

Alois Pichler, Wittgenstein Archives at the University of Bergen (WAB)

Is XML suitable for the transcription and editing of complex manuscript materials? A philosophical solution to an ongoing debate about text encoding

Summary

There is a philosophical, or more precisely, an ontological question about text. It is, in short, the question: What *is* text? In Digital Humanities, the issue has been focused around the more specific question whether texts are hierarchical or rather non-hierarchical structures. Examples of this discussion include the statement that “text is best represented as an ordered hierarchy of content object (OHCO), because that is what text really is” (DeRose, Durand, Mylonas, Renear 1990), or, in opposition to it, the statement that “humanists are trying to represent what they all agree are non-hierarchical structures” (Schmidt 2010). Conflicting conclusions for text encoding have been drawn from these two opposed approaches to the question of what text is. Some have concluded that the hierarchical markup grammar XML can be regarded as adequate for text encoding, because that is, so their view, what texts basically are: hierarchies. Others reject XML and embedded markup more generally as inadequate, precisely on the basis of the view that texts have or can have non-hierarchical structures. I shall look at both of these conclusions, one of my points being that they have both been drawn unjustly, due to a mistaken, “objectifying” ontology of text. In contrast to this view, I shall, starting from an analysis of the differences between writing acts (producing *documents*) and text acts (producing *texts*), try to show that texts ontologically are best conceived of as:

- 1) events (rather than objects or properties);
- 2) specific kinds of events, namely actions;
- 3) specific kinds of actions that are
 - a) jointly produced by both authors and readers (rather than by authors alone)
 - b) symbolizing or semiotic in nature (i.e. actions that are meaning and structure constituting and thus also normative rather than merely meaning and structure depicting).

The application of these three points, together, to the debate about hierarchical vs. non-hierarchical markup shows that both positions are mistaken. “Hierarchical” or “non-hierarchical” describe aspects of our encoding actions rather than aspects of texts “for and in themselves”: i.e., as entities that would exist independently of the encoders’ text actions. As a consequence, the question of hierarchical vs. non-hierarchical markup reveals itself to be a question under control, namely *our* control, rather than an unsolved ontological issue. This helps explain why XML is such a successful markup system for the encoding of complex manuscript materials as well.

Structure of my lecture

1. Writing (Homo litteratus)
2. Scholarship (Humanities)
3. Digital scholarship (Digital Humanities)
4. Philosophy (Philosophy in Digital Humanities)
 - Hierarchical vs. non-hierarchical representation
 - Writing vs. “texting”
 - Sign vs. symbol
5. The normative grammar of “text” (Homo litteratus)

3. Digital scholarship (Digital Humanities)

Select a Nachlass corpus:

Ms-114 (WS) ▼

Sort corpus according to date (optional):

No sorting ▼

Choose basic style for:	<input checked="" type="radio"/> Diplomatic	<input type="radio"/> Normalized
Additions	<input checked="" type="radio"/> Diplomatic	<input type="radio"/> Normalized
Deletions (where deleted part fits in cotext)	<input checked="" type="radio"/> Diplomatic	<input type="radio"/> Normalized
Deletions (where deleted part does not fit in cotext)	<input checked="" type="radio"/> Diplomatic	<input type="radio"/> Normalized
Overwritings	<input checked="" type="radio"/> Diplomatic	<input type="radio"/> Normalized
Editorial corrections	<input checked="" type="radio"/> Diplomatic	<input type="radio"/> Normalized

Include/exclude line breaks	<input type="radio"/> Include	<input checked="" type="radio"/> Exclude
-----------------------------	-------------------------------	--

Show/hide revisions in different writing medium (e.g. pencil in typescript)	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
---	---------------------------------------	----------------------------

Highlight text types	<input type="radio"/> Highlight	<input checked="" type="radio"/> No highlight
Code	<input type="radio"/> Highlight	<input checked="" type="radio"/> No highlight
Notation	<input type="radio"/> Highlight	<input checked="" type="radio"/> No highlight

Show/hide 'section marks'	<input type="radio"/> Show	<input checked="" type="radio"/> Hide
---------------------------	----------------------------	---------------------------------------

Filter 'Bemerkungen' according to 'section mark'	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
[No section mark]	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
/	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
\	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
x	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
—	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
+	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
†	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
*	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
?	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
/?	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
\?	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
✓	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
//	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
\\	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
∴	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
o	<input checked="" type="radio"/> Show	<input type="radio"/> Hide
⊙	<input checked="" type="radio"/> Show	<input type="radio"/> Hide

“Interactive Dynamic Presentation” offered by the Wittgenstein Archives at the University of Bergen for user-steered editing of Wittgenstein’s Nachlass (screenshot of part of <http://wab.uib.no/transform/wab.php?modus=opsjoner>)

4. Philosophy (Philosophy in Digital Humanities)

Hierarchical vs. non-hierarchical markup

The authors argue that text is best represented as an ordered hierarchy of content object (OHCO), because that is what text really is.

