Alice Thornton's Books

Editorial Guidelines

Suzanne Trill   Sharon Howard

25 February 2025

Contents

About this document

This document outlines the editorial policies and practices used in transcribing and tagging Alice Thornton's Books (2025). It should be read in conjunction with the *TEI Element Set*.

The documentation and the TEI XML files to which it refers are stored in the Alice Thornton Digital Edition Github Repository.

```
alice-thornton
|--documentation
| |--atb_elements.pdf
| |--atb_guidelines.pdf
|--entities
  |--events.xml
  |--glossary.xml
  |--glossary_book_one.xml
  |--glossary_book_three.xml
  |--glossary_book_two.xml
  |--people.xml
  |--places.xml
|--texts
| |--00_book_of_remembrances
|    |--book_of_remembrances.xml
| |--01_book_one
|    |--book_one.xml
| |--02_book_two
|    |--book_two.xml
| |--03_book_three
|    |--book_three.xml
```

A note on terminology

- **TEI XML**: XML files marked up using the Text Encoding Initiative schema

- **DSE**: online Digital Scholarly Edition

- **SD**: semi-diplomatic version of the DSE

Transcription Process

Although all of *Book 2* had been transcribed from the original manuscript by Dr Lisa Liddy before the AHRC project began, as Covid-19 protocols were still in place in its early stages, the initial transcriptions of the other books were predominantly made from digital images. Patrick Comber gave permission for Cordelia Beattie to photograph the *Book of Remembrances*, while the British Library supplied images of *Book 1* and *Book 3*. Durham Cathedral Library produced images of *Book 2* which allowed us to cross check transcriptions remotely. Team members noted queries that required checking in the original manuscripts which were updated when it became possible to do so. In order to ensure these provided a firm foundation for future work, the first set of transcriptions were fully diplomatic in format and were saved as word documents for reference.

Workflow has ensured that several different project team members have been involved in the transcription process.

- first transcription: fully diplomatic

    - transcribed by two members of the team
    - cross-checked by transcribers
    - checked by another team member

- second transcription: semi-diplomatic

    - transcribed by two members of the team
    - cross-checked by transcribers
    - checked by another team member
    - shared with the Project Board and the contributors to the volume of essays commissioned to accompany the DSE.

Editorial Principles

Semi-Diplomatic

Our practice was informed by the suggested transcription conventions outlined here: English Handwriting 1500-1700: An Online Course. In our adaptation, the semi-diplomatic version of the text retains some original textual features, including original lineation, paragraphing, punctuation and capitalisation. It also includes evidence of authorial corrections, such as insertions and deletions, and records spaces left intext and blank pages. We also record Thornton's inconsistent use of contractions where these are authorially signalled intext by tildes and note significant forms of damage (such as the cut pages in the *Book of Remembrances*, 57-58).

However, we have silently modernised Thornton's use of u/v, i/j and long S, and expanded her contractions of common words such as 'the', 'which', and 'that' (from 'y^e', 'w^{ch}' and' y^t'), alongside her more unusual use of 'o.^{r}' for 'our'. We have not retained end of line fillers (e.g. ~) and variant forms of hyphenation have been standardised to '-'. Although there is some evidence of later readers annotating Thornton's Books, such intervention is minimal and has not been recorded.

Thornton's hand is predominantly italic and mostly highly legible; however, this is not always the case and, especially where her hand is less legible, Thornton makes frequent use of idiosyncratic abbreviations in which vowels are omitted. Such instances have not been marked up distinctively but are included among substitutions. Two features of Thornton's orthography that may be of interest to historical linguists are her use of 'i' for 'c' and her frequent use of 'hn' for 'him'. While the former is identified in our TEI XML (see '<choice>' below), the latter was silently standardised.

Modernisation

The modernised version of the text follows the conventions of UK English, supplemented by The Chicago Manual of Style (17, notes and bibliography). However, to retain a sense of Thornton's distinctive voice, we have retained some archaic forms for which we have provided glosses. Sometimes Thornton's syntax deviates from modern expectations and where further intervention is required to clarify sense supplied materials are identified by square brackets; to

retain a clean reading experience in the modernised edition of her Books, these are displayed in the SD version.

A major part of the modernisation process involved punctuation. In addition to extensive use of commas and colons, it was often difficult to distinguish Thornton's periods from rest marks. Erring on the side of caution on this point in the SD meant that changes in this area were significant: 17,680 <supplied reason="mod">; 8,930 <surplus reason="superfluous">; 5,800 <pc>@norm. (Supplied – materials added to make sense in modernisation; Surplus – materials deleted to make sense in modernisation; pc@norm – used to convert AT's original punctuation into an appropriate modernised form.)

As is common in Early Modern English (EME), Thornton's sentences frequently commence with a conjunction (most often, 'and', 'but' or 'which') and her paragraphing is inconsistent. Particularly in *Book 2* and *Book 3*, paragraphs can range from a full page or more to several very short paragraphs on one page. In this latter case, sometimes a single sentence can extend over two (or more) paragraphs. As we used parallel markup, this created a tension between the structural hierarchies of TEI XML and the need to make syntactical sense of Thornton's expression in modernised English. To address this issue, KDL implemented a means by which two or more paragraphs in the SD version would display continuously in the modernised DSE (see 'removal' below).

