

# Early Irish Glossaries Project

## Transcription Notes

**Draft: 22 June 2009**

This document aims to describe in detail the conventions followed for manuscript transcription, including conventions for 1) transcribing text, and 2) encoding text as XML following the Text Encoding Initiative (TEI) standard (P5).

This is a draft document.

### 1. Transcription style

#### 1.1 Characters

Transcriptions include the following (Unicode) characters :

- Letters with puncta: *ś f m n*. (Note: *p c t b g d* are silently expanded, see below.)
- Ligatures: *æ œ*. Other ligatures not yet supported in Unicode are recorded as 'entities': *æ* as `&aolig`; and *á* as `&aoligacute`;<sup>1</sup>
- Accented characters: *á é í ó ú æ*.
- Brevigraphs: *7 (et)*,<sup>2</sup> *† (ue)*.
- Other: *ę (e caudata)*.<sup>3</sup>

#### 1.2 Capitalisation

Capitalisation is used in transcriptions where there are capitals in MS.<sup>4</sup> New sentences are not marked with an initial capital, unless where so in MS.

#### 1.3 Punctuation

In general, punctuations follows MS punctuation, but additional punctuation is minimally supplied for readability.

- Hyphens are added: 1) To separate initial mutations before vowels (e.g. *cech n-etag*). 2) To distinguish elements of nonce-compounds (e.g. *Oslucud .i. uas-lecad*).
- Full stops generally used only where in MS. Commas or semi-colons added elsewhere for readability.
- Apostrophe added to mark vowel elision (e.g. *m'forcend*).

Other punctuation is not used (e.g. colons, question/exclamation marks, macrons, middle dots).

---

<sup>1</sup> The *æo* ligature (though not the accented version) was introduced in Unicode 5.1 (April 2008), but is not yet well supported. Practice here follows the Medieval Unicode Font Initiative (MUFI) Character Recommendation version 3.0 (20 March 2009, available from <[www.mufi.info](http://www.mufi.info)>).

<sup>2</sup> The digit 7 is used in place of the Tironian 'et' symbol 7 (Unicode 204A), as the latter is rarely supported in Unicode fonts.

<sup>3</sup> Classified in Unicode as the Polish character 'Latin e with ogonek'.

<sup>4</sup> There is currently some inconsistency with regard to name of people and places, and titles of texts.

## 1.4 Abbreviations

Abbreviations *ʒ t .i. i.e. ʒrl* are retained. *i2* is expanded as *i.e.*

*ph ç t b ġ d* are expanded silently: *ph ch th bh gh dh*.

The following very common abbreviations are also expanded silently:

ṡ	(a)cht, sed	H	enim	—	(e)n
q̄	air	2	est	no,	nomen
ām	amail	f̄	for	nōm	
āp	apud	f̄	firi	nōa	nomina
q	ar, quia, quod	g <sup>c</sup>	gach	n̄	non
ɔ	con	.g. ḡ	grece (for Classical grace)	p̄	per
d̄c̄ns	dicens	ḡc	grece	q <sup>o</sup>	quo
d̄t	dicit	ḡc̄o	greco	qp	quod
d̄r	dicitur	ḡc̄o	greco	rqq	quoque
d̄nr	dicuntur	īm	immorro	ū	uer
dā	dano	īp	interpretatur	uñ	unde
dī	didiu	lā	latine	3	us, as
d̄x	dixit	~	(u)m	X <sup>A</sup>	XrA or XAr
eḃ	ebraice	m̄c	mac, maic, meic		

Other expansions are marked up as such (see below).

## 2. XML/TEI mark-up

### Text divisions

Container for glossary transcript	<code>&lt;div xml:id="[glossary code]"&gt;[text here]&lt;/div&gt;</code>
Glossary entry	<code>&lt;p xml:id=""&gt;Text of entry goes here.&lt;/p&gt;</code>
Headword (single word) <sup>5</sup>	<code>&lt;w type="hw" lemma="[DIL headword]"&gt;Babloir&lt;/w&gt;</code>
Headword (two or more words)	<code>&lt;phr type="hw" lemma="[DIL headword]"&gt;Orc treth&lt;/phr&gt;</code>

### Manuscript description

New column	<code>&lt;cb n="p. 270b"/&gt;</code>
Empty space in MS <sup>6</sup>	<code>&lt;space extent="2.5 lines"/&gt;</code>

