

1474 I

ЕВАНГЕЛІЄ ТОВА

НАЧАЛЪ БЪШЕ СЛОВО

ОВА

ОВА

ГОНИ

ЖИ

УЛИ

ОСВѢЩЕНО БЫ



ЛѦ , Ѧ . ЗѦ , Ѧ .

НАЧАЛЪ БЪ СЛОВО , Н СЛОВО БЪ
ОУ БѦ , Н БѢ БЪ СЛОВО .
БЪ Н СКО НН ОУ БѦ . БѢ ТѢ
БЫША . Н БѢ НЕГО НН ТѢ ЖЕ

ЖИКО ПѢ
УЛИКО МѢ

<hi>&Vos;</hi>&nos;&aos;
&chos;&aos;&los;&jatos;
&bos;&jatos;&sos;&los;&oos;
&vos;&oos;

TEXT VARIETY IN THE WITNESSES OF MEDIEVAL TEXTS

Proceedings of the International Workshop

Institute of Mathematics and Informatics
Sofia, 21-23 September, 1997

DIFFERENT APPROACHES TO TEXT MODELLING

TEXTUAL FLUIDITY AND DIGITAL EDITIONS*

Dino Buzzetti

University of Bologna

Malte Rehbein

Max-Planck-Institut für Geschichte, Göttingen

I. Text and textual representation

I.1 The nature of text can hardly be defined without reference to the form of its material support. It would, however, be misleading to derive an adequate idea of text exclusively from its printed form. Producing a printed edition of a manuscript text, modifying the very form of its representation, may somehow affect its integrity. The edition of an authorial manuscript, for example, poses no minor problems. Sometimes, in a confusion of writing and sketches, the sheet of paper looks like an unwittingly perfect composition. It looks

as if a work would speak to posterity since the earliest scribbles through which it was taking shape. And not only because of what it could communicate by means of words, waiting for a more or less final disposition, but also because of the visual architecture of the page¹.

As a matter of fact, in the case of 'drafts' or outlines with alternative readings, the very placing and spatial arrangement of different portions become very important²; it has been aptly observed that the process of becoming a textual structure is there fixed in the spatial relations of chronologically different, but structurally equivalent textual units³. Now, the case we have been considering is not that exceptional and the kind of remarks here referred to authorial drafts can just as well apply to other examples of textual evolution. Many medieval manuscript traditions hand down texts to us, which far from being fixed and stable evolve over time in a way that very much resembles the process of textual production and composition undergone by a single author's work. The medieval idea of

* This paper, presented in Sofia by Dino Buzzetti, was jointly discussed by the two authors. Dino Buzzetti wrote sections I and III, Malte Rehbein wrote section II.

¹ P. Di Stefano, 'Scarabocchio dunque sono', *Corriere della sera*, 17 April 1996, p. 27.

² D. Buzzetti, 'Image Processing and the Study of Manuscript Textual Traditions', *Historical Methods*, 28(1995), p. 145.

³ H. Kraft, *Editionsphilologie*, Darmstadt, Wissenschaftliche Buchgesellschaft, 1990, pp. 110-111.

textual canonicity includes both the notion of authorship and a variable textuality reflecting scribal 'creativity' and refashioning⁴. If our notion of 'authorship' becomes problematic⁵ here, that is very much due to our inadvertent acquiescence to the textual ideal conveyed by the artifact we call the printed book. It is true that most of us almost automatically equate texts with printed books⁶, but we should be wary of confusing the form of representation, or the properties of the book, with the form of what is to be represented⁷, i.e. the nature of the text. We should not be reminded, therefore, that only as reproduced in a printed book text becomes unchanging and immutable and that a medieval text is in most cases fluid and dynamic, for in customary medieval practice fidelity to an author's work generally involves what we would call changing what the author wrote⁸.

Spatial arrangement and temporal evolution are not easy to render through the linear and unchangeable form of representation provided by the printed book. But the edition, the chosen form of textual representation, should pay due attention to the complexity of the spatial and temporal disposition of textual materials in order to be suitably faithful to the mutable form of the text to be represented. If the aim of the textual scholar remains the *constitutio textus*, the reconstruction of a text, which to a certain extent may bring back its original, having once removed all the encrustations of the centuries and of the different hands by which it has been copied or annotated from time to time, how can the editor recognize *what* is to be represented and establish *in which form* it has to be represented?

The strict correlation between the idea of textual canonicity, on the one hand, and the specific techniques of handling its material support, or the social practices of making use of it, on the other, can help us considerably in disentangling the intricacy of events, of transcriptions from manuscript to manuscript, of readings and readers' interventions, of misunderstandings and errors, of dormancies and library vicissitudes, of loss and successive recovery, of wanderings, through which texts have survived from ancient and medieval times. If it has indeed been possible

⁴ K. D. Uitti, 'Old French Manuscripts, the Modern Book and the Electronic Image', in ACH-ALLC93 Joint International Conference (16-19 June 1993), *Conference Abstracts*, Washington, D.C., Georgetown University, 1993, p. 158.

⁵ D. Buzzetti and P. Denley, 'Maestri e scolari bolognesi nel tardo Medioevo: Per l'edizione elettronica delle fonti', in L. Sitran Rea, ed., *La storia delle università italiane: Archivi, fonti, indirizzi di ricerca*, Atti del convegno: Padova, 27-29 ottobre 1994, Trieste, Edizioni LINT, 1996 (Contributi alla storia dell'Università di Padova 30), p. 206.

⁶ Uitti, 'Old French Manuscripts', p. 158, 157.

⁷ Buzzetti, 'Image Processing', p. 148.

⁸ Uitti, 'Old French Manuscripts', p. 157.

to clear the tangle of hands, libraries, and readings and to discern what manuscripts had been owned, annotated, transcribed, and commissioned by whom, this is because a history of book technologies, of men, ideas, cultures, and literary phenomena, and the mutual influences of material techniques, scripting supports, handwritings and reading manners onto the mutations of intellectual habits and learning practices, of mental attitudes towards written culture, of its production and use — all these facts have been given due attention.

The more complex the task is, the more demanding is the duty of the editor. It is worth insisting on some examples. In classical antiquity an author would prefer to work on wax tablets and nothing would remain of what had been erased. A literary work would be 'launched' in public and private *recitationes* and its written fixation was therefore destabilized by someone else's suggestions and interventions. Hence it is very difficult to analyse significant coeval variants and to distinguish authentic authorial material, possible corrections and refashionings, from material simply introduced by others. Again, in the Middle Ages written books are no more destined to a rhetorical reading, to a reciting voice and an immediate comprehension, but are devoted to an entirely scholastic reception⁹. From around the middle of the 12th century, the written book ceases to be a record of spoken words to become a record of mental thoughts; it ceases to be the trace of a discourse spoken and listened to and becomes a mirror for the mental image of its structure. But the transformation is induced by the introduction of a new material technique. A new textual ideal is brought about by new devices organising the visual arrangement of the page. A new artefact reifies a new notion of the text¹⁰ and with the new visual architecture of the page a 'grammar of the eye' is substituted for 'a grammar of the tongue'. From the 'graphics' of the page — special formulae and punctuation marks, particular scripts and coloured inks pointing out the partitions of the text — from these distinctions, which sometimes remain over long periods of time, the textual scholar can have a clue to old 'editorial' arrangements, to autonomous primitive textual units and to the different ways they combine together in the successive stages of the transmission of a work¹¹.