TEI Preparation and Process

The completed TEI XML contains a number of features, particularly those related to the original appearance of the texts on the page, that are not present in the DSE, or are represented in a simplified manner. The descriptive documentation here focuses primarily on markup that is actually used in the DSE; the Element Set is a more comprehensive record.

Workflows and collaboration

Markup of the transcribed texts involved a number of distinct components. However, after the initial steps, the process was increasingly complex and collaborative and team members could be working on different features at the same time. To facilitate this, the XML files were held in a project Github repository and the team received guidance in using Git and Github.

Initial steps:

- conversion of semi-diplomatic transcriptions in MS Word to TEI XML

- structural markup: divisions of the texts, section titles, headings, marginalia

- textual markup: authorial revisions and editorial insertions

Contextual enhancements:

- named entities (people and places)

- dates

- events

- biblical quotations

Editorial:

- modernisation

- annotation (notes and glosses)

In the collaborative stages, it was important to ensure that the team followed consistent practice as far as possible. Sharon Howard wrote a number of guidance notes on specific aspects of markup, in particular for entities, biblical and modernisation; the policies and practices outlined in those documents have been incorporated into this documentation. The use of Github was also helpful in this respect, as it enabled the sharing of queries and corrections much more readily than would be the case if working on files in isolation.

Files

The base files for XML encoding were the project's semi-diplomatic transcriptions created in Microsoft Word, one file per Book. The transcriptions were converted to TEI XML during the winter of 2021-22, using the TEI stylesheets integrated into the OxygenXML Editor. The schema used is standard TEI P5 with no customisations.

Seven additional entities files were added during the course of the project to store linked data about people, places, events and annotations. The standoff files have been generated from project spreadsheets and other documents, using XSLT, XQuery and R for processing.

The files are stored in the project Github repository:

- Texts

- Entities

DocxToTEI conversion

It's not unusual to begin a TEI project by transcribing in Microsoft Word (or similar software) and then converting the transcriptions into TEI XML. This 'DocxToTEI' approach has the benefit of enabling transcribers to work with familiar software so they can focus on their task without having to grapple with new technologies. However, from an encoding perspective, the method is a compromise. Formatting in word processing software is much more limited than the tag set available in TEI; on the other hand, the converted XML is full of verbose, irrelevant tagging that has to be removed before further work can begin. As a result, careful preparation before conversion begins and cleaning afterwards are both needed.

The conversion used a workflow similar to one described by Dot Porter, which combines Word's built-in formatting and styles with simple 'pseudocodes' to extend the software's capabilities to some extent.

1. The first step before proceeding with conversion was to model the TEI that would be required, with reference to the broad aims of the project and detailed transcription Conventions. Some

protocols were agreed with transcribers so that features like Thornton's textual changes and marginal annotations would be represented using consistent and unambiguous Word formatting.

2. It's equally important to understand MS Word's formatting and styling features, including the use of custom styles. A number of custom styles were added to particular features in the documents: for example, a custom style named *tei:fwPageNum* was applied to Thornton's page numbers which would become **<fw type="pageNum">** and *tei:l* for lines of verse (which would become **<l>**). Small samples were used for testing before applying the process to the full texts.

3. Following the initial conversion, using the TEI stylesheets integrated into the OxygenXML editor, a small "upconverter" XSLT script was used (based on one originally written by James Cummings). This would:

- indent and format the XML consistently

- remove unwanted Word styles markup

- convert the pseudocodes to TEI where possible (if not, it would be done at the next stage using Find and Replace)

4. The upconversion was followed by some manual checks and cleaning.

At this point the XML consisted of a very basic set of tags, mostly those directly converted from Word such as **<pb>** for page breaks, **<lb>** for line breaks and **<p>** for paragraphs, with a few more specific tags converted from the pseudocodes. Nearly all the rest were **<hi>** tags marking features such as superscript, deleted, italic or underlined text.

Data linkage

Data linkage was a crucial element of the project's work in order to enhance browsing and create search tools in the DSE. Databases were created for people, places and events, and every discrete entity or event (as well as annotations) was assigned a unique identifier to which instances of it in the TEI XML could be linked, using several TEI methods to link tagged elements of all kinds together.

1. the @ref attribute was used to link mentions of entities to database entries. The project identifier for Thornton's son Robert is **rt2**, for example, which would be used in the TEI XML files with a **ppl** prefix (to ensure no possible confusion with any other IDs):

- <persName ref="ppl:rt2">Robin Thornton</persName>

2. the @corresp attribute was used to link <anchor> tags for endnotes to the corresponding note texts:

- text: I fell into the smale-pox<anchor n="8" corresp="#p009n02" resp="ednote"/>

- note: <note n="8" xml:id="p009n02" type="note" place="end">
  <p> See Raymond A. Anselment, 'Smallpox in Seventeenth-Century English Literature'...</p>
  </note>

3. **<ptr>** with a @target attribute was used to store event mention IDs in the events metadata.

<linkGrp type="sgl">
<ptr target="bookrem:ev91" type="book" />
<ptr target="book1:ev209" type="book" />
</linkGrp>

The metadata held in the databases was extracted into separate TEI files, known as *standoff* files or, in the case of the endnotes, the notes were placed in a **<standOff>** container at the bottom of the same file. Standoff markup is a highly useful tool that forms a vital element of the edition as a whole. It helps to reduce verbosity within the TEI XML, since contextual information does not have to be repeated every time it is needed for reference in the text.