DeRose, Durand, Mylonas, Renear 1990: What Is Text, Really? In: Journal of Computing in Higher Education 1(2), 3-26; p.3

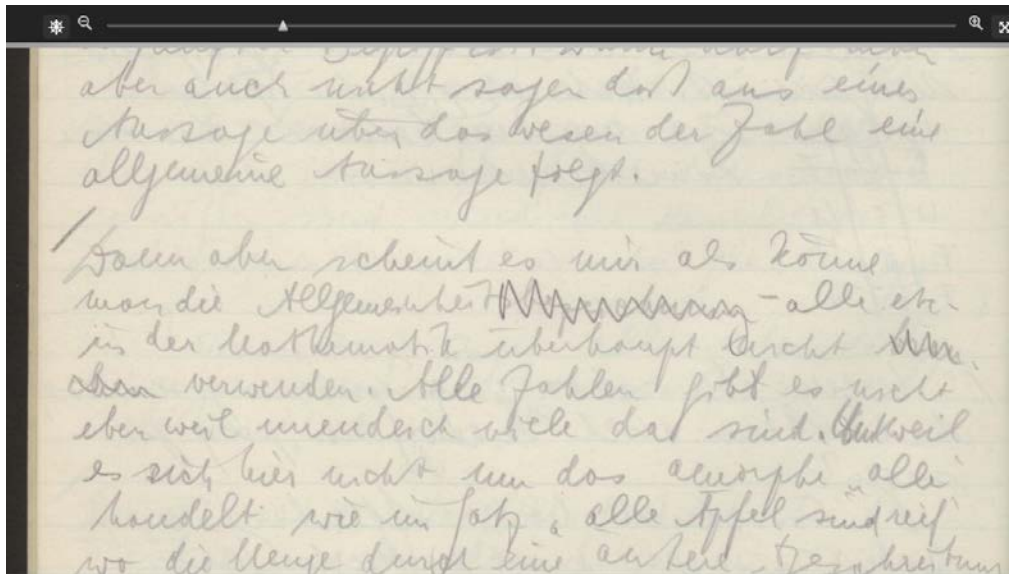
The origin of the overlap problem is simply that humanists are trying to represent what they all agree are non-hierarchical structures using a container whose primary structure is a tree.

Schmidt 2010: The inadequacy of embedded markup for cultural heritage texts. In: Literary & Linguistic Computing 25(3), 337-356; p.344ff

Text ontology (an attempt)

Events (more precisely: <i>actions</i>) rather than objects or properties; caused by authors <i>and</i> readers rather than authors alone; accidents rather than substances; immaterial rather than material; infinite and open-endedly ongoing rather than finite; mind-dependent rather than mind-independent; ...	Texts	Symbol level
Material objects (e.g. sheets of paper) as well as material events (e.g. speech)	Document (NB: Is only <i>one</i> of the text carriers!)	Sign level
	Document carrier	Physical level

Document vs. text



Wittgenstein Nachlass Ms-106,90 (http://wittgensteinsource.org/BFE/Ms-106,90_f).
With the kind permission of © 2015 The Master and Fellows of Trinity College, Cambridge;
The Österreichische Nationalbibliothek, Vienna; The University of Bergen, Bergen

Source transcription

```
<s type="es">Dann aber scheint es mir als könne man die  
<choice type="em"><orig type="em1">Allgemeinheit<del type="d">sbezeichnung</del></orig>  
<orig type="em2"><choice type="dsl"><orig type="alt1">Allgemeinheitsbezeichnung</orig> <orig  
type="alt2">Allgemeinheit</orig></choice></orig></choice> &dash; alle  
<abbr type="abb">etc<corr type="tra">&p.abb;</corr></abbr> &dash; in der Mathematik  
&uuml;berhaupt nicht  
<choice type="dsl"><orig type="alt1"><del type="d">brau<lb rend="shyphen"/>chen</del></orig>  
<orig type="alt2">verwenden</orig></choice>&p.es;</s>
```

Diplomatic transcription

Dann aber scheint es mir als könne
man die Allgemeinheitsbezeichnung — alle etc —
in der Mathematik überhaupt nicht brau-
chen verwenden.

Normalized transcription

Dann aber scheint es mir als könne man die Allgemeinheitsbezeichnung Allgemeinheit — alle
etc. — in der Mathematik überhaupt nicht brauchen verwenden.

Overlap

- The trees are green with white flowers.
- The <underline>trees <italic>are green</italic> with white</italic> flowers.
- The <underline>trees <italic part="I">are green</italic></underline><italic part="F">with white</italic> flowers.

5. The normative grammar of “text” (Homo litteratus)

1. *Augustinus*, in den *Confessionen* I/8: cum ipsi (maiores homines) appellabant rem aliquam, et cum secundum eam vocem corpus ad aliquid movebant, videbam, et tenebam hoc ab eis vocari rem illam, quod sonabant, cum eam vellent ostendere. Hoc autem eos velle ex motu corporis aperiatur: tamquam verbis naturalibus omnium gentium, quae fiunt vultu et nutu oculorum, ceterorumque membrorum actu, et sonitu vocis indicante affectionem animi in petendis, habendis, rejiciendis, fugiendisve rebus. Ita verba in variis sententiis locis suis posita, et crebro audita, quarum rerum signa essent, paulatim colligebam, measque jam voluntates, edomito in eis signis ore, per haec enuntiabam.¹

In diesen Worten erhalten wir, so scheint es mir, ein bestimmtes Bild von dem Wesen der menschlichen Sprache. Nämlich dieses: Die Wörter der Sprache benennen Gegenstände — Sätze sind Verbindungen von solchen Benennungen. — In diesem Bild von der Sprache finden wir die Wurzeln der Idee: Jedes Wort hat eine Bedeutung. Diese Bedeutung ist dem Wort zugeordnet. Sie ist der Gegenstand, für welchen das Wort steht.

