

TEI by example project team meeting, KANTL, 2007/03/05

Present: Edward Vanhoutte (EV), Melissa Terras (MT)

Minutes: Ron Van den Branden (RVDB)

1. State of affairs

1.1 Actions taken

Following actions were taken after the preceding meeting:

- 13/12/2006: call for examples on TEI-L, DM, Humanist, Digital Classicist, projects using TEI
- poster submission for DH2007 (approved)
- EV + RVDB: identify module goals

1.2 License

After debate following the original call for examples on TEI-L, the original Creative Commons Attribution Non-Commercial No Derivatives licence (<<http://creativecommons.org/licenses/by-nc-nd/2.5/>>) was abandoned. Counter-intuitively, the non-commercial restriction appears to have rather restrictive effects. Therefore, a revised call for examples has been issued, adopting the more liberal Creative Commons Attribution ShareAlike license (<<http://creativecommons.org/licenses/by-sa/2.5/>>).

1.3 Examples

So far, the call for examples has only generated 6 responses (see details in appendix A):

- Tim Finney: *"Manuscript Markup," in The Freer Biblical Manuscripts*
1 P5 (+ msDescription), language: Koiné Greek
<http://www.sbl-site.org/Resources/Resources_ManuscriptMarkup.aspx>
- Gabriel Bodard: *Inscriptions of Aphrodisias*
20 P4, sample from 250 texts
- Ross Scaife: *neo-Latin colloquia*
12 P4
<<http://www.stoa.org/colloquia/>>
- Martin Holmes: *Early Modern Literary Studies Vol 1-4*
186 P5 (< P4), born digital
<<http://lettuce.tapor.uvic.ca/cocoon/journals/emls/toc.htm>>
- Janusz S. Bien: *Polish language of the 1960s*
5 P4 collections of +/- 2000 50 word samples, morphologically encoded
1 P4 teiCorpus header
<http://www.mimuw.edu.pl/polszczyzna/pl196x/index_en.htm>
- Eric Lease Morgan : *Alex Catalogue*
85 P4, 50% transcriptions (Gutenberg depth) 50% born digital
<<http://infomotions.com/morgan-tei.tar.gz>>

2. Points of attention

2.1 Increase public response rate

The moderate response to our call for examples calls for action; in the meantime, parties and persons that had expressed their willingness to support have been kindly reminded to their engagements. One factor for this low level of public response is probably the generality and open-endedness of the call for examples. This can be addressed in calls for more specific text samples, depending on the needs identified when developing the tutorial modules. Following approach looks reasonable:

- set out developing the content of the tutorial modules, using whatever is usable in the available examples, or use home-grown examples if needed
- identify specific needs and pro-actively¹ trace possible examples, contact their right holders with specific inquiries to use them
- use home-grown examples if needed

EV suggested to explicitly use multilingual examples (preferably using multilingual TEI markup) in order to maximise the tutorials' appeal to non-English speakers interested in using TEI. It would probably make sense to include such examples systematically as further exercises.

2.2 Discuss didactic approach for tutorial modules

The main aim of this meeting was a more detailed discussion of the didactic approach for the tutorial modules. The following section provides an overview of a suggested approach.

2. Tutorial modules

2.1 Sample module structure

Previously, it has been agreed that each module should consist of:

1. meta-introduction: objectives, prerequisites (,system requirements)
2. thematic introduction: high-level overview
3. step-by-step introduction of new markup: short text excerpts and examples
4. summary
5. download area: sample code (schemas, ODD)
6. resources: links to background information and reference material

2.2 Module goals

A main point of attention is the need to avoid repetition of the TEI Guidelines. Instead of being theory-based, the TBE² modules therefore should explicitly be oriented towards use-cases and examples. By consequence, this implies selecting and cutting down the topics as well as the level of detail of their discussion. All should be well as long as choices are indicated (and/or motivated) and reference is made to the full story in the TEI Guidelines. In this light, it is probably more fruitful to organise the modules around possible topics to be addressed, rather than lists of elements.

