



Conference Programme

Book of Abstracts



Table of Contents

Conference Schedule (Dates and Rooms)	7
Programme	8
Tuesday, 11 October 2011	8
Wednesday, 12 October 2011	8
Thursday, 13 October 2011	8
Friday, 14 October 2011	8
Saturday, 15 October 2011	9
Sunday, 16 October 2011	9
Keynote Presentations	10
Edward Vanhoutte (Gent): "So You Think You Can Edit? The Masterchef Edition"	10
Andrea Rapp (Darmstadt): "From text technology to cultural technology: the role of the TEI in Virtual Research Environments"	11
Papers	12
A TEI schema for the representation of CMC discourse	12
Beißwenger, Michael; Ermakova, Maria; Geyken, Alexander; Lemnitzer, Lothar; Storrer, Angelika	12
Representing genres of computer-mediated communication in TEI	13
Beißwenger, Michael; Lemnitzer, Lothar	13
Challenges of representing genres of computer-mediated communication in TEI: The linguistic perspective	14
Beißwenger, Michael; Storrer, Angelika	14
The electronic edition of the corpus written by Thomas Le Roy about the history of the Mont Saint-Michel, using the TEI	16
Bisson, Marie	16
Two sides of the same medal? Remarks on diplomatic and textual encoding in the Faust Edition .	17
Brüning, Gerrit; Henzel, Katrin; Pravida, Dietmar	17
Improving the Usability of Corpus Markup and Analysis Tools by Studying their Presentation Layer	18
Burghardt, Manuel; Fuchs, Markus; Wolff, Christian	18
Glossing music theory: how to make transparent the web of quotations, authorities and allusions in medieval texts	20
Desmond, Karen	20
Solving Problems for Online Diplomatic Editions of Medieval Manuscripts	21
Fredell, Joel Willis; Borchers, Charles W.; Ilgen, Terri Jo	21



Metadata customization with ODD.....	23
Gaiffe, Bertrand François.....	23
A large scale critical edition: first translation of St Augustine's City of God by Raoul de Presle	24
Gaiffe, Bertrand François; Stumpf, Béatrice.....	24
The Canary in the Text Mine: Analysis of the data mining of TEI-encoded texts in MONK research software.....	24
Green, Harriett Elizabeth.....	24
Measuring the correctness of double-keying: Error classification and quality control in a large corpus of TEI-annotated historical text.....	27
Haaf, Susanne; Geyken, Alexander	27
CMC as a component of a balanced, TEI-encoded corpus representing contemporary German: goals, motivation, design issues.....	27
Lemnitzer, Lothar; Geyken, Alexander; Beißwenger, Michael; Storrer, Angelika.....	27
Collaborative & non-deterministic markup: the CLÉA project.....	28
Meister, Jan Christoph; Petris, Marco	28
Faust: Multiple Encodings and Diplomatic Transcript Layout.....	29
Middell, Gregor; Wissenbach, Moritz	29
Creating lexical resources in TEI P5. Experiences from building multi-purpose digital dictionaries	30
Moerth, Karlheinz; Budin, Gerhard	30
The TEI encoding of textual fragments : dangerous wager or efficient stratagem ?.....	31
Morlock-Gerstenkorn, Emmanuelle.....	31
A Register of Baroque and Enlightenment Slovenian Manuscripts: TEI encoded Analyses and Editions.....	32
Ogrin, Matija; Erjavec, Tomaž; Javoršek, Jan Jona	32
Creating, enhancing and analyzing TEI files: The new, XML-based version of TUSTEP	33
Ott, Wilhelm; Ott, Tobias	33
The Role of Technology in Scholarly Editing.....	34
Pierazzo, Elena.....	34
Reference and Annotation. From Citation to "Watson"	36
Prätor, Klaus	36
Logging the Abbot: Reflection-Oriented XSLT Programming for Corpora Conversion and Verification	38
Pytlik Zillig, Brian L.....	38
Realistic targets in TEI to RDF.....	40
Rahtz, Sebastian	40