Von einem Unterschied der Wortarten spricht Augustinus nicht. Wer das Lernen der Sprache so beschreibt, denkt, so möchte ich glauben, zunächst an Hauptwörter, wie “Tisch”, “Stuhl”, “Brot”, und die Namen von Personen, erst in zweiter Linie an die Namen gewisser Tätigkeiten und Eigenschaften, und an die übrigen Wortarten als etwas, was sich finden wird. ...

¹ Nannten die Erwachsenen irgend einen Gegenstand und wandten sie sich dabei ihm zu, so nahm ich das wahr und ich begriff, daß der Gegenstand durch die Laute, die sie aussprachen, bezeichnet wurde, da sie auf *ihn* hinweisen wollten. Dies aber entnahm ich aus ihren Gebärden, der natürlichen Sprache aller Völker, der Sprache, die durch Mienen- und Augenspiel, durch die Bewegungen der Glieder und den Klang der Stimme die Empfindungen der Seele anzeigt, wenn diese irgend etwas begehrt, oder festhält, oder zurückweist, oder flieht. So lernte ich nach und nach verstehen, welche Dinge die Wörter bezeichneten, die ich wieder und wieder, an ihren bestimmten Stellen in verschiedenen Sätzen, aussprechen hörte. Und ich brachte, als nun mein Mund sich an diese Zeichen gewöhnt hatte, durch sie meine Wünsche zum Ausdruck.

Wittgenstein, Ludwig: Philosophical Investigations / Philosophische Untersuchungen. Ed. by P. M. S. Hacker and Joachim Schulte, translated by G. E. M. Anscombe, P. M. S. Hacker and Joachim Schulte. New York: Wiley. p.5

Some twenty quotes for discussion

[1] St.J. DeRose, D.G. Durand, E. Mylonas, A. Renear 1990: What Is Text, Really?

Journal of Computing in Higher Education 1(2), 3-26; p.3.

<http://link.springer.com/article/10.1007%2FBF02941632>

The way in which text is represented on a computer affects the kinds of uses to which it can be put by its creator and by subsequent users. The electronic document model currently in use is impoverished and restrictive. The authors argue that text is best represented as an ordered hierarchy of content object (OHCO), because that is what text really is. This model conforms with emerging standards such as SGML and contains within it advantages for the writer, publisher, and researcher. The authors then describe how the hierarchical model can allow future use and reuse of the document as a database, hypertext, or network.

[2] P.L. Shillingsburg 1991: Text as Matter, Concept, and Action

Studies in Bibliography 44, 31-82; p.81f. <http://www.jstor.org/stable/4037193>

Texts as Action:

Speech Act = The whole event of creation, production, and reception of a communication at a specified time and place.

Write Act = The complex, never closed, serial event encompassing the creations, productions, and receptions at any and all places and times in which a written work is created, produced and received.

Utterance = A whole Speech Act or a coherent selection of „speaker,“ medium, hearer, time(s), and place(s) employed with regard to a Write Act within which „an understanding“ of the Work is achieved.

Creative Performance = The authorial development of Essayed Versions resulting in Linguistic Texts as found in manuscripts and authorially revised texts.

Production Performance = The scribal and publication development of Material Texts resulting in typists' copies, proofs, and printed books.

Reception Performance = The development of a conceptualized Reception Text in the act of reading.

[3] C. Huitfeldt 1992: Multi-Dimensional Texts in a One-Dimensional Medium

Wittgenstein and Contemporary Theories of Language (Eds. P. Henry, A. Utaker. Bergen: Wittgenstein Archives), 142-161; p.144ff.

<http://wittgensteinrepository.org/agora-wab/article/view/2940/3601>

Why I call texts multi-dimensional will hopefully become clear from the discussion further below ... Even if SGML allows me to have several concurrent hierarchies in a text, I am not convinced that Wittgenstein's manuscripts are basically hierarchical structures. Potentially, for all I know, any feature may overlap with any other feature. Besides, I do not even know what the hierarchies should consist of, or whether the identification of such hierarchies would be particularly illuminating. ... Why on earth should texts by all means be hierarchies? No doubt, there are many hierarchical structures, and no doubt this is important, but there are countless other relations between text elements which are worth while finding and investigating – overlap, substitution, discontinuity, parallel texts, cross-references, etc. ...

[4] A. Renear, E. Mylonas, D.G. Durand 1992: Refining our notion of what text really is: The problem of overlapping hierarchies

Annual joint meeting of the Association for Computers and the Humanities (ACH) and the Association for Literary and Linguistic Computing (ALLC), Oxford University, Oxford.

