

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

Alois Pichler

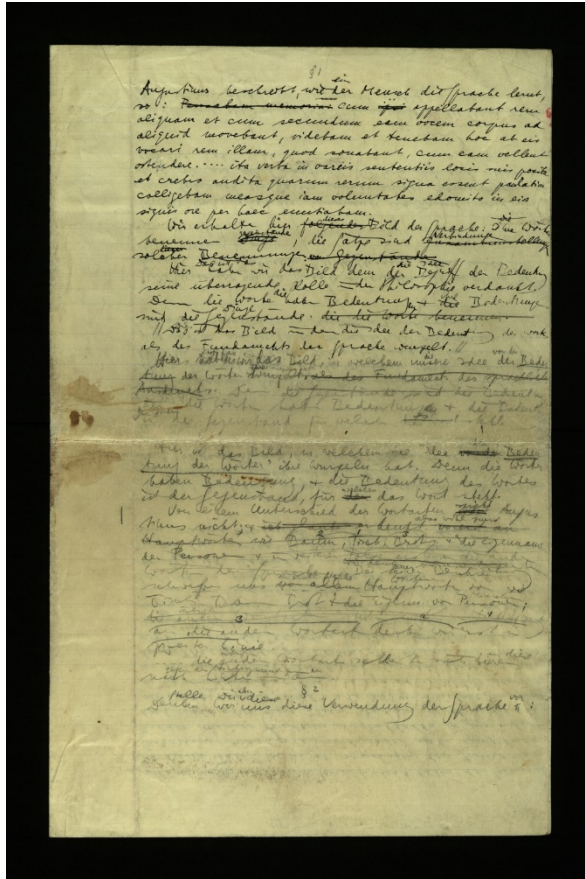
Wittgenstein Archives at the University of Bergen (WAB)

GDDH 6.6.2016

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)

1. Writing (Homo litteratus)



Wittgenstein Nachlass Ms-140,39v (

http://www.wittgensteinsource.org/Ms-140,39v_f).

With the kind permission of © 2015 The Master and Fellows of Trinity College, Cambridge; The University of Bergen, Bergen

2. Scholarship (Humanities)



*Illustration by Christian Erbacher
(Wittgenstein Archives
at the University of Bergen
and University of Siegen).
© 2015 Christian Erbacher*

3. Digital scholarship (Digital Humanities)

Select a Nachlass corpus:

Ms-114 (WS) ▼

Sort corpus according to date (optional):
No sorting ▼

Choose basic style for:	☒ Diplomatic	☐ Normalized
Additions	☒ Diplomatic	☐ Normalized
Deletions (where deleted part fits in context)	☒ Diplomatic	☐ Normalized
Deletions (where deleted part does not fit in context)	☒ Diplomatic	☐ Normalized
Overwritings	☒ Diplomatic	☐ Normalized
Editorial corrections	☒ Diplomatic	☐ Normalized

Include/exclude line breaks	☐ Include	☒ Exclude
-----------------------------	-------------------------------	--

Show/hide revisions in different writing medium (e.g. pencil in typescript)	☒ Show	☐ Hide
---	---------------------------------------	----------------------------

Highlight text types	☐ Highlight	☒ No highlight
Code	☐ Highlight	☒ No highlight
Notation	☐ Highlight	☒ No highlight

Show/hide 'section marks'	☐ Show	☒ Hide
---------------------------	----------------------------	---------------------------------------

Filter 'Bemerkungen' according to 'section mark'	☒ Show	☐ Hide
[No section mark]	☒ Show	☐ Hide
/	☒ Show	☐ Hide
\	☒ Show	☐ Hide
	☒ Show	☐ Hide
x	☒ Show	☐ Hide
—	☒ Show	☐ Hide
+	☒ Show	☐ Hide
†	☒ Show	☐ Hide
*	☒ Show	☐ Hide
?	☒ Show	☐ Hide
/?	☒ Show	☐ Hide
\?	☒ Show	☐ Hide
✓	☒ Show	☐ Hide
//	☒ Show	☐ Hide
\\	☒ Show	☐ Hide
⋄	☒ Show	☐ Hide
o	☒ Show	☐ Hide
Ø	☒ Show	☐ Hide

*“Interactive Dynamic Presentation”
offered by the Wittgenstein Archives
at the University of Bergen
for user-defined presentation of
Wittgenstein’s Nachlass
(screenshot of part of
[http://wab.uib.no/transform/
wab.php?modus=opsjoner](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.”