I.2 All the phenomena of textual mobility we have alluded to can be observed in the teaching texts produced at the university of arts and

⁹ G. Cavallo, 'Quando i libri non si chiudevano', *Il Sole-24 Ore: Domenica*, 14 April 1996, p. 21.

¹⁰ I. Illich, *In the Vinyard of the Text: A Commentary to Hugh's Didascalicon* (1993), It. transl. by A. Serra and D. Barbone, *Nella vigna del testo: Per una etologia della lettura*, Milano, Cortina, 1994, pp. 97-103.

¹¹ Cavallo, 'Quando i libri non si chiudevano', p. 21.

medicine of Bologna in the 14th and 15th centuries. These texts pose serious problems both for their reconstruction and for their interpretation. They are affected by the uncertainties of the written fixation of an oral delivery and their page layout presents all the graphic devices introduced in scholastic practice to visualise the structure of the text. Page layout is very important in establishing the original *divisio textus* and in evaluating successive modifications and interventions on the text. The process of composition, on the other hand, is the result of the teaching methods customary at the University of Bologna during that period.

In most cases, the works of the Bolognese masters take shape from the *reportationes* — or *recollectiones*, to use a term more frequently employed in Bologna¹² — of the students attending the lectures. Moreover, one peculiar character of the Bolognese teaching tradition was the practice of *repetitio*¹³, to which Anneliese Maier has repeatedly drawn the attention¹⁴ in her thorough investigation of these materials. Typical figures in the Bolognese school tradition were, in fact, the *repetitores*, young masters who acted as teaching assistants for the masters appointed to the ordinary courses, with the special duty to explain again to the students in the evening the lectures given by the masters in the morning and to make them practise on their subjects¹⁵. The effect was, in general, that a text was reduced to its essential content, or that, on the contrary, short digressions or other variations on the theme were added to it¹⁶. Traces of the activity of these lesser historical figures are preserved in the manuscripts which reproduce the works of the masters who taught in Bologna, mainly in the form of major accidents occurring throughout their textual tradition, such as anonymous marginal glosses or even interpolations of long passages within the text, but reported only by few

¹² v. D. Buzzetti and A. Tabarroni, 'Informatica e critica del testo: il caso di una tradizione fluida, in *Schede umanistiche*, n.s. 1:2(1991), p. 190.

¹³ On the problem of *repetitiones* in Bologna, besides the works of Anneliese Maier referred to here below, see A. Alichniewicz, 'Matthew of Gubio's 'Commentary on De Anima' and Its Date', *Mediaevalia Philosophica Polonorum*, 28, 1986, pp. 21-25 and especially A. Maier's, *University Training in Medieval Europe*, Leiden, Brill, 1994, pp. 59-62, 69, 122-23.

¹⁴ R. Imbach, *Averroistische Stellungnahmen zur Diskussion über das Verhältnis von Esse und Essentia*, in A. Maier and A. Paravicini Bagliani, eds., *Studi sul XIV secolo in memoria di Anneliese Maier*, Roma, Edizioni di Storia e Letteratura, 1981, p. 329.

¹⁵ A. Maier, *Wilhelm von Alnwicks Bologneser Questionen gegen den Averroismus (1323)*, in *Gregorianum*, 30 (1949), pp. 265-308, now in *Ausgehendes Mittelalter*, I, Roma, Edizioni di Storia e Letteratura, 1964, p. 2, nota 3.

¹⁶ Ead., *Eine italienische Averroistenschule aus der ersten Hälfte des 14. Jahrhunderts*, in *Die Vorläufer Galileis im 14. Jahrhundert*, 2. Aufl., Roma, Edizioni di Storia e Letteratura, 1966, p. 254-255.

or just one copy¹⁷. Thus, with regard to works which present similar phenomena in their transmission, it is appropriate to speak of a mobile or 'fluid' textual tradition¹⁸.

In short, the persistence and transformations of the teaching practices is reflected directly in the evolution of the text during the successive stages of its production and use, a fact that can be documented in a precise and punctual manner¹⁹. A thorough examination of some of the earlier works produced in the University of Arts and Medicine in Bologna²⁰ actually shows a process of continual adaptation and refashioning, a common event, which has been confirmed by a number of explorations of later works. It is therefore plausible to assume that a form of progressive manipulation and rearrangement of the text is a vicissitude common to many, if not all, the teaching books produced in Italian universities during the 14th and 15th centuries. Textual fluidity and mobility appear to be a common feature of the literary production connected with the practice of the ordinary academic activities.

I.3 An interesting case of mutual integration and completion of two parallel textual traditions may be recalled in order to illustrate the editorial and interpretative problems emerging from this kind of literature. The two Bolognese commentaries on *De tribus predicamentis*, the last chapter of William Heytesbury's *Regule solvendi sophismata* (1335), written by Mesino de' Codronchi and Angelo da Fossombrone around the turn of the 14th century, are handed down by a number of extant manuscripts copied before their early printed edition of 1494²¹. The two commentaries are incomplete: Mesino commented all three parts of Heytesbury's work, but left the third part unfinished, whereas Angelo dealt only with the first part, *De motu locali*, treating however its subject in a more thorough and exhaustive manner. Very early, then, the first part of 'Mesino's

¹⁷ cf. D. Buzzetti, P. Pari, and A. Tabarroni, 'Libri e maestri a Bologna nel XIV secolo: un'edizione come database', in *Schede umanistiche*, n.s. 2:2(1992), pp. 165-166.

¹⁸ cf. F. Del Punta, 'La Logica di R. Feribrigge nella tradizione manoscritta italiana, in A. Maier, ed., *English Logic in Italy in the 14th and 15th Centuries*, Napoli, Bibliopolis, 1982, p. 53.

¹⁹ D. Buzzetti, R. Lambertini, A. Tabarroni, 'Tradizione testuale e insegnamento nell'università di medicina e arti di Bologna dei secoli XIV e XV', in *Annali di storia delle università italiane*, 1(1997), p. 78.

²⁰ Cf., for example, the commentary of Gentile da Cingoli on Porphyry's *Isagoge* presumably composed in the first decade of the 14th century and the *Breviloquia* of astronomy and geomancy by Bartolomeo da Parma, written around the turn of the 13th century.

²¹ Messinus, *Questio de motu locali cum expositione tutius tractatus Hentisberi de tribus predicamentis*, in Guillelmus Hentisberus, *De sensu composito et diviso; Regule solvendi sophismata; etc.*, Boneto Locatello, ed. Ottaviano Scoto, Venezia 1494 (Hain *8437; I.G.I. 4618), ff. 52va-64ra and Angelus Forsempronensis, *Scriptum supra tractatu de motu locali*, ibid., ff. 64ra-73rb.

commentary' was replaced by the lengthier *De motu locali* written by Angelo, whereas for obvious didactic reasons, 'Angelo's commentary' was completed with *De motu augmentationis* and *De motu alterationis*, the second and third parts written by Mesino. In more than one codex we even find a transcription of 'Angelo's commentary', completed with a reduced version of Mesino's last two sections, right together with 'Mesino's commentary' in a more extended version comprising all its three original parts as well as the completion of the third one by Gaetano da Thiene. One gets the impression that the need for a satisfactory and complete commentary, determined by the increasing diffusion and systematic study of Heytesbury's *Regule*, might have given birth to two distinct and relatively independent textual traditions, with successive phenomena of cross borrowings and contaminations²².

