when we are discussing interpretations of texts left to us by authors we can not consult any more.

This in turn is highly compatible to an architecture for virtual research environments for manuscript related work, where Humanities' work on historical texts is understood to consist of adding layers of changing and potentially conflicting interpretation unto a set of images of the manuscript to be interpreted. Ebner et al. (2011) have recently described an architecture for a virtual research environment for medieval manuscripts which implements this overall architecture, though using embedded markup for some of the layers for the time being.

To summarize the argument: (1) All texts, for which we cannot consult the producer, should be understood as a sequence of tokens, where we should keep the representation of the tokens and the representation of our interpretation thereof completely separate. (2) Such representations can be grounded in information theory. (3) These representations are useful as blueprints for software on highly divergent levels of abstraction.

2.1. References

Ebner, D., J. Graf, and M. Thaller (2011). A Virtual Research Environment for the Handling of Medieval Charters. Paper presented at the conference *Supporting Digital Humanities: Answering the unaskable*, Copenhagen, November 17-18, 2011 (forthcoming).

Langefors, B. (1995). *Essays on Infology*. Lund: Studentliteratur.

Rowley, J. (2007). The Wisdom Hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science* 33(2): 163-180.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal* 27(3)**:** 379-423 and 27(4): 623-656.

Thaller, M. (2009a). The Cologne Information Model: Representing information persistently. In M. Thaller (ed.), *The eXtensible Characterisation Languages – XCL*. Hamburg: Kovač, pp. 223-240.

Thaller, M., ed. (2009b). *The eXtensible Characterisation Languages – XCL*. Hamburg: Kovač.

3. Bringing together markup and semantic annotation (Dino Buzzetti)

Far from having been convincingly clarified, the relation between *markup* and *semantics* still appears to be a perplexing one. The BECHAMEL project, a consistent and systematic attempt to provide a semantics for document markup (Renear et al. 2002), aimed at introducing mechanisms and rules for mapping syntactic markup structures into semantic domains of objects, properties and relations. In a convincing article, however, Dubin and Birnbaum (2004) acknowledge that 'all the distinctions that we're able to explicate using BECHAMEL' could either 'guide the re-tagging of documents with richer markup' or 'be serialized in the form of RDF or a topic map.' (p. 8) But, in the first case, is the prospect of expressing all semantic information through the markup a viable solution? As it has been pointed out, semantic and interpretative encoding prevents interoperability, and since any 'attempt to make a document interoperational' is 'likely to result in decreased expressiveness,' markup scholars are ready to admit that interoperability is 'the wrong goal for scholarly humanities text encoding.' (Bauman 2011) On the other hand, a purely semantic description is clearly incomplete, for it might disregard equally possible and semantically equivalent textual variants.

Keeping inline markup and semantic information distinctly severed proves to be a more appropriate approach. In the case of scholarly digital editions, the sole concern with markup has left us with 'a problem that still exists,' for 'we need (we still need) to demonstrate the usefulness of all the stuff we have digitized over the last decade and more – and usefulness not just in the form of increased access, but specifically, in what we can do with the stuff once we get it' (Unsworth 2003). How can we proceed 'beyond representation' and build tools that shall 'put us into new relationships with our texts' and enable us to process their information content? (Unsworth 2004) Embedded markup can best serve as a comprehensive information carrier for textual information content, but we need further solutions to process content in an efficient and functional way. For embedded markup provides a data structure, but it does not beget a suitable data model. (Raymond 1992, 1996) It defines a format, not a semantics to process its information content. Whether Semantic Web technologies do provide satisfactory data models for humanities research is still an open question, but the problem yet remains how markup and semantic description techniques can be suitably related, by heeding carefully the basic distinction between data and information content. TEI extensions on the one side and the RDFa syntax on the other, do not seem to provide an adequate approach, failing as they do to keep format and content concerns duly severed. The apparent markup overload they produce carries with it a dubious Ptolemaic flavour.