A web-based application for rapid annotation of TEI documents	41
Ritter, Jörg; Andert, Martin; Molitor, Paul	41
The Descartes Corpus (ProDescartes, ANR 2009-2013) Presentation	42
Roger, Julia	42
The Critical Step in Open Content Greek: Towards a Digital Edition of Athenaeus.....	43
Romanello, Matteo; Berra, Aurélien	43
TEI and DARIAH: Current Activities and Future Work	47
Schöch, Christof; Volkmann, Armin.....	47
Micropapers	49
Digital Editions as the Myth of Sisyphus	49
Burghart, Marjorie.....	49
Converting legacy editions to TEI	50
Cramme, Stefan.....	50
Application of TEI to a biographical dictionary (www.Deutsche-Biographie.de).....	50
Reinert, Matthias.....	50
Mapping metadata of TEI-encoded biographies to CIDOC-CRM	51
Reinert, Matthias; Riechert, Thomas	51
Virtual Scriptorium St. Matthias.....	52
Vanscheidt, Philipp; Scholzen, Sabine	52
Beyond TEI: Returning the text to the reader	52
Wittern, Christian	52
Poster and Demos	55
‘amalia : an eSciDoc based solution to manage the production, processing and publishing workflows of our TEI data.	55
Beaugiraud, Valérie; Gedzelman, Séverine; Ingarao, Maud; Magué, Jean-Philippe; Saïdi, Samantha.....	55
It's all about Integration and Conceptual Change.....	55
Burr, Elisabeth; Kovacs, Pascal; Potapenko, Elena	55
Wandering Jew's Chronicle	57
Cummings, James	57
Patent policies in Dingler's »Polytechnisches Journal« - Exemplary tagging of names, dates and places.....	58
Hug, Marius; Gödel, Martina.....	58



Editing Opera: Challenges of an Integrated Digital Presentation of Music and Text based on “Edirom” and TEI (OPERA – Spektrum des europäischen Musiktheaters, Universität Bayreuth / Edirom – Digitale Musikedition, Universität Paderborn)	59
Münzmay, Andreas; Daniel, Röwenstrunk; Droese, Janine; Seuffert, Janette.....	59
TXSTEP - an integrated XML-based scheme for scholarly text data processing	60
Ott, Wilhelm; Ott, Tobias	60
Virtual Scriptorium St. Matthias.....	61
Vanscheidt, Philipp; Scholzen, Sabine	61
Quality Assurance of Large TEI Corpora.....	62
Wiegand, Frank; Geyken, Alexander	62
Best Practices for TEI in Libraries	62
Kevin Hawkins, Michelle Dalmau, Melanie Schlosser	62
The project "Berlin intellectuals 1800-1830" between research and teaching	62
Baillot, Anne; Seifert, Sabine	62
William Godwin's Diary	64
Cummings, James	64
Workshops and Tutorials.....	65
Analysing Electronic Dictionaries with TEI	65
Dietmar Seipel	65
Tightening the representation of lexical data, a TEI perspective.....	65
Laurent Romary	65
Combining Music Notation and Text – Encoding and Rendering MEI in TEI.....	66
Johannes Kepper	66
Using the Scalable Architecture for Digital Editions (SADE) for the digital presentation of TEI encoded texts.....	66
Alexander Czmil.....	66
Tuning oXygen XML Editor for TEI.....	67
George Bina	67
Using TILE to build links between text and images in TEI projects	67
Dot Porter	67
An introduction to working with the TextGridLab	67
Oliver Schmid, Celia Krause and Philipp Vanscheid	67
Preparing a Critical Edition of an Incunabula with TEI	68
Guenther Goerz, Josef Schneeberger, Klaus Thoden.....	68



At the beginning, these two perspectives were intended to be combined in one complex encoding procedure within a single data file. But very soon, the fact had to be acknowledged that a combination of two widely differing perspectives leads to serious problems, especially the problem of overlapping hierarchies and/or the problem of being continuously obliged to use elements that are mutually exclusive. A special case in point is the rearrangement of segments by changing their respective positions on the manuscript.

