Sequentiaility in Genetic Digital Scholarly Editions. Models for Encoding the Dynamics of the Writing Process

Wout Dillen
wout.dillen@uantwerpen.be
University of Antwerp, Belgium

When TEI P5 version 2.0 was published in 2011, scholar-
ly editors who are interested in the writing process of lit-
ery works gained an important instrument for encoding
their genetic Digital Scholarly Editions in TEI-conformant
XML. After a long process of deliberation, this version of
the TEI’s encoding schema incorporated a large number of
modifications proposed by the TEI MS SIG’s Workgroup
on Genetic Editions that aimed to re-evaluate the existing
TEI tagset in order to facilitate the encoding of genetic
phenomena (TEI Consortium, 2011). The Workgroup’s
‘Encoding Model for Genetic Editions’ (2010) reveals two
major points of interest in this proposal: (1) the need for
the ability to encode features of the document rather than
those of the text; and (2) the need for the ability to encode
time, sequentiality and writing stages in those documents’
transcriptions.

The main answer to the first point of interest was the
introduction of the <sourceDoc> element (as well as its
<surface> children), that was allowed to exist on the
same hierarchical level as the <teiHeader>, <facsimile>,
and <text> elements. Since the Text Encoding Initiative
has (as its name implies) historically favoured ‘text’ over
‘document’, this can be regarded as a powerful statement
to the TEI community that documents are as valuable as
texts in textual scholarship, and that it should be possible
to transcribe them as such. As a result, this encoding model
has been gratefully adopted by editors who are taking a
more document-oriented approach to the transcription of
their materials – like those of the Shelley Godwin Archive
(S-GA) for instance (Shelley, 2013). The question remains,
however, whether the use of this vocabulary is enough to
While the document will take up a central position in any
genetic edition, the use of the ‘Genetic Editions’ document-
oriented transcription model is not a distinctive feature of
the genetic edition in itself.

The Workgroup’s second point of interest (the encod-
ing of ‘time’) is much more central to genetic criticism. In
‘The Open Space of the Draft Page’, Daniel Ferrer makes a
compelling argument that ‘the draft is not a text […]’, it is a
protocol for making a text, comparing it to a musical score
that, though by itself inherently mute, can be interpreted
as a set of instructions for a future performance (1998,
261). Likewise, a draft document leaves the writer with a
set of instructions that help her transport the unfinished
text from one writing stage to the next. The interpreta-
tion of these instructions, and of the distinction between
different versions and writing stages, is one of the most
important tasks of genetic criticism. This is what makes
sequentiality such a key aspect of genetic editing: only
by interpreting the draft materials as an interconnected
sequence of writing acts can we expose the dynamics of
the author’s writing process.

There are many ways of encoding this sequentiality
in the transcriptions of draft materials, across varying
levels of granularity. The Workgroup’s suggestion to use
the <change> element to highlight distinct revision cam-
paigns, for instance, effectively differentiates between
individual versions of the same text when they are found
within a single document. As Pierazzo and André’s ‘Proust
Prototype’ demonstrates, this method can even be em-
ployed to sequence individual stages within a single version
(2012). Going even further, projects like the CD-ROM
edition of Willem Elsschot’s Achter de Schermen (2007)
and the Melville Electronic Library’s TextLab software
(2009) analyze what John Bryant has called the inter-
nal ‘revision sequences’ of individual sentences (Bryant,
2008). The danger of analyzing the writing process on
this small a level, however, is that the mechanics of the
writing process may start to interfere with the dynamics
of that writing process. From a genetic perspective, it is
more important to expose the dynamic relation between
the textual elements involved in a modification (e.g. ‘this
is a substitution’) than the mechanical order in which that
modification was made (e.g. ‘first this word was deleted,
then this other word was added’). Since the exact writ-
ing sequence of such a modification is often impossible
to reconstruct with any degree of certainty, consistently
analyzing and sequencing all the work’s revision sites may
introduce a number of hypotheses in the edition that the
editor is not necessarily comfortable with committing to.

On the other side of the spectrum, analyzing larger
macrogenetic processes across documents, the ‘Encoding
Model for Genetic Editions’ refers to the TEI’s ‘Graphs,
Networks, and Trees’ module, suggesting to encode the re-
lations between documents as the <arc>s between <node>s
in a <graph> element. Depending on the complexity of
the writing process, this <graph> may result in an intricate
data structure that can be used to visualize the chronology
of the writing process on a highly abstract level. For writ-
ing processes that are less complex on the macrogenetic
level, however, this model may be too much pain for too
little gain, as a manually designed timetable could also
do the trick.

The Beckett Digital Manuscript Project’s approach to
encoding sequentiality into its genetic Digital Scholarly
Edition of Samuel Beckett’s works tries to seek a middle
ground between these two extremes: rather than analyzing
the way in which individual sentences were written, the
BDMP’s encoding model allows the user to discover how
those sentences were changed from version to version, across different documents. By linking related semantic clusters on the sentence level across versions, this model allows for the on the fly generation of a chronological overview of all the different versions of each sentence in the corpus. As such, this model combines the ability of comparing different versions of the same work of more macrogenetically oriented approaches with the higher granularity of more microgenetically oriented approaches.

After illustrating the challenges and opportunities of these different models of encoding sequentiality in genetic editions, this paper will demonstrate how the BDMP transcribes its genetic materials in view of visualizing their sequentiality in the edition’s ‘Synoptic Sentence View’ (see ‘Figure 1’). The paper will conclude by presenting an example of how this encoding model may also be used to interpret the macrogenetic writing sequence of individual documents by means of an animated visualization of the writing process of the first draft of Beckett’s L’Innommable.

Figure 1: BDMP Synoptic Sentence View

Bibliography


correspSearch - A Web Service to Connect Diverse Scholarly Editions of Letters

Stefan Dumont
dumont@bbaw.de
Berlin-Brandenburg Academy of Sciences and Humanities, Germany

Letters are an important historical source: First, they may contain comments from contemporaries about the most different events, persons, publications, and issues. Second, letters allow insights about connections and networks between correspondence partners. So, questions occur which can only be answered across the borders of scholarly letter editions due to the fact that these editions are usually focussed on partial correspondences (of a certain person or between two specific persons). But this requires time-consuming searches across various letter editions. This has been a well-known problem for quite some time, now (Bunzel, 2013: 117) and has already evoked work on a few databases dedicated to correspondence, like e.g. “Early Modern Letters Online”. But these databases have