The relation between embedded markup and semantic description languages is an indetermination relationship. Dubin and Birnbaum (2004) fully recognise its very nature: 'the *same markup* can convey different meanings in different contexts,' and 'markup can communicate the *same meaning* in different ways using very different syntax.' It is, on both sides, a one-to-many relation. If you fix the syntax, the semantics may vary in various contexts, and vice versa, if you fix the semantics, you can use a different syntax to express the same content. Contrary to the tenets of hard artificial intelligence – 'if you take care of the syntax, the semantics will take care of itself' (Haugeland 1985: 106) – and of current analytic philosophy of language – 'to give the logical form of a sentence' is to 'bring it within the scope of a semantic theory' (Davidson 1980: 144) – there is no one-to-one correspondence between the logical form of a phrase and the structure of its semantic content. We should not take for granted that by processing a string of characters representing a text, we process its information content, for we can, and often do, process a string without processing the content. And far from being a drawback, this circumstance is actually an advantage, for by dealing with indetermination we can effectively chart variation.

Both the *expression* and the *content* (Hjelmslev 1961) of the text are open to variation. Dealing with textual variants is the task of textual criticism, just as dealing with interpretative variants is that of the literary critic. But we are not at loss in tackling these problems with computational means. We can exploit the ambivalent status of markup to represent the dynamics of variation. (Buzzetti 2009) As a diacritical mark, the markup can be construed either as belonging to the text or as providing an external description of its structure. We may therefore attribute to the markup both a descriptive and a performative function. (Renear 2000) Assumed in its performative capacity, the markup can be seen as an instruction, or as a rule, to determine the semantic structure of the text, whereas taken declaratively it can be equated to a variant of the string of characters that constitutes the text. Referring to a stand-off metadata representation of the textual information content, or to a stand-off markup of sorts with semantic import, we can all the same assume their structural marks in both acceptations, declarative and performative, and get an overall dynamic model of textual and interpretative variation, as shown in Figure 1:

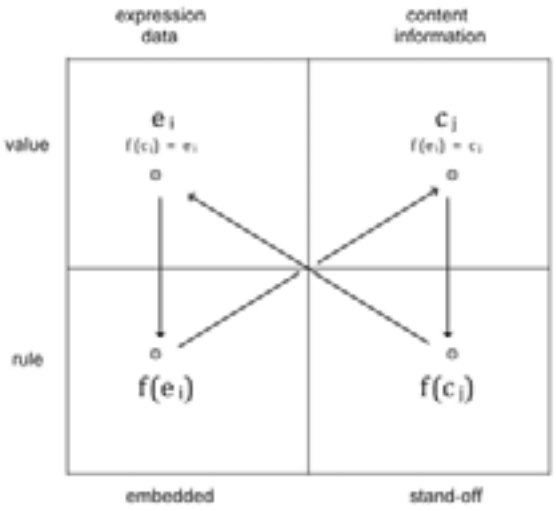

Figure 1

In this diagram, *ei* represents a specific element or construct of the *expression* of the text, conceived of as the set of all tokens that compose it, or E = { *e1* , *e2* , … , *en* }. In its performative capacity that element assumes a different logical status, and can be construed as a function *f (ei) = ci* mapping into the set of all tokens of a given content representation C = { *c1* , *c2* , … , *cn* }, whose specific elements *ci* act in a similar way as a function *f (ci) = ei* mapping into the set E of all the elements of the expression of the text.

Both kinds of variants, textual and interpretative, can be collectively represented, as a kind of 'logical sum' (Thaller 1993: 64), by means of an MVD (Multi-Version Document) graph, as shown by Schmidt and Colomb (2009). Each path of an MVD graph – a directed graph with a start-node and an end-node – represents a different version of the text. A totally isomorphic graph can be obtained also for interpretative variants. In the case of a Topic Maps representation of textual content, an MVD graph was obtained by collating textualized XTM representations of different maps referring to the same text (Isolani et al. 2009).

A comprehensive representation of this kind, of both textual and interpretative variants through MVD graphs, aims at finding efficient ways to determine which paths of the one graph, or which versions of the text, are compatible with specific paths of the other, or with different interpretations of its information content. Both graphs can be used to process the information they represent: the textual variants graph in order to visualise and display different views and versions of the text of a digital edition; the interpretative variants graph in order to process its information content. Promising and different approaches to that end have been proposed by Schmidt (forthcoming), in the context of the HRIT (Humanities Resources, Infrastructure and Tools) project, and by Thaller

(2009), through the development of the XCL (eXtensible Characterisation Language) language. The two methods can offer different implementations of the model here described for specific tasks of the editorial practice, and the pursuit of interoperability between them sets the goal for further development and research.