This example shows two important common facts. With texts of this sort, the author is left with hardly more than an eponymic function in relation to a textual tradition, which develops through frequent contaminations and interchanges, freely moving away from the original version of the text²³. Moreover, one notices immediately, as plain as it is, that in order to achieve a fair understanding and a suitable interpretation of these texts it is absolutely necessary to take into account the vicissitudes of their academic use. The examination of their transmission shows that not only is their origin connected with the teaching practice, but also their form is modelled by their classroom use. It has thus been possible to ascertain

not only that the diffusion and the circulation of the doctrines depended on the production of the texts, but also that the changes and the final shaping of the texts depended on the diffusion of the doctrines and the persistence of their teaching. The ongoing classroom use is what determines, sometimes in a process lasting over many generations, the final form of the works handed down to us²⁴.

Considering these facts, how can the editor provide for a fair representation of textual fluidity and how can all peculiarities of a mobile tradition be accounted for in a suitable edition? The printed book model

²² D. Buzzetti, 'Linguaggio e ontologia nei commenti di autore bolognese al *De tribus praedicamentis* di William Heytesbury', in D. Buzzetti, M. Ferriani, A. Tabarroni, eds., *L'insegnamento della logica a Bologna nel XIV secolo*, Bologna, Istituto per la Storia dell'Università, 1992 (Studi e Memorie per la Storia dell'Università di Bologna, n.s. VIII), p. 587.

²³ Buzzetti and Denley, 'Maestri e scolari bolognesi', p. 206.

²⁴ D. Buzzetti, 'La faculté des arts dans les universités de l'Europe méridionale: Quelques problèmes de recherche', in *L'enseignement des disciplines à la Faculté des arts (Paris et Oxford, XIIIe-XVe siècles)*, Actes du colloque international édités par O. Weijers et L. Holtz, Turnhout, Brepols, 1997 (Studia Artistarum, 4), p. 465.

is clearly not up to such a demand and an obvious alternative seems to be afforded by digital editions. But is it altogether clear what they can really achieve? It very much depends on their functional features, so let us consider an actual implementation, which is not so much contrived as a practical tool, but rather as a new analytic method or a new form of conceptual organisation and structural representation of the text. What results can we expect to obtain by applying that method to the production of a digital edition?

II. The Digital Text Edition

II.1 The text of an edition is based on the manuscript the editor considers to be most important. What is this principal manuscript, or rather: who defines it? The 19th century discussion about the German *Nibelungenlied* shows that very often there is no answer at all. While Karl Lachmann suggested A as the manuscript closest to the original, Karl Bartsch thought the same of manuscript B and Friedrich Zarncke and Adolf Holtzmann of C²⁵. For this reason, the *Nibelungenlied* was edited in three different versions. Meanwhile the discussion about the *Nibelungenlied* has continued but as yet no conclusion has been reached. It is even possible that someday a still unknown manuscript will be found revealing new facts about the original.

The user of a textedition, the philologist, the philosopher or the historian, obviously has to follow the editor's interpretation of the different text witnesses. But what is he supposed to do if he is using Lachmann's edition of the *Nibelungenlied* but is interested in manuscript B? Using the apparatus of variants he has to reconstruct B, which is an avoidable and, therefore, unnecessary effort. Perhaps, this is still possible (albeit difficult) in the case of the *Nibelungenlied*²⁶. Let us introduce another example: the *scala coeli* of Jean Gobi²⁷ has come down in 23 manuscripts which can now be found in different towns, e.g. in Göttingen. It could be

²⁵ For further information about this discussion see: Georg Steer, 'Textkritik und Textgeschichte. Editorische Präsentation von Textprozessen: Das Nibelungenlied. Der Schwabenspiegel. Die Predigten Taulers', in *Methoden und Probleme der Edition mittelalterlicher deutscher Texte*, hg. von Rolf Bergmann und Kurt Gärtner, Tübingen, 1993 (Beihefte zu Editio, 4), pp. 107-119.

²⁶ The *Sachsenspiegel* of Eike von Reggow for instance is handed down to us in about 460 manuscripts. See: Schmidt-Wiegand, 'Überlieferungs- und Editionsprobleme deutscher Rechtsbücher', in *Methoden und Probleme*, p. 65. An apparatus of variants in a printed textedition that will describe this completely is impossible.

²⁷ See: Alain Gurreau, Marie-Anne Polo de Beaulieu, 'Classement des manuscrit et analyses factorielles. Las cas de la scala coeli de Jean Gobi', in *Bibliothèque de l'école des chartes*, 154 (1996), pp. 359-400.

that one user is especially interested in the Göttingen version. Maybe he is dealing with local history and the suggested text, based on the editor's principal manuscript, has never been read in Göttingen and cannot have had any influence in the history of that town.

This question of the principal manuscript has often been discussed and will not be the topic of this paper. But it leads us to the following statement: there is no (printed) text edition that can satisfy the wishes of *all* users. While a historian might be interested in the manuscript of a special town, another (perhaps for the same reason) is interested in that of a different town and a third one is working on the history of textual transmission and needs all the different textvariants. Finally, a fourth historian could be interested in the biography of Jean Gobi and wants, therefore, a reconstruction of the text which is as close as possible to the original.

We would like to present a method that makes it possible (among other things) to handle the above-mentioned problem: an individual interpretation of the apparatus of variants without any additional effort for the user. We call this method a *digital text edition*. In the following, a further problem of printed editions will be examined. It is just as unnerving to work out the text passages various manuscripts have in common or where they differ from one another, as it is to reconstruct one single manuscript using the apparatus of variants. Steer summarises that problem for the *Nibelungenlied* as follows:

Für das Nibelungenlied ist eine Ausgabe denkbar, die auf den Handschriften A, B und C aufbaut, diese in Synopse dargeboten, die durch drucktechnische Akzentuierungen Textgemeinschaften wie Textunterschiede hervorhebt und in einem die Ausgabe begleitenden Kommentar auf erschließbare frühere Textbestände und Schichten, auch auf vermutete originale Positionen aufmerksam macht²⁸.

But mechanically this is only possible if you have no more than two or three text witnesses²⁹. The digital text edition offers the possibility to show text attributes, like different or common parts, in an individual

²⁸ Cf. Steer, 'Textkritik und Textgeschichte', p. 118. He says that it is worth thinking about an edition of the *Nibelungenlied* based on a synoptical presentation of the manuscripts A, B and C. By using the technical facilities of print common and different text passages will be marked. Furthermore, earlier parts of the text, as they can be revealed, and proposed original positions will be explained in a companion commentary.