Consistency and caveats

We regard the encoding of the four Books as a deeply interpretative process, and their complexity and idiosyncracies often defy schemas and standardisation. Towards the end of the project, Howard carried out various consistency reviews and checks of markup. However, the scale and complexity of the final markup made it impossible to ensure complete consistency. Moreover, reviews focused on features that are represented in the DSE, and other markup was likely to receive less scrutiny.

We have endeavoured to create a set of encoded texts with reuse potential beyond the life of the project that created them, but it is important to note that not all markup was applied equally systematically; there are some very experimental and provisional usages that were abandoned due to time constraints, and some markup was intended primarily for internal processing. These may not be specifically documented, or noted with health warnings.

Structure, Text and Page

Book historians emphasise the lack of a strict distinction between manuscript and printed book in the early modern period, and this is well illustrated by the four manuscripts that make up Alice Thornton's Books. They exhibit many features of printed books of the time: title pages, front and back matter, chapters, titles, page numbers and page headings, indexes, decorations and marginal glosses. As a result, the TEI default text structure markup has been used in largely conventional ways, and the Women Writers' Project's guide on encoding early printed books was a key reference text.

However, Thornton's usage of these features was less consistent than would be expected in a printed book, even considering that early modern print was far less uniform than its modern counterpart. For example, she does not always give "chapters" a title; not all books have front and back matter; she does not always number pages (and she quite often makes errors in numbering). Moreover, the books also show many characteristics of scribal rather than printed material. While they are mostly neatly-written fair copies, they contain many minor authorial revisions as well as extensive abbreviations and contracted forms of words. Thornton also made some use of scribal conventions such as *signes-de-renvoi*.

There are also many decorative flourishes that range from small markers resembling slashes, hashes and printers' fleurons to elaborately patterned title decorations including ruled lines, flowers and hearts. Most of these are not represented in the DSE but are encoded (with possibly varying degrees of consistency) in the TEI XML.

Structural divisions

The four TEI XML texts share a standard basic TEI structure: **\<teiHeader\>** (metadata) and **\<text\>** (the transcription of the text), with an additional **\<standOff\>** container for editorial notes. Within **\<text\>**, all four of the texts contain a main **\<body\>** and some kind of **\<front\>** matter. Two of the books also contain **\<back\>** matter.

Subdivisions of the main text

The **<div>** tag is used to mark up narrative chapters or other significant, distinct sections of material; certain sections consisting entirely (or nearly so) of certain kinds of material have a @type attribute to indicate this, although only the types "poetry" and "index" are given any special treatment in the DSE.

- @type="poetry"

Owing to time constraints, poetry has been marked up in only a very basic way. (There is a handful of **<lg>** (line group) tags, but no particular significance should be read into them.)

Lines of verse are marked up with **<l n="x">**, where "x" is the line number within the poem. Lines may contain line break **<lb>** tags.

- @type="index"

Book 1 and the *Book of Remembrances* both contain a section that Thornton labelled as an "Index" in back matter (although it might be more accurate to call it a table of contents). These were marked up in some detail to reflect their complex layout (see the next section on lists).

- @type="prayer"

There are a number of substantial sections of prayer across the Books and these have been marked up as such. However, it is important to be aware that these headlined prayers do not represent all the prayers to be found in the Books; there are many shorter prayers interpersed with narratives.

- @type="supplementary"

In *Book 1* and the *Book of Remembrances*, Thornton inserted additional material at some point after writing the main text; in the latter this is in a distinct gathering *after* the Index.

- @type="correspondence", "documents" (etc)

Towards the end of *Book 3*, Thornton compiled a number of lists of letters, legal documents and other items.

Lists and indexes

There are various lists in the book, marked up with **<list>**, and Thornton's indexes have also been treated as lists for markup. Individual items in a list are marked up with **<item>** and numbered labels (and page references in the index) with **<label>**.

- @type="index"

Items in the indexes are frequently grouped together with a single page number. To capture this, we used a special attribute @rend="braced({})". A @corresp attribute links the entry to the corresponding <div> in the text, though this linkage was not used in the DSE.

Example:

<item>
<list rend="braced({})">
<item corresp="#div-024">Mr Edmund Norton died. 1648 </item>
<item corresp="#div-025">King Charles the Blessed. 1648 </item>
<item corresp="#div-028">Sr Edward Osborne died</item>
</list>
<label place="right mid"><ref type="page">23.</ref></label>
</item>

Note that the page numbers in <ref> tags refer to Thornton's written numbers, not the edition's pagination. Thornton sometimes made errors, and where that is the case her actual page number is noted in an @n attribute.

Titles

There are four forms of markup used to distinguish titles and headings from the main body of text:

- **<docTitle>** is used on title pages only

- **<head>** is used for chapter and section titles

- **<fw type="header">** is used for page headings, primarily in *Book of Remembrances*

- **<fw type="head">** is used to make some more significant page headings (where <head> is inappropriate or not allowed) stand out in the DSE

The standard formatting applied to <head> markup in the DSE simplifies the varied forms of decoration that Thornton applied to titles, which are encoded in more detail in @rend attributes.

Textual features

As already noted, the Books are largely fair copies of narrative prose, with occasional verse, and as such markup of the transcribed text didn't present many major challenges.

Changes by the author

Thornton's deletions were marked up with **\<del rend="(deletion)"\>**. In the DSE these are all indicated with a single strikethrough, but occasional more forceful deletions are noted in the @rend attribute.

