No Representation Without Taxonomies: Specifying Senses of Key Terms in Digital Humanities

Caton, Paul

pncaton@gmail.com
INKE Project

INKE Research Group

inke.project@gmail.com
INKE Project

Digital humanities practitioners typically deal with polysemous terms by specifying the intended sense of a term in accompanying documentation (when it is one of the set of terms in a schema) or by giving a localized qualification (when the term is being used in a scholarly article). Granted, practitioners do interrogate their use of ubiquitous terms: 'theory,' 'model,' and 'text,' for example, have all been critically examined.[1] These discussions, however, have not visibly affected the prevailing ad hoc, localized approach to sense disambiguation.

In ordinary language use multiple senses are the norm: we might hope for greater precision in an academic field, but cannot assume it. "After all," writes Allen Renear apropos of conflicting views on the essential characteristics of textuality, "there is not even a univocal sense of 'text' within literary studies: Barthes's 'text' can hardly be Tanselle's 'text'" ("Out" Note 1 124). The more finely senses are distinguished, though, the greater the need for documentation to point to, the greater the amount of documentation there must be, and the greater the requirement that digital resources make all the necessary pointers available.

There is a case, then, for relieving the polysemous burden carried by terms like 'text'. This could be done either by shifting some senses onto different terms or by adding an agreed upon set of clearly defined qualifiers to the original term. One example of different terms being available is the FRBR Group One entity types (IFLA Study Group 3.2). It may not have been the *intention* of the IFLA Study Group to provide alternatives for 'text', but unquestionably each Group 1 entity type - work, expression, manifestation, and item - corresponds to an existing sense of 'text' and can therefore be used in place of it. However, while these types do capture some broad distinctions, the set is very small.

More ambitious is the taxonomy of texts proposed by Shillingsburg as part of his overall concept of a 'script act.'[2] Here the semantic burden is shifted to a qualifying phrase and 'text' has the constant sense of a sign sequence (in material or immaterial state), whose existence is established by at least one material instantiation, and which is intended as a unitary communication (whether actually finished or not). Extrapolating from this, we can say that--in relation to this taxonomy--'textuality' is the exhibiting of such properties, and 'text' as a general phenomenon (that is, as a mass noun rather than a count noun) is some quantity of that which exhibits 'textuality'.

These definitions are ours and not Shillingsburg's, but derive from his definitions and are consistent with the principles upon which his taxonomy is based. Furthermore, they accord with common senses of those terms. We emphasize this both because it has methodological implications and because it helps us rethink a notion of 'text' that is well-known in the digital humanities community and to see its proper relation to the senses just described.

The quote from Renear given earlier comes from his discussion of "theory about the nature of text" coming out of the electronic text processing and text encoding localities ("Out" 107). The view Renear himself espouses--"Pluralism"--developed as a refinement of the earlier view--"Platonism"-- associated with the assertions made by de Rose et al in the paper "What is Text, Really?" This line of thinking has presented itself as *definitional*, offering a sense to associate with 'text.' Also, by emphasizing its origins in work on automated document processing, it presents this sense of 'text' as *fundamental*: that is, a more universal sense of 'text' than any sense coming from the traditional humanities localities, because it is as applicable to tax forms, memos, and technical manuals as to novels, plays, and poems. The third thing to note is that

this approach has used 'text' in both mass noun and count noun senses interchangeably, and so whatever is said about one applies equally to the other. In the Pluralist view, what defines text is the presence of one or more structures of content objects. We believe this view actually has the opposite effect of what it originally intended because, despite its avowedly universal scope, it actually imposes a greater restriction on what qualifies as a text than Shillingsburg's taxonomy does. Shillingsburg's categories have the form QualifyingLabel+'text', where 'text' has the sense of a sign sequence as described earlier. The sentence "Call me Ishmael." clearly counts as 'text' in Shillingsburg's sense, and equally clearly does not count as 'text' in the Pluralist sense - unless we dilute the sense of the phrase 'content object' until it includes standard linguistic structural units such as the clause, in which case the Pluralist sense simply becomes the same as Shillingsburg's sense.

What that line of thinking about text, texts, and textuality that runs from "What is Text, Really?" through "Out of Praxis" actually describes is a property that many--indeed most--texts exhibit, but that is not an *essential* property of a text. In a footnote to the discussion in "Out of Praxis" Renear acknowledges that the various meanings 'text' has in the various disciplinary localities do share a common ground, namely that "they all are efforts to understand textual communication." But he continues "I think that taxonomies of sense are best deferred until after we have a better understanding of actual theory and practice" (124). We think the conceptual help afforded by the clarity of Shillingsburg's distinctions shows the opposite is true: having taxonomies in place first betters our theoretical understanding.