1 In this respect, EV pointed out that the Women Writers Project (<<http://www.wwp.brown.edu/>>) will open up its text base during the month March 2007. MT suggested to contact Wendell Piez and one of her students for poetry examples.

2 In order to distinguish the notions *TEI modules* and their treatment in the *TEI by Example modules*, the latter is indicated as *TBE module* in the remainder of this text.

2.2.1 Module 1: Introduction

Proposed concepts	Proposed elements
Text encoding	
Procedural vs descriptive markup	
XML – well-formed XML – valid XML (DTD / Schema ~> ODD)	
TEI: history and structure	

During the meeting, the initial proposal to include treatment of global TEI phrase-level elements in the introductory module was reorganised. The introductory TBE module will be restricted to a general introduction to text encoding and representation, XML and a situation of TEI as a specific XML vocabulary. The second module, then, will be reoriented as an introduction to 'TEI essentials' (general text structure, the header, general phrase-level elements), in order to provide a well-balanced structure integrating all common TEI elements.

2.2.2 Module 2: TEI essentials

Minimal structure	Mandatory elements
<pre> <TEI> <teiHeader> <fileDesc> <titleStmt> <title></title> </titleStmt> <publicationStmt> <p></p> </publicationStmt> <sourceDesc> <p></p> </sourceDesc> </fileDesc> </teiHeader> <text> <body></body> </text> </TEI> </pre>	<ol style="list-style-type: none"> 1. TEI 2. teiHeader 3. fileDesc 4. titleStmt 5. title 6. publicationStmt 7. p 8. sourceDesc 9. text 10. body
Proposed concepts	Proposed elements
TEI document structure	<ol style="list-style-type: none"> 1. TEI, teiCorpus 2. teiHeader 3. text
TEI header: minimal – proposed – maximal	<ol style="list-style-type: none"> 1. teiHeader 2. fileDesc 3. titleStmt 4. title 5. author editor funder principal respStmt

	<ul style="list-style-type: none"> 6. publicationStmt 7. publisher distributor authority pubPlace date idno availability 8. sourceDesc 9. p 10. bibl/title bibl/author bibl/publisher bibl/date 11. encodingDesc 12. projectDesc, projectDesc/p 13. editorialDecl, editorialDecl/p editorialDecl/list 14. tagsDecl 15. tagsDecl/rendition, tagsDecl/namespace 16. profileDesc 17. profileDesc/creation, profileDesc/langUsage, profileDesc/textClass 18. revisionDesc, revisionDesc/change
General TEI text structure	<ul style="list-style-type: none"> 1. text, group 2. body, front, back 3. div 4. head
Global attributes	<ul style="list-style-type: none"> 1. @xml:id 2. @xml:lang 3. @rend 4. @n
General phrase-level elements:	
data structures	name / rs / title / address / date / abbr – expan / q / bibl* / ...
typographical phenomena	hi / foreign / emph / term / gloss / mentioned / ...
simple editorial changes	corr - sic / reg – orig / add / del / unclear
structural elements	opener / closer / epigraph / table / list / figure / graphic / pb / lb
TEI modular architecture	

This TBE module will introduce the common ground for all TEI documents. If this is done in general terms, other TBE modules can either refer to or literally include these sections (or at least provide that option to the user).

Due to its high level of abstraction and organisation (featuring many optional but dependent structures) treatment of the TEI header requires a different approach from other TEI concepts, by probably following the TEI Guidelines more closely. Since the TEI header consists of 4 major parts (file description, encoding description, profile description and revision description), it makes most

sense to organise the tutorial module along this structure. One option could be to illustrate how the bibliographic information on a sample title page of a book can be encoded in a minimal TEI header, and extend this to a proposed header (featuring a proposed selection of optional parts) and even a full header (featuring all optional parts). This can be illustrated either via gradual elaboration of this same example, or via separate examples. If possible, EV suggested that packages with further exercises could be requested by email, and solutions be made accessible on non-linked parts of the website.