²⁹ This has been tried for the *Nibelungenlied* by M. S. Batts using A, B and C. Batts mentioned (p. VII) that this would not be possible for seven manuscripts (A, B, C, D, I, b, d). See: *Das Nibelungenlied. Paralleldruck der Handschriften A, B und C nebst Lesarten der übrigen Handschriften*, hg. von Michael S. Batts, Tübingen, 1971.

manner, i.e. referring to the problem that the user is working on. We must emphasize that this is *not* an electronic tool for making a printed edition (although this would be possible) but a new, computer based method.

II.2 First of all, an explanation of what a digital textedition looks like for the user will be given. As a common wordprocessing programme, the computerscreen is divided mainly into two parts: one window containing the text and another containing a command area, where the different functions can be selected. The textwindow — the working area — is comparable to the text of a traditional, printed edition. It contains a version of the text, or rather, an interpretation of it, with its additional attributes, i.e. the textual criticism that one would normally find in the critical apparatus of annotations and variants of a printed edition. But these are visualised in a way the user requires.

Although this aspect will be dealt with in greater detail further on, let us consider the following example: a typewritten manuscript often contains handwritten marginal notes — additional annotations by the writer or the readers. In a conservative text edition this would be marked, for instance, as handwritten notes in the critical apparatus. In a digital textedition, however, there is the possibility to interpret the characteristic handwritten by entering a simple command. For instance, it can be made visible by using another background colour or Italics for this part of the text. Should there be various annotations in different colours (possibly made by different people), they can be displayed in the original textcolour. This makes it possible to see the additional information at once. Furthermore, if the editor or the user recognises that the manuscript has been written by different hands, these can be visualised as well. While these features are quite normal in modern printed editions, the digital textedition goes one step further. If you are *not* interested in the mentioned details (any more), you may easily conceal them, if they are not essential for the question you are dealing with, and they will not distract you while working on the text. The digital textedition restricts itself to the kind of information required at that moment and the user's attention is focused on what is really needed. Thus it is possible to include every kind of information in the digital edition. Neither the clarity of the layout, nor the user's ability to absorb matters are being overexerted.

Another important, very central possibility offered by the digital textedition, is the connection with a database system. As databases are not the topic of this paper, one example will be sufficient to show what integrating a digital edition into a database means. The digital textedition itself offers the chance to visualise all available attributes. This means, it occasionally can 'answer questions' such as: 'Show me all handwritten

annotations in manuscript B, C1 and E'. In this, however, it is restricted to formal characteristics. Together with an underlying database structure, which relates to the content of the text, a more refined search is possible, such as 'show me all references to persons in all manuscripts, which have been added in another shade of ink'.

II.3 The problems of handling variants in printed texteditions have already been characterised in the first paragraph. In the following we assume that the text we are analysing has already been edited digitally. The corresponding file — there is only *one* single file containing all the manuscripts — has been opened in a programme for digital texteditions, called *DTEprogram*. Now, we are considering the problems from the user's point of view, the one who wants to work with the edition and look at the possibilities that are offered.

Let us assume that the *scala coeli* of Jean Gobi we referred to at the beginning has been edited in a digital edition containing all 23 witnesses. The historian working on Göttingen had so far to reconstruct the manuscript important for him by using the principal manuscript and the apparatus of variants³⁰. Now he is able to switch to the manuscript in question (H in this case) by pressing just one key in the DTEprogram and sees precisely this (Göttingen) manuscript displayed in the textwindow. Without wasting any more effort or time, it is now possible to read any text witness you like — even only parts of it.

Not always, however, is the user of an edition interested in all or in specified variants. Often he merely wants to follow the editor's interpretation and read the text worked out by the latter. This is just as simple as choosing the option editor's text in the DTEprogram. By analogy with the above-mentioned case of the Göttingen manuscript, the text based on the editor's principal manuscript is then displayed in the textwindow.

These possibilities of looking at different variants are executed straight away by the programme without opening up any additional textfile. Because all manuscripts are stored in only one file, comparisons between different witnesses can be effected very quickly and easily. Making the differences and similarities of the manuscripts demanded by Steer for the *Nibelungenlied* visible can be achieved with no problem. Choosing the textcolour black for the manuscript in question (e.g. A), the DTEprogram shows all the passages of this manuscript that differ from B and C on user's request in another colour, for example in red. In one view the user sees the identical and different textparts of the three

³⁰ This would be, by the way, impossible with the edition of Marie-Anne Polo de Beaulieu, because she only works on three main manuscripts in the apparatus of variants: Marie-Anne Polo de Beaulieu, *La scala coeli de Jean Gobi*, Paris, 1991.

witnesses. In another view he can see the difference between the Göttingen and the Madrid manuscript of the *scala coeli*. In the same way, the programme can show all the textpassages that differ in *any* possible manuscript using different colours (or different fonts, for instance), and the programme will visualise the smallest common ground of the history of textual transmission, i.e. the textparts that are identical in *all* manuscripts.

Working out the text, the editor often has to take elements from different witnesses and bring them in line. The digital textedition offers the possibility to make this evident by setting off the corresponding textparts, for instance with different textcolours. If the editor mainly follows manuscript A (i.e. A is the principal manuscript), but uses parts of B and C in between, passages from B will be displayed in a different colour from the parts of C, while the rest (A) is displayed in the basic colour.

The use of different colours to make textvariants visible leads us immediately to another possibility the digital textedition provides. The making of a stemma for the *scala coeli*³¹ raises the following question: In the textparts used for the analysis 363 different variants were found that occur altogether 896 times in the 23 manuscripts. How are these variants distributed referring to their number and the manuscripts? GUERREAU and POLO summarised their result in a statistical survey. For example, 227 textpassages occur only in *one* manuscript (*hapax*). K contains 38 of such differences, O, however, does not contain any. But where and in which text witnesses are these *hapax* to be found? Do they occur very often in particular parts of manuscript K? This *might* be a hint that the text from which K had been copied from was damaged in that part — an important result for analysing the history of textual transmission. As in the above examples, these questions can easily be visualised in the digital textedition.

But the last aspect, in particular, can easily lead to wrong conclusions, if one exclusively relies on such — completely quantitative — results. The digital textedition is not meant to spare either the editor, the historian or the philologist any exertion of the brain. It is only a tool by which the analysis of certain, individual questions can be made easier by displaying textual criticism that otherwise would take much more time to elaborate.

Whereas so far we have only described the benefits for the user, we shall now change to the work from an editor's point of view and explain the principles of building a digital edition.

First, we assume that the text to be edited is already available in a machine-readable format, such as an ASCII file. Even this raw text can

³¹ See: Gurreau, Polo, 'Classement des manuscrit et analyses factorielles', pp. 364ff.

be used as a digital edition. However, to make use of the features described above and those described later, you have to extend the textfile, i.e. the additional information has to be integrated. This is basically the same work that the editor of the printed textedition has to do while constructing the apparatus of annotations and variants.

For the following explanations some technical expressions are essential. A mark-up is, in our definition, the assigning of certain (initially invisible) characteristics of a textpassage³². The integration of these additional textattributes, i.e. comments on textcolour, hands, variants and so on (see part IV), is done by including such a mark-up into the text. The textparts showing a specific characteristic, for instance that they occur as a handwritten marginal note, will be selected and assigned an attribute called 'handwritten note'. In an ASCII representation, this attribute would presumably occur as a 'tag'; whether this is so, however, need not bother the digital editor.