<http://cds.library.brown.edu/resources/stg/monographs/ohco.html>

We examine the claim that 'text is an ordered hierarchy of content objects'; this thesis was affirmed by the authors, and others, in the late 1980s and has been associated with certain approaches to text processing and the encoding of literary texts.

[5] A. Pichler 1995: Transcriptions, Texts and Interpretation

Beiträge des 18. Internationalen Wittgenstein Symposiums, Kirchberg am Wechsel, 690-695; p.690f. <http://bora.uib.no/handle/1956/1874>

I think, the essential question is not about a true representation, but: Whom do we want to serve with our transcriptions? Philosophers? Grammarians? Or graphologists? What is "correct" will depend on the answer to this question. And what we are actually going to represent, and how, is determined by our research interests (philosophical, grammatical, philological, graphological ... interests), and not by a text, which exists independently and which we are going to depict.

[6] A. Pichler 1995b: Advantages of a Machine-Readable Version of Wittgenstein's Nachlaß

Beiträge des 18. Internationalen Wittgenstein Symposiums, Kirchberg am Wechsel, 770-776; p.774f. <http://bora.uib.no/handle/1956/1875>

Machine-readable versions make it more clear to us what texts are and what text editing means: Texts are not objectively existing entities which just need to be discovered and presented, but entities which have to be constructed. They are products of both the author and the reader. All that exists in the case of Wittgenstein's Nachlass are scripts which first of all need to be identified, interpreted and organized. Having a machine-readable version of Wittgenstein's Nachlass provides a multiplicity of ways to organize and construct texts, it makes this easy - and it makes it obvious that there is an element of construction.

[7] M. Biggs, C. Huitfeldt et al. 1997: Philosophy and electronic publishing. The theory and metatheory in the development of text encoding

The Monist. Interactive Issue 80(3), 348-367; p.358ff.

https://www.jstor.org/stable/27903534?seq=1#page_scan_tab_contents

Raymond concentrates on the structural approaches epitomised in Renear's OHCO discussion. Raymond rejects the efficiency arguments which objectify OHCOs in preference to other analytical constructs. In particular, he finds the transfer of computer terminology such as "text file" insidious. The implication that the text file "contains" text is as misleading as the suggestion that a payroll database "contains" the payroll. In general he [Raymond] draws the distinction between manipulating texts successfully using frameworks such as OHCOs, and the lack of any necessity to also provide a coherent theory of text. For example, advocates of OHCO seem to suggest that the structure of a document is a property of the document that is independent of the operations to be performed on it, or of other issues such as how we decide the equivalence of documents. He suggests that to "structure" information means to encode it in such a way that certain operations are efficient and others are not. By the time one has

structure, one is already halfway to operations. ... Raymond's external argument criticises advocates of OHCO for suggesting that OHCO-like structures should capture only structure and not semantics (e.g., SGML). He argues that structure always involves semantics, and there is no such thing as "pure" structure, because structure always has mathematical and combinatorial properties that make certain types of operation possible or efficient. Hence structures are chosen which support the operations we think we want to do, which in turn reflect the semantics that we implicitly attach to the text. ... Broderick proposes the thought-experiment that after knowledge of English has been completely lost, an archaeologist digs up an issue of the *Monist*. Is it still a text? If the inhabitants of this age exchange knowledge in electronic format and no longer read the appropriate kind of character strings? Does a text contain knowledge if there is no one around to read it? These questions point out what seems to be the most credible Realist/Antirealist distinction. The Realist would answer "yes" to the above questions, the Antirealist, a "no," qualified only by the possibility that the future archaeologist might figure out some way to decode the artifact and return its textuality to it. ...

[8] S. Hockey, A. Renear, J.J. McGann 1999: Panel: What is text? A debate on the philosophical and epistemological nature of text in the light of humanities computing research

Annual joint meeting of the Association for Computers and the Humanities (ACH) and the Association for Literary and Linguistic Computing (ALLC), University of Virginia, Charlottesville, June 1999. <http://www2.iath.virginia.edu/ach-allc.99/proceedings/hockey-renear2.html>

The Text Encoding Initiative and other recent markup research projects in humanities computing have fostered lively debate and discussion on the nature of text. If we are to represent text in a form suitable for electronic processing, we need to define what text is and what structures we want to be able to recognize in text. ...

1. Position statement from Renear

I will present and defend a particular view of textuality. It is a view which has roots in common sense and tradition, and which has also proven to be heuristic in directing research and practical for guiding system and software designers. It is a view that seems to illuminate textual practices in the humanities and it is arguably implicit in work of the Text Encoding Initiative. However it is, at least in parts, coming increasingly under suspicion. This is too bad, as this view is not only a rather good account of textuality, it is the best account we have. I will advocate this theory by defending five theses; I will argue that texts are:

1. real: they have properties independent of our interests in them and our theories about them.
2. abstract: the objects which constitute texts are abstract, not material, objects.
3. intentional: texts are, necessarily, the product of mental acts
4. hierarchical: the structure of texts is fundamentally hierarchical
5. linguistic: texts are linguistic objects; renditional features are not parts of texts, and therefore not proper locations for textual meaning.