For treatment of the common TEI phrase-level elements in the this TBE module, the TEI Guidelines could suggest following topics:

- common phrase-level elements (chapter 6):
 - Highlighting and Quotation
 - Simple Editorial Changes
 - Names, Numbers, Dates, Abbreviations, and Addresses
 - Simple Links and Cross References
 - Notes, Annotation, and Indexing
 - Graphics and other non-textual components
 - Reference Systems
 - Bibliographic Citations and References

Another issue is the treatment of more specific phrase-level elements that are defined in separate TEI modules. Since the project proposal lists these modules for nearly all TBE modules, it makes sense to include them in the this TBE module. The relevant chapters of the TEI Guidelines suggest following structure:

- linking, segmentation and alignment (chapter 14):
 - Links
 - Pointing mechanisms
 - Blocks, Segments and Anchors
 - Correspondence and Alignment
 - Synchronization
 - Identical Elements and Virtual Copies
 - Aggregation
 - Alternation
 - Stand-off Markup
 - Connecting Analytic and Textual Markup
- names and dates (chapter 20):
 - Personal Names
 - Place Names
 - Organization names
 - Biographical and Prosopographical Data
 - Dates and Times

Although these TEI modules merely provide extensions to core TEI elements, they are considered relevant to be included in this TBE module. This does not imply that all TEI elements should be treated exhaustively. On the other hand, it should be stressed in the tutorials that the use of these elements requires selection of the relevant TEI modules in the TEI Customisation phase.

2.2.3 Modules 3 to 5: prose – poetry – drama

- Prose:

Minimal structure	Mandatory elements
-------------------	--------------------

<body> </body>	1. body group/text/body 2. div p
Proposed structure	Proposed elements
<front> </front> <body> <head></head> <div type=""> <p></p> </div> </body> <back> </back>	1. front / back 2. body 3. div 4. head 5. p

- Poetry:

Minimal structure	Mandatory elements
<body> </body>	1. body group/text/body 2. div lg l
Proposed concepts	Proposed elements
Poetry structure	1. body group/text/body 2. div lg l
Specific poetry elements	1. caesura 2. rhyme 3. @met, @real, @rhyme, @enjamb

- Drama:

Minimal structure	Mandatory elements
<body> </body>	1. body group/text/body 2. div sp
Proposed concepts	Proposed elements
Drama structure	1. body group/text/body 2. div lg l
Specific drama elements	1. set 2. prologue, epilogue 3. performance 4. castList, castGroup, castItem 5. role, roleDesc 6. actor 7. move 8. view 9. camera 10. sound 11. caption

With the treatment of common TEI phrase level elements isolated in the TBE essentials module, these 3 modules can have a similar structure. By illustrating different aspects of encoding texts from each genre in TEI, lots of genre-specific TEI elements should be introduced. For the general text-level elements, reference can be made to the relevant section in the TBE essentials module (as assumed knowledge), or this content can be included (perhaps at user request) into these modules. An important issue is the identification of topics to be treated: specific genre types (different structural types), illustrations of genre-specific text phenomena,.... For poetry and drama, the organisation of current TEI Guidelines' chapters could be informative:

- Poetry (chapter 9):
 - Structural Divisions of Verse Texts
 - Components of the Verse Line (<seg>, <caesura>)
 - Rhyme and Metrical Analysis (@met, @real, @rhyme)
 - Rhyme (<rhyme>)
 - Encoding Procedures For Other Verse Features
- Drama (chapter 10):
 - Front and Back Matter (<set>, <prologue>, <epilogue>, <performance>, <castList>)
 - The Body of a Performance Text (<sp>, <speaker>, <stage>, <move>)
 - Other Types of Performance Text (<view>, <camera>, <caption>, <sound>, <tech>)

Since the TEI Guidelines consider prose the default genre, there are nearly no genre-specific elements apart from paragraphs and lists. This leaves following topics from the TEI Guidelines that could be informative for the contents of the TBE prose module:

- Prose (chapter 6):
 - Paragraphs
 - Treatment of Punctuation
 - Lists (and tables)

2.2.4 Module 6: manuscript transcription

Minimal structure	Mandatory elements
<body> </body>	1. body group/text/body 2. div p lg l sp
Proposed concepts	Proposed elements
Primary source markup tags	1. addSpan 2. delSpan 3. restore 4. supplied 5. hand 6. handShift 7. handList 8. damage 9. space 10. fw

This TBE module will cover the TEI modules for the transcription of primary sources. Use of these elements should be illustrated in combination with that of the common TEI phrase-level elements discussed in the TBE essentials module.

2.2.5 Module 7: scholarly editing

Minimal structure	Mandatory elements
<body> </body>	<ol style="list-style-type: none"> 1. body group/text/body 2. div p lg l sp
Proposed concepts	Proposed elements
Apparatus + parallel segmentation (OR all methods?)	<ol style="list-style-type: none"> 1. app 2. rdg 3. lem 4. witList 5. witness 6. variantEncoding

This TBE module will cover the TEI modules for critical apparatus. The major point of interest will be the illustration of the concepts involved, rather than in-depth analysis of all possible mechanisms to encode textual variation. This probably means that the *parallel segmentation* method will be taken as the standard practice in the discussion, while both alternatives (viz. the *location-referenced* method and the *double-end-point-attached* method) can be illustrated depending on available examples.

2.2.6 Customizing TEI

Minimal structure	Mandatory elements
<body> <schemaSpec></schemaSpec> </body>	<ol style="list-style-type: none"> 1. body group/text/body 2. schemaSpec 3. moduleRef specGrpRef classSpec elementSpec listRef macroSpec moduleSpec specGrp
Proposed concepts	Proposed elements
Roma	Some use cases for customizing TEI: <ol style="list-style-type: none"> 1. select modules + elements 2. change element names 3. add elements
ODD	Treat tagsdoc module fairly cursory (to provide notions) <ol style="list-style-type: none"> 1. schemaSpec 2. moduleRef

The main focus of this TBE module will probably be situated in explaining possible use cases for customisations of TEI. Considering the complexity and technical nature of the TEI tagsdoc module, I believe this TBE module can suffice by providing insight into the most important elements of a TEI ODD file. Probably Laurent Romary's talk³ at the Methods Network workshop *Advanced Text Encoding with TEI P5* (Oxford, 19/09/2006) could provide a good angle to tackle this subject:

- introduce the Roma interface
- present increasingly complex usage scenarios, introducing different concepts of customising

3 Laurent Romary, *Uncovering the TEI and ODD, a pedagogical strip-tease*. Talk at the AHRC ICT Methods Network workshop *Advanced Text Encoding with TEI P5*, 19/09/2006. <<http://tei.oucs.ox.ac.uk/Oxford/2006-09-methNet/Talks/RomaJourney.ppt>>

- TEI (selecting TEI modules, selecting TEI elements, adding and changing elements)
- discuss how these concepts are expressed in the generated ODD files

3. Time line

The goal was set to complete 3 TBE modules for the DH2007 conference in June 2007 (completion before the end of May). This prioritised list was agreed on:

1. Poetry
2. Essentials
3. Prose
4. Drama
5. Customisation
6. Manuscript transcription
7. Scholarly editing
8. Introduction

Half April 2007 was set forth for completion of first test versions of the Poetry module.

4. Miscellaneous

- MT proposed to look for a dedicated logo for the project, depending on the price.
- For testing procedures, MT reported that her students could be a good test panel, and that Julia Flanders should be contacted.

5. Action points

- All: more pro-active search for poetry examples
- EV: develop a proof of concept for the TBE Poetry module
- EV: inquire for dedicated logo
- RVDB: develop a proof of concept for the common phrase-level elements
- RVDB: develop first test versions for the TBE Poetry module