Thus, all the possible available information will be included in the text. Nevertheless, neither the editor nor the user has to worry about the internal structure of the digital textedition. No one has to learn or use the described mark-up language³³. The DTEprogram provides a user-interface through which the mark-up can be constructed. This interface is similar to word-processing programmes as far as functionality and use are concerned.

Assuming that the text is only available in one witness, the edition is carried out by simply describing and including the characteristics of this manuscript. This digital textedition can be used with the DTEprogram. But in order to handle different textwitnesses, further steps are required. Generally there are two different possibilities:

1. Using the procedure for marking-up single textpassages as described above, one digital edition for every manuscript is created. As a provisional result there are n different textfiles. A programme for text collation, like *Collate 2*³⁴ connects the single files and generates one file, which is then translated into a digital textedition by the DTEprogram. Since the mark-up of the single textwitnesses has been done previously, the collation programme is able to recognise even differences in the

³² Probably the best known mark-up is HTML, the language of the world-wide-web.

³³ The purpose of the mark-up language is mainly to make sure the digital textedition will be readable in future times. They consist only of pure ASCII characters and are independent from any internal format, like the ones you will find in common word-processing programs. Such special formats of word-processing programs become outdated rapidly and can then no longer be read.

³⁴ P.M.W. Robinson, *Collate: Interactive Collation of Large Textual Traditions, Version 2*. Computer Program distributed by the Oxford University Centre for Humanities Computing, Oxford, England, 1994.

additional text characteristics. For example, it is able to discover whether a word is written in small letters in manuscript A but in capitals in B. The final result is the digital textedition including all variants. It is the logical sum of the different textwitnesses and textattributes.

2. In many cases, however, the different witnesses differ only in a few passages, in single words or parts of sentences. While the possibility described as number 1 assumes that every manuscript already exists in a machine readable format, we now want to introduce a method that can do without a manual input of all manuscripts. Obviously it is necessary that at least *one* textwitness is available in a format that can be read by the DTEprogram, that means as a textfile (in ASCIIformat). This is the *basic manuscript* that forms the basis for further processing³⁵. In this basic file, textpassages that are different in other manuscripts are marked and the corresponding information (text in manuscript X =...) is included. For example, consider the following sentence of the *Gesta Frederici*³⁶:

(A): *non longe a Danubio super fluvium Werenza dictum
tabernacula locavit*

(B): *non longe a Danubio super fluvium Werenza dictum
castra posuit*

Assuming that A is the basic manuscript then we have to mark *tabernacula locavit*, choose manuscript B from the command area and type in *castra posuit*. Thus, all the differences in all the manuscripts are taken into consideration and we obtain the same result as before: the logical sum of all text witnesses.

Having worked out the digital edition by using one of these procedures, the editor still has to define the principal manuscript and to choose the variants he wants to use in order to create a readable text. This is also done by marking up the corresponding textpassages referring to all manuscripts. Internally, the editor's text is, thus, treated like just another text witness. It is possible to switch between the variants (and to the editor's text) and to visualise the differences in the text without any additional work.

Of course this paper cannot deal with all the possibilities of handling variants offered by the digital textedition. But the authors hope to give at least an impression about what working with this new method can be like.

³⁵ It need not be the principal manuscript, although this would be helpful.

³⁶ *Gesta Frederici*, I, 19.

II.4 In this concluding section we shall now concentrate on further advantages of the digital textedition. We shall discuss in particular the meaning of *interpretative environment*.

Let us recall the example of the second paragraph. There we learnt one of the possibilities of the digital textedition, namely the fade-in and -out of handwritten marginal notes in a typewritten text. Furthermore, we pointed out the possibility to visualise annotations in different ways. If the user of the edition is interested in this information, i.e. the annotations, he selects the corresponding option and decides, furthermore, *how* the annotations are to be displayed. The digital textedition offers, for instance, the chance to show these marginal notes in a different type or with another background colour. Thus, the required information can easily be grasped. When it is not needed, the user simply turns them off and they will no longer distract him.

Using the mark-up described earlier in this paper, the editor can include any characteristic of the manuscripts in the edition. In a printed edition he had to find a compromise between the essential textual criticism and the restrictions of the layout³⁷. It has already been shown that the digital textedition does not have such limitations.

The concept of an *interpretative environment* means the following:

1. Every critical comment provided by the editor can be asked for, and visualised by, the user.

2. It is the user who selects the information he wants to be displayed, how this is to occur shall happen and what information should remain invisible.

3. This is done *interactively*, i.e. the *interpretative environment* can be varied permanently and the result will be immediately visible.

If one is determined to create a critical apparatus as detailed as possible in a printed edition, the clarity of the layout automatically suffers. When considering, for instance, the 1898 edition of the *Liber Pontificalis* of Theodor Mommsen³⁸, one will notice that on many pages the critical apparatus is much greater than the actual text. To satisfy the various witnesses, the layout is sometimes divided into two or three columns even in the apparatus of variants. Square brackets or small apostrophes mark additional information about the history of textual transmission. It is quite difficult to read such an edition while analysing the critical annotations at the same time. But in not providing this apparatus, however,

³⁷ Sometimes it is impossible to find such a compromise. An example is given by G. Schmitz, 'Unvollendet Eingestampft Kassiert. Nie Erschienenes und Mißglücktes', in: *Zur Geschichte und Arbeit der Monumenta Germaniae Historica*, Ausstellung anlässlich des 41. Deutschen Historikertages, München, 17.-20. September 1996, München, 1996, pp. 64-73.

³⁸ Unveränderter Nachdruck, *Monumenta Germaniae Historica*, München, 1982.

the scientific standard of the edition is questionable. The digital textedition can cope with both: the clarity of the layout (using the interactive interpretation) and the scientific standard (offering a complete textual criticism).

Some things are difficult or even impossible to print. A simple example of such a situation occurs later, when specific colours of ink are associated with specific authors. A more complex case: if the editor is uncertain about the reading of a specific textpassage, in a printed edition he has only the possibility to mark it, e.g. with a question mark. The digital textedition offers a link to the corresponding part of the original manuscript³⁹. When the user follows this link in the digital edition, he gets a faithful copy of the uncertain reading in question on screen and can decide for himself whether to follow the editor's interpretation or not. In the same way, it is possible to complement or replace the verbal description of a *chrism* with a detail out of the original manuscript. Again, it depends on the user if and how he wants to use this possibility or whether a stylised figure or a verbal description is sufficient.

There are many more examples of the interpretative visualisation of the critical apparatus of a text that cannot be dealt with in this paper. On the whole, the digital textedition offers the chance to represent a text *individually*, which in a printed edition is not possible.

II.5 It is, furthermore, not possible to explain all the advantages the digital textedition has in detail. Nevertheless, we must emphasise that in comparison with a classical edition no information, no component of the textual criticism is lost. Every element of the printed edition can be brought into the digital edition but the user need not use all the features provided. Moreover, it is the editor's responsibility to decide the kind of information to be included or excluded. Basically, however, he is able to include as much information as he wishes because he need not worry about the clarity of the layout or the technical restrictions of print.