Text inserted by Thornton at some point after initial writing was marked up with **\<add place="(location)"\>**. Most of these insertions were above the line and are indicated using @place=above; some were in the left margin. A few longer insertions were written inline.

Where an addition and deletion represent a single substitution, they are grouped together using the **\<subst\>** tag.

Contractions and abbreviations

Contractions in words that were signalled with a tilde were transcribed in curly braces and marked up with **\<ex\>** so that they could be handled distinctly in the DSE.

Other forms of abbreviation, written above the line in superscript but not marked with a tilde, were marked up with **\<am rend="superscript"\>** and were given additional handling during the modernisation process.

Perhaps one of the most notable features in the Books is Thornton's frequent use of a heart "emoji" in place of the word "heart". These are marked up with **\<g ref="#heart"\>**.

Materiality and the page

Rendition and layout

Markup across various elements (including \<fw\>, \<head\>, \<div\>, \<note\> and \<milestone\> markers) made extensive use of @rend (rendition) and @place attributes to record as much detail about the appearance of the text on the page as possible. We made use of the WWP's rendition keywords strategy to encode some complex appearance features, so for example a full-width ruled horizontal line before a heading would be **pre(rule)** and a similar ruled line decorated with hearts and flowers after the heading would be **post(rule_pattern)**. Ultimately much of this detail was not used or was simplified in the DSE, but it has been retained in the TEI XML.

Page images

Selected images for each book are displayed in the DSE; the markup for this is **\<figure\>\<graphic\>**, inserted at the beginning of the respective page.

Page numbering

There are two different sets of page numbers marked up in the texts:

The edition page numbering uses **<pb id="page-xxx" />**, where xxx is the page number.

Meanwhile, Thornton's original written numbers are marked up with **<fw type="pageNum">**. These were usually located at the top corner of a page, but were not always present or correct.

Intentional space

The **<space>** tag is used to mark up noticeable blank space left by the author. The attributes @unit (page, line, or word) and @quantity additionally record the extent of the space.

Gaps, omissions and illegible material

In additional to intentional blank space, there are also areas of missing or otherwise unreadable text due to, for example, ink smudges or physical damage.

Where missing or illegible text could not be reconstructed, the **<gap>** element was used, along with the @reason attribute to record the nature of the gap and, as with <space>, @quantity and @unit to indicate its size. These were usually small areas.

A number of areas of physicl damage were marked up with the **<damage>** element (usually also wrapping one or more <gap>). In this case the @agent attribute encoded some more detail about the cause of the damage.

In some cases where a single word or part of a word was illegible but could be inferred from context, **<supplied>** has been used.

Marginalia

There is no specific or exclusive TEI tag for "marginalia"; usage generally depends on the nature or function of the marginal text. We used a number of tags to encode Thornton's varied practices:

- marginal glosses or comments on the text are marked up with **<note>** and @place, along with @resp="#awt1" to ensure differentiation from editorial <note> tags.

- marginal text that functions as a heading or label uses **<fw>** or **<label>** and @place

- some textual insertions in the margin were marked upt with **<add>** and @place

The more substantial marginal items in <note> tags were ultimately recorded only in editorial notes in the DSE, due to space limitations and other technical issues, but as with other features the markup has been retained in the TEI XML.

Entities

We tagged a number of types of entity to provide enhanced contextual information for readers: dates, people and places and biblical quotations and references.

Dates

Calendar dates in the Books (such as "1 December 1662") were tagged using **<date>** as long as a year was stated or could be inferred with confidence from context. Dates written only as feast days, (eg "Michaelmas 1662"), were also tagged. Relative dates (such as "the next week") might be tagged, but only if they were particularly significant and there was no other nearby date mentioned for a significant event. Dates that did not state a year (and it could not be inferred) were not tagged.

Additional attributes were added to date tagging to provide standardised metadata. The dates were encoded in Extended Date/Time Format (EDTF) format which allows for varying degrees of precision: for example "1668-08-01" for "1 August 1668" or "1668-08" for "August 1668". Ian's English Calendar was used to find the calendar date for feast days.

Thornton was not consistent in beginning a new year on 1 January or 25 March ("Lady Day" dating). A method used by the Map of Early Modern London was adopted to encode the variations; they are also noted in editorial notes in the DSE. There are also some errors and inconsistencies in Thornton's dates, including inconsistencies across the books or dates that do not match external sources. No attempt has been made to correct these in markup, though they are noted in editorial notes.

Examples:

- <date when-custom="1648-11-30">30th of November: 1648</date>

- in the yeare <date when-custom="1641">1641</date>

- <date when-custom="1640-11">at the latter end of November</date>

- <date when-custom="1661-02-11" datingMethod="#JulianMar">feb 13th 1661</date>

People

Markup and linkage

Within the Books, we tagged potentially identifiable people mentioned in the Books who were Thornton's contemporaries or historical figures, including:

1. named individuals (eg "Mr Thornton")

2. unnamed individuals who could be identifiable (eg "my mother", "the midwife")

3. grouped people (eg "my two daughters"), as long as the individual members of the group could be identified from context.

We did not tag names from Biblical and fictional sources, abstract figures (including God or Jesus) or non-specific collective entities such as "my family".