### Readings

Text not transcribable	<code>&lt;gap extent="4 letters" reason="illegible"/&gt;</code> <code>&lt;gap extent="2 lines" reason="erasure"/&gt;</code> <code>&lt;gap extent="3 letters" reason="damage"/&gt;</code>
Text omitted (MSS H <sup>1a</sup> , H <sup>1b</sup> , K only)	<code>&lt;gap extent="rest of entry" reason="sampling"/&gt;</code>
Text unclear (but partly legible)	<code>&lt;unclear&gt;some barely readable text&lt;/unclear&gt;</code>
Text supplied by editor	<code>&lt;supplied&gt;A&lt;/supplied&gt;dam .i. homo</code>
Expanded abbreviation <sup>7</sup>	belach gab<ex>ruan</ex>
Text marked in MS to be deleted	<code>&lt;del&gt;text&lt;/del&gt;</code>
Text transposed (e.g. AB to BA) <sup>8</sup>	BA<note>Transposed from AB</note>
Text inserted later (inline)	Tigrad<add place="inline">us</add>

### Glosses

Supralinear	<code>&lt;add&gt;gloss&lt;/add&gt;</code>
Infralinear	<code>&lt;add place="infralinear"&gt;gloss&lt;/add&gt;</code>

<sup>5</sup> The headword may be analysed editorially as a single word even where two words are written in the manuscript (e.g. Sanas Cormaic M 'brac cille' refers to *DIL braccaille*).

<sup>6</sup> For spaces left by scribes (e.g. after letter block). Not for deleted, damaged or illegible text (see under Readings).

<sup>7</sup> Other than those silently expanded in the list above.

<sup>8</sup> In preference to the use of `<add>` and `<del>` as recommended in the TEI guidelines (11.3.6).

Left margin	<code>&lt;add place="margin-left"&gt;gloss&lt;/add&gt;</code>
Right margin	<code>&lt;add place="margin-right"&gt;gloss&lt;/add&gt;</code>
Upper margin	<code>&lt;add place="margin-top"&gt;gloss&lt;/add&gt;</code>
Lower margin	<code>&lt;add place="margin-bottom"&gt;gloss&lt;/add&gt;</code>
Corrections <sup>9</sup>	<code>ga&lt;add type="correction"&gt;o&lt;/add&gt;bur</code>
<b>Editorial notes<sup>10</sup></b>	
Editorial note	<code>&lt;note&gt;A general editorial note may be included here.&lt;/note&gt;</code>
Non-Irish word <sup>11</sup>	<code>&lt;w xml:lang="la"&gt;ab eo quod est&lt;/w&gt;</code>

Features not described:

- Large letter initials (e.g. at the start of letter blocks).
- Metrical divisions (lines or stanzas of verse).

Notes on language mark-up:

- At present we mark languages in MSS Y, OM<sup>1</sup>, D<sup>1</sup>, *Loman*, *Irsan* only.
- The extremely common and potentially ambiguous abbreviations *ſ. i. i. e.* are not marked as Latin.
- Words introduced with *quasi* are marked as Latin only if they appear to be actual Latin words (in some cases possibly scribally corrupted), rather than invented Latinate nonce-words.
- Words are marked as Greek, Hebrew, etc. if and only where the text labels them as such. (Editorial indentifications of Greek, etc. are discussed in the editions.)

### 3. Misc to do

Uncertain readings are currently provisionally marked {?}. These need to be checked against microfilms/manuscripts and either resolved or else marked with `<unclear>[uncertain reading here]</unclear>` or `<gap extent="..." reason="illegible"/>`.

<sup>9</sup> Corrections for TEI purposes are regarded as additions to the text, inserted by 'an author, scribe, annotator, or corrector'.

<sup>10</sup> Elements `<corr>`, `<sic>`, `<colophon>` not used.

<sup>11</sup> Codes for foreign languages: *lat* Latin, *grc* Greek, *heb* Hebrew, *non* Norse, *cym* Welsh (for Brittonic), *eng* English, *qpi* Pictish. (Follows the ISO 639-2 standard. *qpi* is an invented code within the section 'reserved for personal use'. See <http://www.loc.gov/standards/iso639-2/>.)