References

Bauman, S. (2011). Interchange vs. Interoperability. In *Proceedings of Balisage: The Markup Conference 2011. Balisage Series on Markup Technologies*, vol. 7, doi:10.4242/BalisageVol7.Bauman01 (accessed 13 March 2012).

Buzzetti, D. (2009). Digital Editions and Text Processing. In M. Deegan and K. Sutherland (eds.), *Text Editing, Print, and the Digital World*. Aldershot: Ashgate, pp. 45-62.

Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.

Dubin, D., and D. Birnbaum (2004). Interpretation Beyond Markup. In B. T. Usdin (ed.), *Proceedings of the Extreme Markup Languages 2004 Conference*, Montreal, Quebec, August 2-6, 2004 http://www.ideals.illinois.edu/bitstream/handle/2142/11838/EML2004Dubin01.pdf (accessed 13 March 2012).

Haugeland, J. (1985). *Artificial Intelligence: The very idea*. Cambridge, MA: MIT Press.

Hjelmslev, L. (1961). *Prolegomena to a Theory of Language*. Madison, WI: U of Wisconsin P.

Isolani, A., C. Lorito, Ch. Genovesi, D. Marotta, M. Matteoli, and C. Tozzini (2009). Topic Maps and MVD for the Representations of Interpretative Variants. In *Digital Humanities 2009: Conference Abstracts*, Proceedings of the 2nd ALLC, ACH and SDH-SEMI Joint International Conference (University of Maryland, College Park, June 22-25, 2009), College Park, The Maryland Institute for Technology in the Humanities (MITH), pp. 8-11.

Raymond, D. R., et al. (1992). Markup Reconsidered. Paper presented at the *First International Workshop on Principles of Document Processing*, Washington, DC, 22-23 October 1992. http://www.cs.uwaterloo.ca/~fwtompa/.papers/markup.ps (accessed 13 March 2012).

Raymond, D. R., F. Tompa, and D. Wood (1996). From Data Representation to Data Model: Meta-Semantic Issues in the Evolution of SGML. *Computer Standards and Interfaces* 18(1): 25-36.

Renear, A. (2000). The descriptive/procedural distinction is flawed. *Markup Languages: Theory & Practice* 2(4): 411-20.

Renear, A., M. Sperberg-McQueen, and C. Huitfeldt (2002). Towards a semantics for XML markup. In R. Furuta, J. I. Maletic, and E. Munson (eds.), *DocEng'02: Proceedings of the 2002 ACM Symposium on Document Engineering*, McLean, VA, November 8-9, 2002, New York, NY: ACM Press, pp. 119-126.

Schmidt, D., and R. Colomb (2009). A Data Structure for Representing Multi-version Texts Online. *International Journal of Human Computer Studies* 67(6): 497-514.

Thaller, M. (1993). Historical Information Science: Is There Such a Thing? New Comments on an Old Idea. In T. Orlandi (ed), *Discipline umanistiche e informatica: Il problema dell'integrazione*. Roma: Accademia Nazionale dei Lincei, pp. 51-86.

Thaller, M., ed. (2009). *The eXtensible Characterisation Language: XCL*. Hamburg: Kovač.

Unsworth, J. (2003). Tool-Time, or 'Haven't We Been Here Already?' Ten Years in Humanities Computing. Paper presented at the conference *Transforming Disciplines: The Humanities and Computer Science*, Washington, DC, 17-18 Jnuary 2003. http://people.lis.illinois.edu/~unsworth/carnegie-ninch.03.html (accessed 13 March 2012).

Unsworth, J. (2004). Forms of Attention: Digital humanities beyond representation. Paper delivered at *The Face of Text: Computer-Assisted Text Analysis in the Humanities*, the third conference of the Canadian Symposium on Text Analysis (CaSTA), McMaster University, November 19-21, 2004. http://people.lis.illinois.edu/~unsworth/FOA/ (accessed 13 March 2012).