2. Position statement from McGann

The question framing this ACH session involves a misconception. It assumes that "text" is a unitary phenomenon and that its concept can be thought as self-identical. But while both of these assumptions may be undertaken for heuristic purposes, neither represents what Wittgenstein called "the case". ... There is no question but that most of our textual archive is hierarchically organized. On the other hand, there is also no question but that poetical texts comprise a key, perhaps even a defining, part of the corpus of our humanities archive. ... Unlike expository text, poetry is not organized in a determinate hierarchy. TEI and SGML markup, therefore, while reasonably adequate vehicles for expository and informational texts, fails to render those features of poetic text that are most salient for its makers and users. Poetical texts are recursive structures built out of complex networks of repetition and variation. No poem can exist without systems of "overlapping structures", and the more developed the poetical text, the more complex are those systems of recursion. ... This essential character of poetical text helps to explain why content in poeisis tends to involve more broadly "semiotic" rather than narrowly "linguistic" materials. The sonic and visible features of text are, so far as the poets who make these texts are concerned (or the readers who engage them), nearly as apt for expressive poetical purposes as the semantic, syntactic, and rhetorical features. Each of these features represents a field of textual action, and while any one field may be individually (abstractly) framed in a hierarchized scheme, the recursive interplay of the fields produces works whose order is not hierarchical. Of course a governing hierarchy can be imposed upon such works. TEI and SGML create, as Renear shows, a certain type of "linguistic" hierarchy, one that privileges text as a container for storing information. But even that linguistic hierarchy is highly specialized (it does not consider, for example, the rhetorical structures that overlap and infect the syntax and semantics). ... Not without reason has the Bergen Wittgenstein project abandoned TEI/SGML as a system for marking up the corpus of Wittgenstein's texts; and the scholars setting out now to "edit" the Peirce archive are well aware that TEI/SGML does not lend itself to an adequate treatment of Peirce's work, and least of all to his existential graphs. "Text" in Kant "is" one thing, but in Peirce it "is" something else again.

[9] D. Buzzetti 2002: Digital Representation and the Text Model

New Literary History 33(1), 61-88; p.71f. <http://muse.jhu.edu/article/24607>

An "OHCO structure" is not a model of the text, but a possible model of its expression. The "OHCO thesis," however, is no longer accepted without reservation. In fact, the act of encoding texts itself has given rise to "some practical problems that seem to call this thesis to question" (RN 263). In essence, "the way in which texts were analyzed into objects by the text processing theorists and standard developers," a method founded upon the notion that the "text" is a string of characters, has revealed itself to be "fundamentally different from the way in which they were analyzed into objects by the literary and linguistic encoding community." So, "the new literary history tendency of SGML to assume that documents could be represented as a single hierarchical structure created real practical problems for text encoding projects." But it is not only a matter of concluding that the OHCO thesis "is false" because you can assign "many hierarchical structures," all of which have "a plausible claim to be the 'logical' structure" (RN 269), to a single document, but rather, it is a matter of admitting that "the relevant logical structures of a text are not without exception hierarchical" (RN 279, n.13). Indeed, not only must we agree that "the same document conforms to several overlapping structures," each of which is "strict hierarchical," ... but we must also recognize that textual structures are not usually of this type.

[10] P. Eggert 2007: The Conservator's Gaze and the Nature of the Work
Library Trends 56(1), 80-106; p.84. <https://muse.jhu.edu/article/223252>

"Our aim in transcription," Pichler reasoned, "is not to represent as correctly as possible the originals, but rather to prepare from the original text another text so as to serve as accurately as possible certain interests in the text." And he added: "what we are going to represent, and how, is determined by our research interests . . . and not by a text which exists independently and which we are going to depict" (1995, p. 690). In 1997 Allen Renear, a prominent member of the same community, objected to what he called this antirealist view of text, but his arguments seem finally to rest on the unproblematized notion that texts are or must be objective realities that encoders would do well to represent truthfully (Renear, 1997, pp. 117–124).

[11] P. Robinson 2009: What text really is not, and why editors have to learn to swim
Literary & Linguistic Computing 24(1), 41-52; p.46f.
<http://llc.oxfordjournals.org/content/24/1/41.full.pdf+html>

Yet, while I am clearly more comfortable with Pichler's formulation than with those of Sperberg-McQueen and Renear, there are elements in Pichler's analysis with which I am uncomfortable, too, and where I find myself closer to Renear than to Pichler. In the last paragraph, I argue that our function is to help the text speak to its readers; that we perceive the text as saying something, and that we seek to help readers discover what this something is. This takes us away from what one might term Pichler's hard-line 'antirealism' to something nearer Renear's argument: the text does have an independent existence; it is saying something; we are not just 'constructing' an artefact for the use of our readers, but we are trying to interpret an utterance.

[12] M. Nassourou 2010: Markup Overlap: Improving Fragmentation Method
<https://opus.uni-wuerzburg.de/frontdoor/index/index/docId/4091>

Overlapping is a common word used to describe documents whose structural dimensions cannot be adequately represented using tree structure. Non-hierarchical structures allow some children to share simultaneously several parents. . . . The Non-XML approaches introduce new syntaxes and semantics that need to be learned and understood. Not being XML constitutes a major disadvantage for these solutions. . . . XML is an easy, human readable language that has tools available for processing it. XML users would definitely prefer solutions within XML itself than new syntaxes.