That last statement brings out the 'chicken and egg' nature of this problem with terminology, as many scholars would doubtless argue that specifying a taxonomy like Shillingsburg's *presupposes* one's holding to a particular theory of text/textuality. Debating that, however, would in turn be helped by having a taxonomy of 'theory' available, because what that term means in digital humanities is itself hotly contested.

As helpful as we believe Shillingsburg's taxonomy to be, it only clarifies a few items of the "essential vocabulary," and while we think his overall 'script act' framework a good place to start, it needs adding to--for example, in the area that Shillingsburg calls "reception performance" (*Resisting* 77-80). Though he emphasizes his debt to McGann he doesn't attempt a taxonomy of the bibliographic codes that McGann considers such an important feature of production texts (*Textual* passim). Nor does he really say what happens to the notion of illocutionary point when we move from speech act to script act.[3] This is work still to be done.

References

Caton, Paul (2003). 'Theory in Text Encoding'. *ACH/ALLC Annual Conference*. University of Georgia, Athens, Georgia, May 2003.

Caton, Paul (2004). *Text Encoding, Theory, and English: A Critical Relation*. Dissertation. Providence, RI: Brown University.

DeRose, Steven J., et al. (1990). 'What is Text, Really?'. *Journal of Computing in Higher Education*. **2**: 3-26.

Eggert, Paul (2005). 'Text-Encoding, Theories of the Text, and the Work-Site'. *Literary and Linguistic Computing*. **4**: 425-435.

IFLA Study Group on the Functional Requirements for Bibliographic Records (2009). *Functional Requirements for Bibliographic Records: Final Report*. International Federation of Library Associations and Institutions, Amended and corrected version.

McCarty, Willard (2009). *Humanities Computing*. Basingstoke, Hampshire: Palgrave Macmillan.

McGann, Jerome (1988). *Social Values and Poetic Acts: The Historical Judgment of Literary Work*. Cambridge, Mass.: Harvard University Press.

McGann, Jerome (1991). *The Textual Condition*. Princeton Studies in Culture/Power/History. Princeton, New Jersey: Princeton University Press

Renear, Allen (1997). 'Out of Praxis: Three (Meta)Theories of Textuality'. *Electronic Text: Investigations in Method and Theory.*

Sutherland, Kathryn (ed.). Oxford: Clarendon Press, pp. 107-126.

Renear, Allen (1997). 'Theory Restored: A Response to Caton'. *ACH/ALLC Annual Conference.* University of Gothenburg, Gothenburg, Sweden, June 2004.

Renear, Allen, Durand, David, Mylonas, Elli (1996). 'Refining our Notion of What Text Really Is'. *Research in Humanities Computing.* Ide, Nancy, Hockey, Susan (eds.). Oxford: Oxford University Press.

Robinson, Peter (2009). 'What Text Really Is Not, and Why Editors Have to Learn to Swim'. *Literary and Linguistic Computing.* **1**: 41-52.

Searle, John (1979). *Expression and Meaning.* Cambridge: Cambridge University Press.

Shillingsburg, Peter (2006). *From Gutenberg to Google: Electronic Representations of Literary Texts.* Cambridge: Cambridge University Press.

Shillingsburg, Peter (1997). *Resisting Texts: Authority and Submission in Constructions of Meaning.* Ann Arbor, Michigan: University of Michigan Press.

Shillingsburg, Peter (1991). 'Text as Matter, Concept, and Action'. *Studies in Bibliography.* **44**: 32-83.

Tanselle, G. Thomas (1995). 'The varieties of scholarly editing'. *Scholarly Editing: A Guide to Research.* Greetham, D.C. (ed.). New York: The Modern Language Association of America.

Notes

1. As a representative selection from the existing literature, see Caton "Theory"; Caton "Text Encoding", *passim*; DeRose et al; Eggert; McCarty, *passim*; Robinson; Renear "Out of Praxis"; Renear "Theory Restored"; Renear, Durand, and Mylonas.

2. See *Resisting* ch. 3. This is a revised version of his "Text" where the term originally used was 'write act.'

3. On illocutionary point see Searle 2. We find no treatment of it by Shillingsburg in either *Resisting* or *From Gutenberg*.