Only the most radical solution proved to be manageable, the separation of the two levels (inscription and text) by using two separate data files. Now, both markups – the one for rendering the inscriptional record and the other one for encoding the text – can be applied without any conflict. However, the separation of the two levels on the other side does not imply that both are meant to be completely independent from each other. On the contrary, the interrelationship between both levels is of great importance for the genetic analysis.

This way of transcribing is exactly the TEI conformant model of multiple encodings of the same information described in the TEI guidelines (see chapter 20.1: www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html). As every method it has its advantages and disadvantages. The advantages, especially for dealing with Faust manuscripts, will be explained in our presentation. The splitting of the encoding rules in two different bodies of rules and concomitantly the division of most of the markup into two different types of markup shall be summarized. It will be illustrated with examples that will give some more insight into the asserted necessity of distinguishing the two perspectives.

And finally there are many questions of how to deal with the obvious disadvantages that come along with the division of transcripts, that is to say the “the maintenance of multiple copies of identical textual content” as well as the the missing “explicit indication that the various views, which might be in separate files, are related to each other” (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html#NHME>). What are the practical consequences of separating the diplomatic from the textual transcript for further steps in generating the edition and how will it be possible to avoid inconsistency? How are we going to evaluate and how to relate all information distributed on both levels? What do we have to keep in mind for the implementation of the genetic reconstruction?

Bibliography:

- <https://faustedition.uni-wuerzburg.de>
- <http://www.tei-c.org/Activities/Council/Working/tcw19.html>
- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html>

Improving the Usability of Corpus Markup and Analysis Tools by Studying their Presentation Layer

Burghardt, Manuel; Fuchs, Markus; Wolff, Christian

The TEI plays an outstanding role as a first approach towards a standardized representation of annotations. While the standardization of the representation (encoding) of annotations has steadily evolved over the last years, the presentation or visualization of markup and its



implications for the usability of markup tools has been treated with significantly less attention. In many cases the representation of markup and its actual presentation to the user are but the same thing: plain text (original text) and markup tags (annotation) at the code level. Peter Flynn has observed that “markup experts” have a different idea about the structure of a text than “conventional writers”, the one group seeing a document as a hierarchical tree with different kinds of nodes, and the other group seeing a text as a “continuous linear narrative, broken into successive divisions” (Flynn 2009). This divide between markup experts and plain annotators is of particular importance for the case of the TEI, which was designed as a representation standard for the humanities, social sciences and linguistics. Due to the research tradition and prevailing methods in this field, humanists often lack deep technical skills, i.e. many of them aren’t aware of basic markup concepts such as document types or document trees. At the same time, it has become clear that tool and ICT usage is as indispensable for the humanities as for any other field of research (Toms & O’Brien 2008). Santos & Frankenberg-Garcia claim that “most existing corpora today are only available to and understood by a small, restricted community of users” (Santos & Frankenberg-Garcia 2007). This makes usability and user experience on the presentation side a vital component for markup- and analysis-tools. Consequently, several tools try to hide the actual representation of markup from the user, by providing different interface designs and visualizations of the data. Unfortunately, current approaches to these issues are often not in accordance with existing usability standards like ISO 9241-110:2006 for dialog principles or ISO 9241-151:2008 for the usability of web interfaces (Dipper et al. 2004, Burghardt & Wolff 2009).