[13] D. Schmidt 2010: The inadequacy of embedded markup for cultural heritage texts
Literary & Linguistic Computing 25(3), 337-356; p.344ff.
<http://llc.oxfordjournals.org/content/25/3/337.full.pdf+html>

The origin of the overlap problem is simply that humanists are trying to represent what they all agree are non-hierarchical structures using a container whose primary structure is a tree. This seems to apply to all markup languages that are embedded in the text, not only those based on XML. . . . Humanists seem united in their view that all markup is interpretation (e.g. Eggert, 2005; Fiormonte, 2003, p. 163; McCarty, p. 27; Sperberg-McQueen, 1991, p. 35). If that is so, then why is it embedded in the text? The text is the thing being interpreted, not the markup. Once embedded, markup obscures and biases what a new scholar, who didn't carry

out the initial markup, can see. ... So the big question becomes, is it appropriate to embed the technology and interpretations of today into texts that will be archived for the future? If not, then archives of cultural heritage texts should be in a plain text format. The ASCII standard, which became part of ISO-LATIN-1, which in turn became part of Unicode, is one of the most stable forms of digital data known. If one can predict anything about the software industry it is that texts of the future will be in Unicode or something compatible with it.

[14] D. Schmidt 2012: The Role of Markup in the Digital Humanities

Historical Social Research / Historische Sozialforschung 37(3), 125-146; p.126.

<http://www.ssoar.info/ssoar/handle/document/37836>

The question then becomes how good is the data model of the transcriptions upon which all these operations depend? Does it facilitate or inhibit the development of all that functionality? Can we share applications built upon it? Can we automate the process of creating a digital edition of a work or a set of documents? Can we make it easy to use so that training is minimised? All of these questions depend upon the data model we choose to represent the text, which until now has been embedded markup. ... Most digital humanists have heard of the “overlapping hierarchies” problem made famous by the 1993 paper by Renear, Mylonas and Durand. What was noticed very early on in the application of generalised markup languages to humanities texts was their apparent inability to represent anything except tree-structures; whereas what humanists often wanted to record seemed to be non-tree structures.

[15] H.W. Gabler 2012: Wider die Autorzentriertheit in der Edition

Jahrbuch des Freien Deutschen Hochstifts 2012, 316-342; p. 317

<http://epub.ub.uni-muenchen.de/17080/>

Allein auf Dokumenten schlagen sich Texte nieder. Ohne Trägern eingeschrieben zu sein - ob Stein, Ton, Papyrus, Pergament oder Papier - wären Texte als Texte niemals materiell fassbar. Folglich gehen Texte und Dokumente in unserer überkommenen Schreib- und Schrifttradition eine anscheinend untrennbare Symbiose ein, die so weitreichend ist, dass im alltäglichen Sprachgebrauch, ja selbst konzeptionell, beide Begriffe gegeneinander austauschbar sind. Verträge ebenso wie Testamente beispielsweise sind in Sprache als Texte formuliert. Dennoch wird ihre Rechtsgültigkeit und -Verbindlichkeit üblicherweise damit begründet, dass wir sie als Rechtsdokumente (unterschrieben, bezeugt und gesiegelt) besitzen und vorweisen können. Das materielle Dokument als solches als >Vertrag< oder >Testament< zu bezeichnen, ist eine alltagssprachliche Verkürzung. Analytisch betrachtet sind Text und Dokument distinkte, logisch trennbare Entitäten.

[16] C. Huitfeldt, F. Vitali, S. Peroni 2012: Documents as Timed Abstract Objects

Proceedings of Balisage: The Markup Conference 2012. Balisage Series on Markup Technologies 8. <http://www.balisage.net/Proceedings/vol8/html/Huitfeldt01/BalisageVol8-Huitfeldt01.html>

We believe that this account of documents as Timed Abstract Objects has a number of strengths, and that it gives a coherent explanation of document change as well as our practices of identifying documents at different levels of abstraction. Notwithstanding its attractiveness, however, it also has some weak points. One of its consequences is that two documents which would perhaps on other accounts be regarded as identical would have to be regarded as different if they have different histories. In other words, two series of identical changes to "the

same" document at slightly different times will constitute two different documents. But perhaps this is just a minor quirk, perhaps we simply don't have clear intuitions in such cases. However, the fact that its change history becomes so to speak a constitutive part of the document itself also implies that documents are not really accounted for as changeable objects, but as "events". Events, though they do take time, don't change. This attempt to account for document change has the seemingly paradoxical consequence that documents do not really change at all. (A poker may exemplify change by being first cold, and then warm. But a poker which is cold at the one end and warm at the other does not exemplify change.)