Examples:

<persName> is used for references to people that use proper names

- <persName ref="ppl:et1">Elizabeth Thornton</persName>

<rs type="person"> or **<rs type="group">** for unnamed but identifiable references

- <rs ref="ppl:aow1" type="person">my mother</rs>
- <rs ref="ppl:cw1 ppl:aow1" type="group">my parents</rs>

Once tagging was completed, the tagged entities were extracted from the XML in order to build a list of distinct *people*. This entailed a) merging variant names for the same person (deduplication) and b) disambiguation of different people with the same name. Each person was assigned a short unique identifier, usually based on their initials (eg "awt1" for Thornton herself), which was subsequently added to every tagged mention in the Books using a @ref attribute.

The list formed the basis of all data linkage and biographical research. Some individuals (especially unnamed servants) were not identifiable; the markup has been retained in the TEI XML, but they are not linked.

Metadata

Short narrative biographies were written for as many individuals as possible. In addition, where available, more structured metadata was recorded for spouses, parents, children, birth and death dates, titles and married women's birth surnames, as well as the main sources used. These are

not present in the DSE but (along with lists of sources) have been included in the metadata files for reference.

We have striven to identify people as accurately as possible, but there may be errors in identification, deduplication or disambiguation, especially for obscure individuals and those with common names.

The metadata is held in a standoff file in the project Github repository, **people.xml** .

Example:

<person xml:id="dcl1">
<persName type="label">Daphne Lightfoot (died 1689)</persName>
<persName>
<forename>Daphne</forename>
<surname type="birth">Cassell</surname>
<surname type="married">Lightfoot</surname>
</persName>
<death type="burial" when-custom="1689-09-20">buried 20 September 1689</death>
<gender>female</gender>
<noteGrp>
<note type="bio">Daphne Lightfoot entered the service of the Wandesford family as an unmarried woman with the family name Cassell; the earliest reference is to her moving to Hipswell with the family in 1644...</note>
<note type="marriages">
<persName role="spouse" n="1" ref="#gl1">George Lightfoot</persName>
<persName role="child" n="1" ref="#ml1">Mary</persName>
</note>
<note type="sources">
<listBibl>
<bibl>Whiting, <hi rend="italic">Autobiographies of Thomas Comber</hi>, 2:53 </bibl>
<bibl>'The Will of Dame Alice Wandesford', in McCall, <hi rend="italic">Wandesforde Family</hi>, 357–58 </bibl>
<bibl>'Gilling West Parish Registers: Baptisms, Marriages and Burials, 1639–1782', PR/GIL 1/1, NYCRO</bibl>
</listBibl>
</note>
</noteGrp>
</person>

Places

Markup and linkage

The general approach to tagging and linking places mentioned in the Books was similar to that used for people's names.

We tagged potentially identifiable and mappable places mentioned in the Books,

- named ("Hipswell", "the river Swale")

- unnamed but identifiable ("the town", "the river")

Biblical and fictional places, abstract and indeterminate place mentions were not tagged.

Tagged mentions of places were linked to a reference list of distinct place names in a similar process to that used for people. The metadata is held in a standoff file in the project Github repository, **places.xml** .

Examples:

<placeName> for places described using proper names:

- <placeName ref="place:Kirk_131">Kirklington</placeName>

<rs type="place"> for places not described with a proper name but which are identifiable:

- <rs ref="place:Kirk_031" type="place">the Church</rs>

<geogName> for geographical features described using proper names:

- <geogName ref="place:Swal_123">the Swale</geogName>

<rs type="geog"> for geographical features not using names but identifiable:

- <rs ref="place:Iris_130" type="geog">that Tempestious Sea</rs>

Metadata

Short narrative descriptions were written for 44 of the more significant places. Additionally, places have wherever possible been linked to external reference data to provide additional metadata. Many places could be linked to Wikidata, which usually supplies geocoordinates and quite often links to other databases such as British History Online, Vision of Britain and Historic England. Wikidata linkage was carried out using OpenRefine and then checked manually. In some

cases, the National Library of Scotland's digitised historic OS maps were used to help find places that can no longer be located or have changed names. Please note that, as with people's names, there may be occasional errors in either identification or linkage to a map location.

The metadata is held in a standoff file in the project Github repository, **places.xml** .

Example:

<place xml:id="East_019" type="unit">
<placeName type="label">East Newton Hall, Stonegrave, Yorkshire</placeName>
<location>
<placeName>East Newton Hall</placeName>
<settlement>Stonegrave</settlement>
<region>Yorkshire</region>
<country>England</country>
<geo>54.207155481 -1.014613603</geo>
</location>
<noteGrp>
<note type="desc">East Newton had been home to the Thornton family since at least the fourteenth century but was fully rebuilt by William Thornton between 1652–62. It was Alice Thornton's home from 1662 until her death in 1707. The rear wing was entirely demolished c.1984–90 and replaced with a pastiche. Some seventeenth-century panelling and door furniture remains. Today, it is a privately-owned farm.</note>
<note type="sources">
<list>
<item>Pevsner and Grenville, <hi rend="italic">North Riding</hi>, 261-62.</item>
<item>George R. Keiser, 'Robert Thornton: Gentleman, Reader and Scribe', in <hi rend="italic">Robert Thornton and His Books: Essays on the Lincoln and London Thornton Manuscripts, ed. Susanna Fein and Michael Johnston</hi> (Woodbridge: Boydell & Brewer, 2014), 67.</item>
</list>
</note>
</noteGrp>
<linkGrp>
<ptr target="http://www.wikidata.org/entity/Q105790653/" type="wikidata"/>
<ptr   target="https://www.british-history.ac.uk/vch/yorks/north/vol1/pp561-566#p11/" type="victoria_county_history"/>
</linkGrp>
</place>

Biblical/Book of Common Prayer

While Thornton's expression is infused with biblical allusions, she only rarely provides specific references. For a modern reader, these allusions may pass unnoticed; using TEI, they not only

become visible but help to identify the specific sources she used.