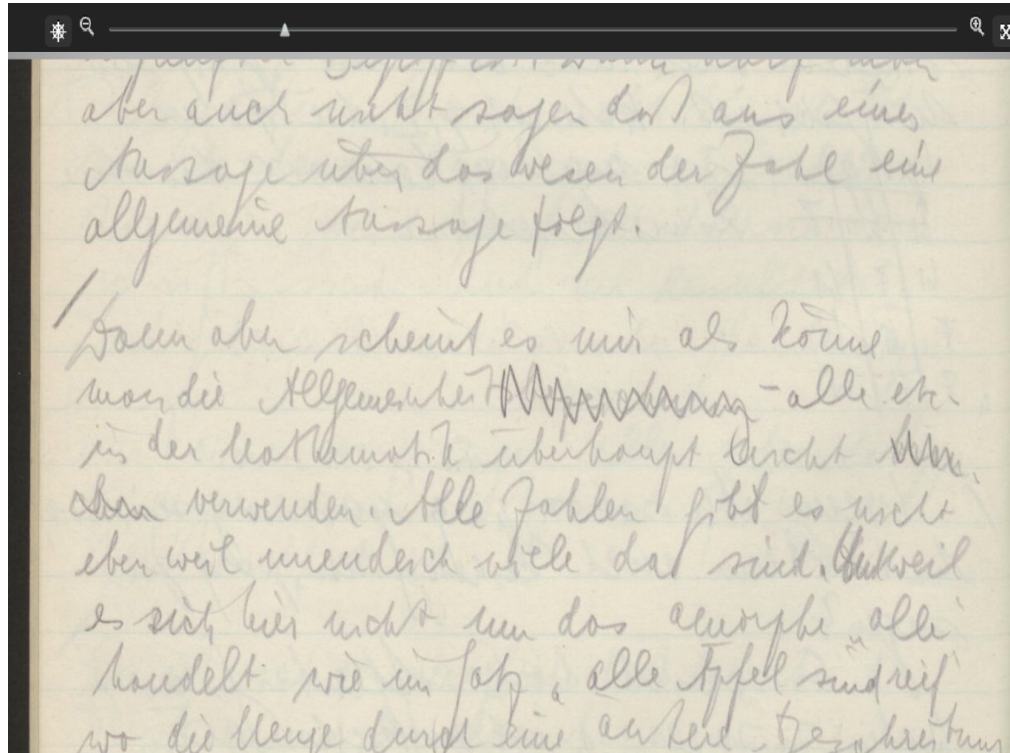
- *Schmidt 2010: The inadequacy of embedded markup for cultural heritage texts. In: Literary & Linguistic Computing 25(3), 337-356; p.344ff*

4. Philosophy (Philosophy in Digital Humanities)

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; ...	Text	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

4. Philosophy (Philosophy in Digital Humanities)



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

4. Philosophy

(Philosophy in Digital Humanities)

Diplomatic output:

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

4. Philosophy

(Philosophy in Digital Humanities)

Normalized output:

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

4. Philosophy (Philosophy in Digital Humanities)

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</orig> <orig type="em2"><choice type="dsl"><orig
type="alt1">Allgemeinheitsbezeichnung</orig> <orig type="alt2">Allgemeinheit</
orig></choice></orig></choice> ‐ alle
<abbr type="abb">etc<corr type="tra">&p.abb;</corr></abbr> ‐ in der
Mathematik überhaupt nicht
<choice type="dsl"><orig type="alt1"><del type="d">brau<lb rend="shyphen"/
>chen</orig> <orig type="alt2">verwenden</orig></choice>&p.es;</s>

4. Philosophy

(Philosophy in Digital Humanities)

Overlap

- The trees are green with white flowers.
- The <underline>trees <italic>are green</underline> 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 aperiebatur: 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 ...

- *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*