-

2.

5. Natura est pri[n]c[ip]ium motus 6. et quietis primo per se et non 7. per accidens
 .i. is ead as he 8. ann do reir an feallsam ceadna 9. tosach cumsanaid , agus cumsgaigthe maille 10.
 fein ni co haicidach \\

Diplomatic Text

1. {resine} {racione}{nul}
 2. la est .i. an {raed bis}
 3. gan resun {is}-gap-
 4. {ni is{cuis}}-gap-
 5. Naturaest pricum motus
 6. et quietis primo per seetnon
 7. per accidens .i. iseald ashe
 8. ann do reir anfeallsamcead na
 9. tosach cumsanaid , agus cumsgaigthe maille
 10. fein ni cohaicidach\\