Identification

Suzanne Trill was able to identify many of the biblical references. However, as it became clear that Thornton was familiar with several different translations, she also drew on the resources of 'e-sword' to search the Bishops' Bible (BB), the Geneva Bible (GB) and the King James Bible (KJV) to identify the appropriate source. She is also grateful to Bo Van Broekhoven for her assistance in identifying these materials.

In addition to the three bibles, Thornton draws heavily on the Book of Common Prayer (BCP): we used a 1693 EEBO-TCP edition (for which a TEI-XML version is available) as the main reference text for identification.

It was also evident that Thornton's references ranged from word-for-word quotation to general allusion, so our DSE distinguishes these as five different types (using the @type attribute):

- **direct**: verbatim/near verbatim quotation of an entire verse

- **paraphrase**: imprecise citation of a verse or more

- **partial**: verbatim/near verbatim quotation of at least half a verse

- **phrase**: short phrases (e.g., 'throne of grace'), precisely quoted

- **allusion**: refers to a verse, story or event but uses own words

Markup and linkage

The TEI **<quote>** element is used to markup passages of text drawn from a biblical source, while **<ref type="biblical">** is used for citations in text. Specific references are included using the @source attribute.

For biblical citations, the referencing is based on canonical names and conventional reference formats (book, chapter and verse). These appear in pop-up notes in the DSE.

In the TEI XML, the standard format of each reference is #source_book_chapter:verse(s). There may be more than one reference per @source; they are separated by a space.

Where the materials match a specific source, this is indicated by the following prefixes

- BB_
- GB_
- KJV_
- BCP_

Where the phrase is not specific to one translation, it is noted simply with an underscore in @source references and no prefix appears in the DSE.

No spaces were permitted within each reference, so for example "1 Kings" would be shortened to "1Kings". Otherwise, canonical references for book names were used.

Slightly different formats were required for the Book of Common Prayer. Psalms in the BCP Psalter could be simply referenced using the Psalm numbers. Elsewhere in the book, custom short codes were added to the project's TEI copy of the 1693 BCP to use in markup, which were later expanded to shortened versions of titles in the BCP to enable readers to identify the relevant section.

Example references:

- a single verse in a non-specific bible: #_Genesis_4:1

- a single verse in the Geneva Bible: #GB_Genesis_4:1

- a verse range in the King James Bible: #KJV_Genesis_4:1-2

- a Psalm in the BCP Psalter: #BCP_Psalter_Psalm1

- BCP, "the Litany, or General Supplication, to be sung or said after Morning Prayer upon Sundays, Wednesdays and Fridays...": #BCP,_The_Litany

- BCP, "Prayer with Thanksgiving to Almighty God, For having put an end to the Great Rebellion by the Restitution of the King and Royal Family, and the Restauration of the Government after many Years Interruption... ": #BCP,_Thanksgiving_for_the_Restoration

Events

Identification of events

For our purposes, an "event" was considered to be some significant phenomenon recorded by Thornton in at least one of the Books and which occurred at a (potentially) identifiable time and place. This could include inner experiences such as dreams or spiritual awakenings, though the majority were "external" events such as illnesses, births and deaths, wars and rebellions, and legal disputes.

An initial list of events was compiled using section headings and the indexes in *Book 1* and the *Book of Remembrances.* This starting point was then expanded by detailed reading and comparison of the four texts. It should be emphasised that there were significant subjective elements in the process, including the identification of events and choices made in breaking up some events into grouped sub-events.

Simply defining individual events within the texts could be difficult. Descriptions of events can be highly varied, encompassing anything from one sentence to several pages. Thornton's prose is complex, and one event may flow into another - and back again - in ways that defy pinpointing for markup. Events can become fragmented; one event might be mentioned very briefly in the middle of a more substantial narrative of another event, or Thornton might interrupt describing an event with a prayer or meditation and then resume her narrative flow. Sometimes she would jump around chronologically, even in the mainly linear narrative of *Book 1* and the *Book of Remembrances*; *Book 2* and *Book 3* were even less straightforwardly organised.

Not all possible events were of equal significance to Thornton; her pregnancies, the births and deaths of her children, the legal and financial challenges she faced and her family's experiences of the Wars of the Three Kingdoms are among a number of major events that she describes repeatedly and often in considerable detail. It is these important and repeatedly described events that are of most interest to the project, but they're often also the most complex and challenging to model as data.

Some simplification of this complexity was necessary. In particular, overlapping markup, or marking up any passage of text as more than one event, was completely avoided. The markup of very short mentions within longer event narratives were also avoided unless the fragment was felt to be particularly important.

Finally, the dating of events - including some of vital importance in Thornton's life - was also not always straightforward. Most events mentioned at any length in *Book 1* and the *Book of Remembrances* could be dated with reasonable accuracy and precision (although Thornton's own dating was not always entirely reliable). However, a number of significant events were primarily or exclusively related in *Book 2* and *Book 3* and Thornton recorded dates much less consistently in these two books.