We identify two major challenges for the presentation layer of markup software which should be considered by tool designers in order to enhance the acceptance and the actual usage of standardized markup like the TEI guidelines, and to prevent corpora from becoming expensive data graveyards (Soehn et al. 2008), as corpus creation and especially intellectual annotation are extremely cost- and labor-intensive tasks. These challenges are at the same time general and domain-independent requirements for tools which strive for a high level of usability and user experience. The first requirement is an adequate visualization of data and annotation, the second requirement calls for appropriate interface and interaction design for markup- and analysis-tools. These requirements affect different stages in the typical workflow for the creation and use of corpora, which we call corpus pipeline. The corpus pipeline describes all steps necessary to fulfill an information need by querying an annotated corpus, starting from the creation and annotation of the actual corpus and ending with the query building and visualization of results. As presentation and representation often can’t be separated precisely during the first two stages (“digitization” and “normalization”) the presentation of markup is mainly an issue in the succeeding stages: “annotation”, “query building” and “visualization of results”. In the paper, we will derive and explain specific requirements for the presentation and interaction layer of each of these three stages (e.g. visualization of original text and multiple layers of annotation as well as the underlying annotation scheme during the annotation stage) by looking at existing tool solutions and by comparing different user interface design models with each other (recent examples of visually enriched tools including e.g. WordTree (Wattenberg & Viegas 2008) and DoubleTree (Culy & Lyding 2010) in the stage of “visualization of concordances”).



We argue for a user-centered presentation of markup, starting from the annotation of text, and ending with the querying of a corpus of documents and the presentation of query results. Future work will include a detailed user study and evaluation of different presentation aspects, such as the best presentation of multilayer annotation or complex queries.

Bibliography:

- Fuchs, Markus (2010). Aufbau eines linguistischen Korpus aus den Daten der englischen Wikipedia. 2010. In: Pinkal, M. & Rehbein, I. & Schulte im Walde, S. & Storrer, A. (Hrsg.).
- Semantic Approaches in Natural Language Processing. Proceedings of the Conference on Natural Language Processing 2010 (KONVENS 10). Saarbrücken: Universitätsverlag des Saarlandes. S. 135-139. Online verfügbar unter: <<http://wikicorpus.com/ErstellungWPKorpus.pdf>>
- Burghardt, Manuel & Christian Wolff (2009). Werkzeuge zur Annotation diachroner Korpora. In: Hoepfner, Wolfgang (Hrsg.).
- Proc. GSCL-Symposium Sprachtechnologie und eHumanities. Technische Berichte der Abteilung für Informatik und Angewandte Kognitionswissenschaft, 2009-01. Abteilung für Informatik und Angewandte Kognitionswissenschaft, Universität Duisburg-Essen, Duisburg, S. 21-31. Online verfügbar unter: <<http://epub.uni-regensburg.de/6756/>>
- Burghardt, Manuel & Christian Wolff (2009). Stand off-Annotation für Textdokumente: Vom Konzept zur Implementierung (zur Standardisierung?). In: Chiarcos, Christian et al. (Hrsg.).
- Von der Form zur Bedeutung: Texte automatisch verarbeiten. Proceedings of the Biennial GSCL-Conference 2009 in Potsdam, S. 53-59. Online verfügbar unter: <epub.uni-regensburg.de/14223/>

Glossing music theory: how to make transparent the web of quotations, authorities and allusions in medieval texts

Desmond, Karen

This paper takes as its starting point a music theory text known as the **Ars nova**. This text has been considered foundational to our understanding of medieval music history. In the fourteenth century there was a profound shift in musical style from the previous century's **ars antiqua** (the "old art") to what was termed the **ars nova** (the "new art"). The **Ars nova** was the medieval "avant-garde" with a sound that combined new rhythms, harmonies and texts in complex structural and formal layers. This complexity was due in large part to the expansion and reformulation of the musical notation system. The **Ars nova** theory treatise was a short technical manual that contained rubrics on how to interpret this new notation system. In the traditional historical narrative, the supposed author of this treatise was the composer and poet, Philippe de Vitry (1291-1361), who wrote music in the new style, and was quickly crowned through the annals of music history as the figurehead and putative creator of the *ars nova* movement.

This narrative is extremely simplified. There is no "one" complete text of the **Ars nova**, but in fact, a small handful of related, but widely divergent, texts extant in manuscripts dating from the fourteenth and fifteenth centuries. The sixteen or so texts that present these new notational theories vary in many ways: in levels of completeness (many of the texts start or