[17] P. Sahle 2013: Digitale Editionsformen (Schriften des Instituts für Dokumentologie und Editorik 7-9)

Teil III: Textbegriffe und Recodierung; p. 409. <http://kups.ub.uni-koeln.de/5353/>

Selbst ein vorwissenschaftliches Textverständnis integriert bereits drei konstitutive Hauptaspekte des Textes: (1.) Gedanken werden (2.) sprachlich formuliert und (3.) medial ausgeformt. Daraus lassen sich drei fundamentale Sichten auf den Text ableiten: Text als Inhalt (Idee, Gedanke, Intention), Text als sprachliche Äußerung und Text als Dokument (mediale Ausformung). Wenn man hier nun jeweils Zwischenpositionen (Verbindungspositionen) berücksichtigt, dann lassen sich insgesamt sechs Textbegriffe klassifizieren, die durch Indizes gekennzeichnet werden sollen. In Ergänzung zum Indexfreien „Text“ als Bezeichnung eines konkreten Textes ohne explizite Angabe des dahinter stehenden Textverständnisses und zum Text_M als dem Metabegriff des „Text als reden über den Begriff des Textes“ kämen dann ...

zunächst als Hauptperspektiven:

- Text_S – Der Text als sprachliches Gebilde, als sprachliche Äußerung.
- Text_I – Der Text als Inhalt, Idee, Intention.
- Text_D – Der Text als Dokument, als Materialisierung in einem Medium.

und dann als Zwischenpositionen:

- Text_W – Der Text als Werk. Als sprachlich variable aber bereits strukturierte Ausdrucksformen eines Gedankens.
- Text_F – Der Text als einzelne sprachliche Fassung. Als medial fixierte Äußerung.
- Text_Z – Der Text als zeichenhafter Ausdruck. Als nicht-nur-sprachliche Äußerung.

[18] P. Sahle 2013: Digitale Editionsformen (Schriften des Instituts für Dokumentologie und Editorik 7-9)

Teil III: Textbegriffe und Recodierung; p.511. <http://kups.ub.uni-koeln.de/5353/>

Texte sollten auf der Basis der OHCO-These recodiert werden, weil dies die meisten praktischen Vorteile bieten würde. Und darin sieht man dann in einer merkwürdigen logischen Schlussweise den eigentlichen Beweis für die ontologische Identität von „Text“ und OHCO-Modell: “the reason this modell of text is so functional and effective is that it reflects what text really is” oder später etwas anders ausgedrückt „the comparative efficiency and functionality of treating texts ,as if“ they were OHCOs is best explained [...] by the hypothesis that texts ‘are’ OHCOs“. Weil sich die OHCO-These in der Praxis bewährt und es angeblich keine ernsthaften Alternativmodelle gebe, sei der Beweis erbracht, dass Texte tatsächlich „geordnete Hierarchien von Inhaltsobjekten“ seien. Weil das Modell adäquat zur Wirklichkeit seiner Anwendung sei, sei es auch ontologisch zwingend für den Gegenstand seiner Anwendung! *Die OHCO-These als spezieller Textbegriff.* Wenn die Vorstellung des Textes als OHCO dadurch begründet wird, dass sie die größte *praktische* Mächtigkeit nach sich ziehen würde, dann liegt der Verdacht nahe, dass auch hier nur einmal mehr aus den

gerade aktuellen technischen Möglichkeiten der zu ihnen passende Textbegriff abgeleitet worden ist. Dass auch hier die Technologie einfach manche Vorstellungen eher begünstigt und bevorzugt realisiert und andere behindert oder ausblendet.

[19] Y. Liu 2014: Ways of Reading, Models for Text, and the Usefulness of Dead People
Scholarly and Research Communication 5(2), 1-13; p.1ff.
<http://src-online.ca/index.php/src/article/view/148/296>

The definition of text is still a live issue with important implications for theorizing, developing, recognizing, and using emerging forms of digital textuality. This article proposes that no single definition of text is sufficient to account for all manifestations of textuality and presents medieval textuality as a test case. ... But while scholars argue over what text really is, around them are millions of people creating and using texts without committing themselves to any kind of textual metaphysics and negotiating – often with ease and very often unreflectively – multiple, variable, and complex texts and textual functions. Furthermore, people have been doing this for thousands of years. ... If we ought not to insist upon a single model of text, recognizing the diversity not only of text forms but also of approaches to text and uses for text, value remains in attempting to describe possible models of text and their implications.

[20] C. Kanzian 2015: Kunstwerke als Artefakte
Revista Portuguesa de Filosofia 71(4), 895-912; p.897f.
http://www.publicacoesfacil.pt/product.php?id_product=819

Wenn man nach einem kategorialen Rahmen sucht für die Interpretation unserer alltäglichen Lebenswelt, kommt man nicht vorbei an *Partikularien* oder konkreten Individuen, das sind einmalige, unwiederholbare Entitäten: *Ereignisse* und *Zustände* sowie, als eigene Kategorie, die *Dinge*, worunter im Allgemeinen *materielle Gegenstände* verstanden werden können. Ereignisse sind, im Unterschied zu Dingen, Bewegungen bzw. Änderungen materieller Gegenstände; Zustände bestehen darin, dass einem materiellen Gegenstand Eigenschaften zukommen. Eine Autofahrt z.B. ist ein Ereignis, das Auto ein Ding, das Blausein des Autos aber ein Zustand.