Markup and linkage

There are often tensions between the complexities of these flowing narratives of life events and the demands of data structures for order and precision. They pose a particular problem for XML because it has a strictly hierarchical nested structure, but the events don't always fit neatly into that: for example, a mention of an event might start in the middle of one paragraph and finish in the middle of another. TEI offers a number of methods for handling the problem of overlapping structures, but they often have downsides, particularly that they're less straightforward to work with and to process than nested tags. We used a boundary marking technique in which the beginning and end of an event mention are marked up with "empty" tags that are linked using a special @spanTo attribute, so that every marked up mention of an event in the TEI XML has a unique identifier. These event mention identifiers were in turn linked to event identifiers in our database.

Metadata

Event metadata was kept quite minimal.

- each event was assigned a unique event identifier

- if a major event was split into sub-events, each group was also assigned an identifier

- events in the database were categorised as "single" (standalone events), "sub", or "parent"

- each event was given a short descriptive title

- each event was given a date, date range or notBefore/notAfter pair of dates, though some of the dates had to be very imprecise or uncertain

- each event was assigned a thematic keyword (eg "illness", "legal/financial")

The metadata is held in a standoff file in the project Github repository, **events.xml** .

Example:

A mention in the texts:

```
<milestone spanTo="#ev1-end" xml:id="ev1-start" n="ev1" unit="event"/>
```

text of event narrative...

<anchor xml:id="ev1-end" n="ev1" type="event"/>

Metadata:

<event xml:id="awt1_1626_birth" type="sgl" when-custom="1626-02-13">
<desc>Birth of Thornton</desc>
<label>birth/baptism</label>
<linkGrp type="sgl">
<ptr target="bookrem:ev1" type="book" subtype="bookrem"/>
<ptr target="book1:ev1" type="book" subtype="book1"/>
<ptr target="book2:ev64" type="book" subtype="book2"/>
<ptr target="book3:ev3" type="book" subtype="book3"/>
</linkGrp>
</event>

TEI Modernisation

VARD and regularisation of spelling

We used VARD 2 (version 2.5.3), a software tool developed by Lancaster University, to help with the modernisation of variation in Thornton's spelling. The software was used in single-text, manual mode. Given an early modern text, VARD detects variant spellings and offers a list of alternative modern spellings as replacements. It is possible to normalise all instances of a word simultaneously, if there is only one possible replacement, or to check and change them individually. Alternative options include marking words as not variants, which could be used for words that need some other attention or query. When finished, the VARDed version of the text is saved as an XML file. The XML tags added by VARD are not TEI-compliant, so the final step in the process was to convert it back to TEI with an XSLT script.

VARD has a number of limitations. It is essentially a spell-checker, so it can't find variations that also happen to be valid modern spellings. For example, Thornton often writes the word "borne", as well as frequently splitting some words into two, including "a bout" and "with out". These had to be searched and corrected manually afterwards. We also did not try to regularise capitalisation using VARD, as it would have been much more time-consuming and not always clear at this early stage what the final result should be. VARD was a useful tool, but it was only the first step in the modernisation process, as detailed in editorial principles above.

Other limitations and caveats cannot be blamed on the software. There could be grey areas between regularising variation and correcting error, considering just how variable early modern spelling could be. A broad view of variation was taken: if a non-modern form was consistent with the kinds of phonetic variation understood by VARD and known to the team from other early modern manuscript and printed sources, then it would be treated as a variation rather than error. This includes, for example, "e" at the end of a word, the use of "'d" or simply "d" rather than "ed" for past tense, "ei" instead of "ie" (notably in words like "freind"), "y" for "i", and the omission of apostrophes in possessive nouns.

Examples of error, on the other hand, would include clear misspellings like "pumishet" for "punishment" or "tradegy" for "tragedy". However, in practice the distinction between variation and error was not applied quite as consistently as an editor would like (and late reviews indicated small numbers of errors and inconsistencies that there was insufficient time to correct). Addi-

tionally, some of Thornton's interesting orthographic idiosyncrasies, such as writing "i" for "c", were not approached as consistently as they could have been.

Modernisation: method and markup

Rather than creating two separate versions of the text that would be difficult to keep in sync, our approach was to use "parallel" markup (TEI <choice> and other forms of markup that enable the recording of multiple versions). An alternative could have been to use (or abuse) TEI's critical apparatus for recording variants of a text, but the team lacked the expertise to evaluate whether this method would be appropriate or practicable. The approach we did take was flexible enough to allow for a multitude of very small edits in order to produce a readable modernised DSE version.

Broadly speaking, markup for modernisation involved three types of action: *removal*, *addition* and *substitution*. The bulk of modernisation actions fell into two categories: spelling including capitalisation and punctuation.

Removal

<**surplus**> was used where we wanted to completely hide material from the modernised view. The @reason attribute encodes different reasons for removal:

- "superfluous", mainly used for punctuation

- "repetition", specifically for erroneously repeated words (highlighted in SD)

There were two significant markup choices implemented to resolve specific technical difficulties.

<**pc**> without any attributes was used to wrap hyphens before and after <lb> tags, in order to resolve a problem with closing up spaces in the modernised version.

<**p rend="modernised(append-following-sibling)"**> was used to enable us to remove paragraph breaks as needed in the modernised version, without breaking the TEI XML or the SD version. It was applied to the first of any pair of paragraphs that needed to be amalgamated.