The digital textedition is more than an interactive toy. Principally, the connection with a database system, and the possibilities to visualise effortlessly individual problems, turns it into a serious, promising instrument of research for the humanities.

³⁹ This had to be digitalised, i.e. scanned, before. Since scanning and including in a database is the first step in digitization of historical sources, this step will have been done often in advance. But this is not the topic here.

III. Conceptual and computational models of the text

III.1 Digital editions present an alternative to printed editions. In abstract terms the form of representation provided by the printed book is basically linear and synchronic. To an extent, the printed book also makes use of the bidimensional arrangement of the page. The very placement of the words defines titles, notes, marginal glosses, captions, and so on. Such forms of the visual organisation of the page show that text cannot be represented in a one-dimensional form. The fundamental problem of textual representation is to seize upon the multidimensional nature of textual content. The very idea of textual mobility makes us realise that text is not to be identified with any of its several forms of representation. A textual representation is by necessity concrete and needs a physical support. Text, on the contrary, is abstract and may be conceived as the *invariant* content of all its material representations⁴⁰. But a text comes into existence only through the representation of its content. Therefore, what we know about a text, what is made explicit about it, depends on the form of its representation. The sense of what is signified, Pascal reminds us⁴¹, is changed by the words, which express it. A different form of representation conveys different information about its content. The digital edition is a step forward in the direction of a multidimensional form of representation. It captures different aspects of textual mobility such as spatial arrangement and temporal evolution. Spatial disposition and change over time are features of textual representation, which visibly reveal the abstract and mobile nature of a fluid text.

The capability of a digital edition to visualise spatial and temporal modifications depends on the way textual information can be represented and processed in machine-readable form. Therefore, the limitations of 'text' as a digital data type have to be overcome. In describing the storage and processing of textual materials by means of a digital computer, text is commonly understood as information coded as characters or sequences of characters. But this is not text in the sense of literary material: a mere sequence or string of characters is simply not able to represent all the information conveyed by a literary work⁴². Textual mark-up has been

⁴⁰ Cf. C. Segre, *Avviamento all'analisi del testo letterario*, Torino, Einaudi, 1985, p. 29: If we think of graphic signs (characters, punctuation marks, etc.) as meaning sounds, pauses, etc. and we consider that these signs may be transcribed many times and in many ways (for instance by means of different scripts and fonts), whereas their value remains the same, we can conclude that text is the invariant, the sequence of values, with respect to the variables for characters, script, and so on.

⁴¹ Cf. B. Pascal, *Pensees et Opuscules*, ed. L. Brunschvicg, Paris, Hachette, 1957, 23 and 50.

⁴² A. C. Day, *Text Processing*, Cambridge, Cambridge University Press, 1984, pp. 1-2.

introduced precisely for that purpose. By textual *mark-up* here we mean all the information in a document other than the 'contents' of the document itself, viewed as a stream of characters. In a document, therefore, any information present about formatting (margins, font shifts, page breaks) is given by mark-up. All structural information, all relationships among texts, apparatus, and notes, is given by mark-up. All analytic or interpretative information to be included in an electronic text is by definition expressed as mark-up.

So, claiming that the only essential part of a text is its sequence of graphemes represents a misguided and inadequate theory of texts⁴³. On the contrary, text for philologists is not just a sequence of characters; it has structure(s) that can be exploited by researchers⁴⁴. But can these structures be suitably described by mark-up? Is textual mark-up up to its task? Can it make a digital representation fully match a literary text? Some doubts have already been cast upon its adequacy — or at least upon the adequacy of its currently accepted practice. The basic difficulty seems to reside in the inability of mark-up to provide a suitable data model. Mark-up is not in itself a data model, or a mathematical formalism⁴⁵; it is not a mathematical abstraction⁴⁶ for textual structures, an abstract object defined independently of its representation, quite apart from a data structure⁴⁷ such as a marked up stream of characters. And in the absence of a proper data model, the form of the representation, the marked up stream of characters, is mistaken for the form of what is to be represented, the surface and deep structures of the text.

III.2 The inadvertent shift from data as *representation* of a given content, as information coded in a special way⁴⁸, to data as *information* modelled according to a given abstract structure, as a formal object to be represented through a specific form of coding, seems to occur in the very definition of what is mark-up⁴⁹ endorsed by the editors of the Text Encoding Initiative (TEI), the most widespread scholarly attempt to provide a format

⁴³ C. M. Sperberg-McQueen, 'Text in the Electronic Age: Textual study and text encoding, with examples from medieval texts', in *Literary and Linguistic Computing*, 6(1991), p. 35.

⁴⁴ D. J. Birnbaum, 'How Slavic Philologists Should Use Computers', in *Computer Processing of Medieval Slavic Manuscripts*, Proceedings of the First International Conference, 24-28 July 1995, Blagoevgrad, Bulgaria, Sofia, Akademichno izd-vo "Marin Drinov", 1995, p. 20.

⁴⁵ D. R. Raymond, F. W. Tompa and D. Wood, 'Markup Reconsidered', paper presented at the First International Workshop on Principles of Document Processing, Washington DC, October 22-23, 1992, p. 4.

⁴⁶ *Ibid.*, p. 16.

⁴⁷ *Ibid.*, p. 4.

⁴⁸ Day, *Text Processing*, p. 1.

⁴⁹ L. Burnard and C. M. Sperberg-McQueen, *Living with the Guidelines: An introduction to TEI tagging*, Text Encoding Initiative, Document Number: TEI EDW18, March 13, 1991, p. 2.

for data interchange in humanities research and to suggest principles for the encoding of texts in the same format⁵⁰. If mark-up is to be defined as all the information contained in a computer file other than the *text* itself, how can indeed *any* aspect of a text of importance to a researcher ever be signalled by mark-up⁵¹? Either mark-up represents that information, which *is not* part of the text⁵² and is *other* than text, or mark-up represents aspects of that information which *is* part of the text, and is *the same as* text. But not both, unless *text* is taken, as it is here, in two different senses, as a stream of characters contained in a file on the one hand, and as the textual content of a written document on the other. But the representation of any information content is not the information content, which is represented by that representation. A picture is not a landscape and a word is not its meaning. Information, inasmuch as it is coded as characters or sequences of characters⁵³, is not information inasmuch as it is conceived as the textual content of a written document. Text as a digital expression, cannot be text as the content of the expression⁵⁴. Otherwise one inevitably mistakes text for a data type, or a text for its representation, a text for a document⁵⁵, be it a printed document or a digital one.

This very last confusion seems to be induced by the choice of SGML as a metalanguage for the description of text-encoding schemes⁵⁶. Somewhat inappropriately for a formal syntax such as it is really, SGML is sometimes described as a metalanguage, for it has syntax, but not semantics⁵⁷. In other words, SGML does not itself define a mark-up language, but is a language for defining *mark-up languages*, i.e. sets of mark-up tags with rules defining when they are applicable and how they can interrelate⁵⁸. It does not specify a particular set of tags, but rather provides a way for declaring which tags are to be used, along with their

⁵⁰ L. Burnard, 'An Introduction to the Text Encoding Initiative', in D. Greenstein, ed., *Modelling Historical Data*, St. Katharinen, Max-Planck-Institut für Geschichte i.K.b. Scripta Mercaturae Verlag, 1991, p. 83.