Dinge zeichnet, im Unterschied zu Ereignissen und Zuständen, aus, *drei- und nicht vierdimensionale* Entitäten zu sein. Sie haben also eine räumliche, nicht aber eine zeitliche Ausdehnung. Sie sind somit zu jedem Zeitpunkt ihrer Geschichte nicht nur teilweise, sondern als Ganze da. Darin begründet sich ein weiteres grundlegendes ontologisches Merkmal, nämlich ihre *diachrone Identität* im strikten Sinn. Dinge sind „endurer“. Die Autofahrt z.B. kann man, aufgrund ihrer zeitlichen Ausdehnung, in zeitliche Teile gliedern; nicht aber das Auto selbst. So besteht zwischen Linz und St. Pölten ein zeitlicher Teil der Autofahrt von Innsbruck nach Wien, allerdings jederzeit das ganze Auto.

[21] H.W. Gabler 2015 forthc.: The Draft Manuscript as Material Foundation for Genetic Editing and Genetic Criticism
Variants 12 (forthc.). <https://textualscholarship.eu/journal/issues/>

Text is the result of a writing-for-reading and is pre-conditioned by the rules and habits of reading: it advances linearly, two-dimensionally, from upper left-hand to bottom right-hand corner of a given material support, e.g. a page or sheet, and thence through a sequence of pages. But writing in draft documents is not so vectored. The prime function of draft documents, and the writing in them, is not to record text for reading, but to record, support

and further engender composition. For the processes of composition, a writing space is not predetermined by expectations of linear text reading. What we encounter as writing in the pages of original draft documents, therefore, are the traces of how the document space was filled in the course of composition. ... At its best, the “reading” of a draft inscription amounts to a process of deciphering. This requires both a spatial comprehension and a comprehension of the temporal succession, the diachrony, of the inscription.

In draft manuscripts, consequently, the writing and its material support form an inseparable unity. ... Writing, then, is not just inscribed on, but inseparably grafted into its material support. It is visually traceable *within* (rather than merely *from*) the document. Its essence lies in its appearance bodied forth in its materiality. The documents thus, quite simply, do not host or harbour texts, or “text”, in the sense of linearly consecutive reading matter. Text as linear reading matter is always what is already copied off from the draft document, whether in acts of reading or acts of transcribing.

Some relevant websites at the Wittgenstein Archives at the University of Bergen (WAB)

Examples of WAB's XML transcriptions of the Wittgenstein Nachlass can be downloaded from http://wab.uib.no/wab_hw.page/, e.g.:

- http://wab.uib.no/cost-a32_xml/Ms-141_OA.xml

Examples of WAB's own editorial outputs from these XML transcriptions can be found on http://wab.uib.no/wab_hw.page/, e.g.:

- http://wab.uib.no/cost-a32/Ms-141_diplo.html
- http://wab.uib.no/cost-a32/Ms-141_norm.html

or on Wittgenstein Source <http://www.wittgensteinsource.org/>, e.g.:

- http://wittgensteinsource.org/BTE/Ms-141_d
- http://wittgensteinsource.org/BTE/Ms-141_n
- [http://wittgensteinsource.org/BTE/Ms-141,1\[1\]_d](http://wittgensteinsource.org/BTE/Ms-141,1[1]_d)
- [http://wittgensteinsource.org/BTE/Ms-141,1\[1\]_n](http://wittgensteinsource.org/BTE/Ms-141,1[1]_n)

Variant writing can be found in most of the Wittgenstein Nachlass items, incl. Ms-115, Ms-140 and Ms-141, all available from http://wab.uib.no/wab_hw.page/.

On http://wab.uib.no/wab_hw.page/ see esp.:

- p. 39v in Ms-140 (http://wab.uib.no/cost-a32/Ms-140,39v_diplo.html, http://wab.uib.no/cost-a32/Ms-140,39v_norm.html)
- pp.118-292 in Ms-115 (http://wab.uib.no/cost-a32/Ms-115_diplo.html, http://wab.uib.no/cost-a32/Ms-115_norm.html)

or, on Wittgenstein Source:

- [http://wittgensteinsource.org/BTE/Ms-115,118\[3\]_d](http://wittgensteinsource.org/BTE/Ms-115,118[3]_d)
- [http://wittgensteinsource.org/BTE/Ms-115,118\[3\]_n](http://wittgensteinsource.org/BTE/Ms-115,118[3]_n)
- [http://www.wittgensteinsource.org/BTE/Ms-140,39v\[1\]_d](http://www.wittgensteinsource.org/BTE/Ms-140,39v[1]_d)
- [http://www.wittgensteinsource.org/BTE/Ms-140,39v\[1\]_n](http://www.wittgensteinsource.org/BTE/Ms-140,39v[1]_n)

WAB offers “Interactive Dynamic Presentation” of the Wittgenstein Nachlass at:

- <http://wab.uib.no/transform/wab.php?modus=opsjoner>

Other examples of putting WAB's XML transcriptions of the Wittgenstein Nachlass to use include:

- in cooperation (2011-) with the Centrum für Informations- und Sprachverarbeitung an der Ludwig Maximilians Universität München (CIS), advanced Nachlass search with WiTTFind at <http://wittfind.cis.uni-muenchen.de/>
- in cooperation with the European Agora (2011-14) and DM2E (2012-14) projects, a Semantic navigation demo for the Nachlass at <http://141.20.126.236/dm2e/ajax-solr/examples/wab/>