Addition

<**supplied**> was used to add material. Again this was used with the @**reason** attribute:

- "mod" - the main general purpose value, primarily used for punctuation

Smaller numbers of other specific reasons:

- when a word or letter had been omitted from the original (and could be reconstructed from context)

- when a word or letter was obscured by marks or damage on the page (and could be reconstructed from context)

Examples:

- again<supplied reason="omitted">st</supplied>

- not to bid him<supplied reason="mod">,</supplied> but could

Substitution

Various methods were employed to replace original text with modernised forms.

<**w** @**norm="(modern spelling)"**> was used to regularise spellings, including capitalisation.

- <w norm="soldier">Souldier</w>

<**pc** @**norm="(modern punctuation)"**> was used to regularise punctuation.

- <pc norm=",">.</pc>

This usage of the @norm attribute, although it's technically valid TEI, represents a departure from the TEI guidelines, which explicitly state that it is "meant for strictly lexicographic and linguistic uses, and not for editorial interventions". Our pragmatic justification for this is that it was clear from the beginning that there would be many thousands of changes needed (in fact there are more than 70,000 uses of <w>/@norm), and it's considerably simpler to apply (and less verbose) than the approved method (using <choice><orig><reg>).

We did nonetheless make extensive use of the <**choice**> "family" of tags, where <choice> is used as a wrapper to group "alternative encodings" of a text together.

For correction of the author's errors: <**sic**> is used to tag the original and <**corr**> for the editorial correction.

To expand abbreviations, <**abbr**> tags the original and <**expan**> the expanded version.

Examples:

- his <choice><sic>Kingome</sic><corr>kingdom</corr></choice>

- my <choice><abbr>Hon.<am rend="superscript">rd</am></abbr><expan>honoured</expan></choice>

- <choice><abbr>42</abbr><expan>1642</expan></choice>

<**measure**> has been used specifically to tag sums of money so that they could be reformatted in the DSE. There are two main variants: the first replaced Thornton's use of superscript l with modern £, and the second replaced money written out as words with £xxx.

Examples:

- <measure type="currency" unit="£">900<am rend="superscript italic">l</am></measure>

- <measure type="currency" unit="£" quantity="100">one hundred <unit>pound</unit></measure>

Editorial Notes and Glosses

Much information that would be provided using editorial annotation in a print edition is handled rather differently in a digital edition, using the affordances of markup. Unless specific detail needs to be added to a particular instance, the following types of information are not usually indicated with editorial notes in the DSE:

- names, places and biblical references

- word errors, omissions, accidental repetitions and spelling variations

- authorial insertions and deletions

- areas of damage, gaps and blank space

There are two types of editorial annotation: notes and glosses. (In both cases, the actual text of the annotation is held in a linked standoff container or file rather than at the location of the annotation itself.) Both types of annotation are displayed in popups in the DSE.

Notes

Notes are used to supply specific detail at a certain point in the text, including

- to record the author's marginal notes, which could not be included in the DSE itself because of technical constraints on space

- to add historical and documentary context

- to clarify ambiguity

- to note cross-references to the other Books

There are two linked pieces of markup for notes. Firstly, the point in the text where the annotation applies is tagged with an empty tag **<anchor resp="ednote" corresp="#**

id

"/>. Then, in a **<standOff>** section at the bottom of the XML file there is a corresponding **<note xml:id="**

$$id$$

" type="note" place="end"> which contains the text for the note.

Glosses

These are annotations that gloss the meaning of words or short phrases in the text, such as archaic meanings, legal terms, some specialised or regional terms and so on. Most definitions are drawn from the Oxford English Dictionary. Instead of a number pointer in the text, the relevant segment of text is tagged with **<term ref="gloss:**

$$id$$

">. The definitions are stored in a standoff file for each Book, in which each definition is tagged with **<item xml:id="**

$$id$$

">.

The standoff files are stored in the entities folder of the Github repository, named **book_(number).xml** .

Resources

Transcription and editing

Barber, Charles Laurence. *Early Modern English.* Edinburgh: Edinburgh University Press, 1997.

Chrystal, David and Ben Crystal. *Shakespeare's Words: A Glossary & Language Companion.* London: Penguin, 2002.

Marcus, Leah S. *Unediting the Renaissance: Shakespeare, Marlowe and Milton.* London & New York: Routledge, 1996.

Ross, Sarah C. E., and Paul Salzman, eds. *Editing Early Modern Women.* Cambridge: Cambridge University Press, 2016.

Key TEI references

- TEI Guidelines
- Women Writers Project Guide to Scholarly Text Encoding
- TEI By Example
- Lou Burnard, What is the Text Encoding Initiative?

Other editions

Apart from the guides noted above, various TEI projects' public guidelines or editorial policies were consulted at various stages. Particularly helpful examples were:

- Newton Project
- Casebooks

- Bess of Hardwick

- Map of Early Modern London

- Walt Whitman Archive

- Mary Hamilton Papers

Books and manuscripts

- Early Printed Books

- ArchBook

- Andrew Dunning, Transcribing medieval manuscripts with TEI

- Laura Estill, 'Encoding the Edge: Manuscript Marginalia and the TEI', *Digital Literary Studies* 1 (2016)

- Erik Kwakkel, *Books Before Print* (Amsterdam University Press, 2018).