⁵¹ Burnard and Sperberg-McQueen, *Living with the Guidelines*, p. 2.

⁵² J. H. Coombs, A. H. Renear and S. J. DeRose, 'Markup Systems and the Future of Scholarly Text Processing', in *Communications of the ACM*, 30(1987), p. 934.

⁵³ Day, *Text Processing*, p. 1.

⁵⁴ Coombs et al., 'Markup Systems', p. 934.

⁵⁵ D. Buzzetti, *Il testo 'fluidi': Sull'uso dell'informatica nella critica e nell'analisi del testo*, in Luciano Floridi, ed., *Filosofia & informatica*, Atti del primo incontro italiano sulle applicazioni informatiche e multimediali nelle discipline filosofiche (Convegno Nazionale della Società Filosofica Italiana: Roma, 23-24 novembre 1995), Torino, Paravia, 1996, p. 90.

⁵⁶ Burnard, 'An Introduction to the Text Encoding Initiative', p. 83.

⁵⁷ J. M. Smith, *SGML and Related Standards: Document description and processing languages*, New York, Ellis Horwood, 1992, p. 15.

⁵⁸ Burnard and Sperberg-McQueen, *Living with the Guidelines*, p. 2.

permissible relationships⁵⁹. SGML is designed for *descriptive* rather than *procedural* mark-up⁶⁰: it does not directly specify how to format or process a document, but describes the document structure; it defines a powerful language for describing [...] structured documents of arbitrary complexity with simple character stream files⁶¹; it declares that a portion of a text stream is a member of a particular class⁶². SGML applies then to documents and describes structures made up of sequentially ordered marked up portions of processable streams of characters. But how much can the structure of a *document* be equated with the structure of a *text*? Only inasmuch as by text we simply mean information coded as a stream of characters. It is only taking this premise for granted, that we can say that texts are composed of discrete content objects⁶³, or, in a more assertive way, that text is best represented as an ordered hierarchy of content objects (OHCO), because that is what text really is⁶⁴. But this definition really prompts

to speculate whether the widely accepted view that documents are hierarchical is a result of deep thought about document structure, or simply a result of years of experience with marked up text⁶⁵.

The assumption, however, that text is essentially made up of an ordered hierarchy of discrete components, cannot always ensure the adequacy of textual analysis. In most cases, the same document conforms to several overlapping structures and indeed humanists involved in computer-assisted textual analysis often need to view a document as conforming to several overlapping structures. But SGML is not designed to accommodate overlapping structural elements, and any kind of non-hierarchical structure constitutes a problem for an SGML-based encoding system for literary texts⁶⁶. SGML is a representation language designed to describe the logical structure of a document⁶⁷, which is by necessity discrete and sequential; but the logical structure determined by sequentially ordered portions of the document, does not necessarily coincide

⁵⁹ S. J. DeRose, D. D. Durand, E. Mylonas, A. H. Renear, 'What is Text, Really?', in *Journal of Computing in Higher Education*, 1:2(1990), p. 12.

⁶⁰ Burnard and Sperberg-McQueen, *Living with the Guidelines*, p. 2.

⁶¹ DeRose et al., 'What Is Text, Really', p. 12.

⁶² Coombs et al., 'Markup Systems', p. 936.

⁶³ R. Cover, N. Duncan and D. T. Barnard, 'The Progress of SGML (Standard Generalized Markup Language): Extracts from a Comprehensive Bibliography', in *Literary and Linguistic Computing*, 6(1991), p. 198.

⁶⁴ DeRose et al., 'What Is Text, Really', p. 3.

⁶⁵ Raymond et al., 'Markup Reconsidered', p. 9.

⁶⁶ D. Barnard, R. Hayter, M. Karababa, G. Logan and J. McFadden, 'SGML-Based Markup for Literary Texts: Two Problems and Some Solutions', in *Computers and the Humanities*, 12(1988), pp. 266-67.

⁶⁷ Smith, *SGML and Related Standards*, p. 15.

with any structure a given interpretation may assign to the text. An essential property of interpretation is creativity, and thus the need for modelling is potentially infinite. But each model of a text can point to different structures⁶⁸ and structure is not always reducible to a functional description of a system's subcomponents. It is impossible, therefore, to reduce any high-level structure to a composition of low-level features⁶⁹ of discrete document components. The structure of the document is the structure of a particular representation of the text, and we should recall once again that the structure of a representation is not the structure of what is represented thereby.

This conclusion, however, seems to be contradicted by mark-up practitioners, who claim that SGML is a data model for *any* kind of data, since it is possible to express any data structure as an SGML-conformant stream of mark-up and data. This alleged contradiction could be easily explained away just by pointing out that for a markup practitioner 'data model' means a common language in which to express the structure of data, whereas for a database practitioner 'data model' means a common language for describing constraints on data and the effect of operations on that data⁷⁰.

And, from the latter point of view, mark-up is not a data model, it is a type of data representation⁷¹. However, by saying that mark-up belongs not to the world of formalisms, but to the world of representations, we do not recount the whole story. SGML-based encoding reduces the structural properties of a text to the structural properties of a document, because it is normally employed as strongly embedded mark-up and its position in the data is information bearing. This means that the properties of the structures which this kind of mark-up can describe are largely derivative of the properties of the document, the linear stream of characters in which it is embedded. By contrast with strongly embedded mark-up, a weakly embedded tag is informative, but its location within the stream of characters is not information bearing. It could be placed at any point in the text, or even outside it and for this reason so-called *out-of-line* mark-up is more properly considered a specific type of external structure. We may therefore wish to distinguish between the internal and external structure⁷² of a 'text', but it should not be overlooked that it is text as data representation, text as a data structure, text as a stream of characters,

⁶⁸ Raymond et al., 'Markup Reconsidered', p. 14.

⁶⁹ Ibid., p. 9.

⁷⁰ D. Raymond, F. Tompa and D. Wood, 'From Data Representation to Data Model: Meta-Semantic Issues in the Evolution of SGML', in *Computer Standards and Interfaces*, 10(1995), [p. 6].

⁷¹ Raymond et al., 'Markup Reconsidered', p. 16.

⁷² Ibid., pp. 3-4.

text as a digital document, text as information represented as coded characters, what we are here referring to. Any linear or non-linear structure can be represented by a linear stream of mark-up and data⁷³, which conforms to the syntactic requirements of a given mark-up standard such as SGML⁷⁴, but a non-linear structure cannot be strongly embedded in a linear stream of characters *representing* a text, or in a 'text' as a linear data representation. Because it shares the data's representation, it is difficult for a strongly embedded kind of mark-up to express structure that is not a subset of characters positions in the text. The order of text, i.e. the order of the linear string of characters, inherited by mark-up is what creates the problem⁷⁵.

III.3 Fluid textuality is mobile and mutable and changes in many ways according to multidimensional patterns. For this reason, a processable representation of textual fluidity requires a non-linear model. The data model for a 'text engine' implementing the *digital textedition* method, described here in Section II, is a non-linear one. Such a model treats 'text' as a non-linear data type⁷⁶, as a general representation of information comprising both running texts, consisting simply of a collection of linearly ordered information strings, and data bases and knowledge representations, consisting of texts connected in a non-linear way. So a specific realisation of this data type need not be a linear structure, and a running text can be described just as a trivial case of a non-linear one⁷⁷. A non-linear *external* structure defined as a general model for the representation of texts, independently of any linear or non-linear realisation, can always be embedded within a running text. Thus, even a linear text can be assigned a non-linear structure, as it would be impossible otherwise, by defining an *internal* structure, generated from within, strongly embedded in its stream of characters and inheriting its sequential order.

But why, from a scholarly point of view, should non-linear data models be hidden or built into 'natural texts'? Why should a non-linear external structure such as a database be embedded or integrated into a running text? This practice, unusual⁷⁸ as it may seem, can indeed recommend

⁷³ Id., 'From Data Representation to Data Model', [p. 6].

⁷⁴ Id., 'Markup Reconsidered', p. 4.

⁷⁵ Ibid., pp. 9-10.

⁷⁶ M. Thaller, 'The Processing of Manuscripts', in Id., ed., *Images and Manuscripts in Historical Computing*, St. Katharinen, Max-Planck-Institut für Geschichte in Kommission bei Scripta Mercaturae Verlag, 1992 (Halbgraue Reihe zur historischen Fachinformatik, A14), p. 55.

⁷⁷ Ibid., p. 57.

⁷⁸ Ibid., pp. 68-70

itself, for it actually reflects the conceptual decisions of a scholar within the context of a specific discipline⁷⁹. Let us consider the case of history. As it is usually understood, a database is an interpretation of the world⁸⁰; it assumes that we already know the world and we have already abstracted a model out of it. But databases providing collections of information, which define their own reality⁸¹, are of limited use in historical research. In general, historians are concerned with a world yet to be discovered. For an historian, the world is not simply there to be understood by means of an abstract model, neither is it already there for a text just to describe it. A text has yet to be interpreted to disclose information about the world. As a matter of fact,

we consider a text to be 'historical', when it describes a situation, where we do neither know for sure, what the situation has been 'in reality', nor according to which rules it has been converted into a written report about reality⁸².

Thus, texts are themselves worlds to be interpreted⁸³ and an interpretation can be seen as an attempt to structure information that has been tradited in a text⁸⁴ by embedding a non-linear database structure into it. Since historical work deals with structures in a text, which we want to *discover*, a text needs to be understood as the formally treatable representation of the current assumptions of a researcher about what his documents actually contain. In historical research, a text forms an intermediate layer between two other layers of information, the purely graphical characteristics of the written document and abstract factual knowledge about the various entities described in a text. The first layer poses a genuine mark-up problem, whereas the second poses the problem of defining the relationship between a 'text' as a running representation of a tradited document and a 'text' as converted into a database according to some abstract model⁸⁵. Embedding a non-linear external structure into a text is therefore an attempt to realise a form of textual representation, which can alternatively but synchronously be interpreted as a running text and as a database⁸⁶. Such a running representation, on the one hand, should not only make it possible to handle variants⁸⁷, considered as

⁷⁹ Id., 'Text as a Data Type', in *ALLC-ACH'96 Conference Abstracts*, University of Bergen, 1996, p. 253.

⁸⁰ Raymond et al., 'Markup Reconsidered', p. 14.

⁸¹ Thaller, 'The Processing of Manuscripts', p. 70.

⁸² Ibid., p. 55.

⁸³ Raymond et al., 'Markup Reconsidered', p. 14.

⁸⁴ Thaller, 'The Processing of Manuscripts', p. 70.

⁸⁵ Ibid., pp. 55-57.

⁸⁶ Ibid., p.60.

⁸⁷ Ibid., pp. 56-57.

debased and lower-rank⁸⁸ alternatives of a given portion of text, but it should also enable us to treat all streams of tradition as equally valid⁸⁹ and all the so called alternative readings as essential parts of the same text⁹⁰; accordingly, it should provide a single and coherent representation of several layers of possibly conflicting streams of tradition, as if it would consist of the logical sum of two or more manuscripts or printed records. On the other hand, each specified portion of the text should contain linkages to representations of knowledge, and to each other portion of the running representation which deals with the same abstract concept, irrespective of the natural language context⁹¹ and the particular expressions used to describe it.

III. 4 It has very appropriately been maintained that the task of encoding the different strata of a manuscript tradition poses formally identical problems with those of encoding in parallel multiple linguistic analyses of a given sentence⁹². But it is only the availability of a non-linear model, connecting a concrete running representation of the text with an abstract and externally structured knowledge representation of its content that can make this assumption true. It is only by embedding external non-linear structures into the running stream of characters *representing* the information content, or the author's understanding and factual knowledge of what is described and expressed through the text, that we can adequately solve the problem of encoding in parallel multiple analyses of a discursive unit, or the different strata of a manuscript tradition.

These two problems arise at two different information layers, that of the written *document* witnessing a text and that of the *text* witnessed or represented by a given document. In this respect, an edition, or a form of textual representation⁹³, can only be adequate if it is expedient to both purposes. The task of a digital edition is therefore best understood as an attempt to match the computational model of text representation and processing with the conceptual model of text reconstruction and textual criticism, on the one side, and, on the other, with the procedures and methods of text analysis and interpretation. Both kinds of textual scholarship, the editorial and the interpretative practices, seem to require a non-linear form of text representation. An editor is faced with a variety of documents witnessing the text, which form a multiplicity of distinct

⁸⁸ Buzzetti, 'Image Processing', p. 149.

⁸⁹ Thaller, 'The Processing of Manuscripts', pp. 56-57.

⁹⁰ Kraft, *Editionsphilologie*, p. 107.

⁹¹ Thaller, 'The Processing of Manuscripts', pp. 56-57.

⁹² Burnard, 'An Introduction to the Text Encoding Initiative', p. 84.

⁹³ Buzzetti, 'Image Processing', p. 148.

sequential representations severally implying different and possibly overlapping hierarchical representations of the text, and has to reduce them to a consistent unity, a unique structural representation of a non-linear kind. The interpreter, on the other hand, is faced with a single sequential representation of textual content, and may derive from it a multiplicity of structural representations, possibly of a non-sequential kind (Fig. 1). But

the editor considers a text representation primarily as a set of data yet to be interpreted concerning material aspects of the text, whereas the interpreter considers a text representation as a set of interpreted data concerning abstract textual information⁹⁴.

Accordingly, the editor has to care about structural representations, both sequential and non-sequential, of the *documents* representing the text, whereas the interpreter has to care about structural representations, both sequential and non-sequential, of the textual *content* represented by a document⁹⁵. These two problems have to be solved at two different levels of textual information, which are precisely the two different layers of information, which have been distinguished before. And it is the edition that has to provide the intermediate layer, a form of text representation connecting the other two, the layer of the concrete graphic characteristics of a written *document*, on the one hand, and the layer of the abstract information contained in a *text*⁹⁶, on the other. It should now be clear that in principle the two problems couldn't be solved at a simple mark-up level, by assigning an internal strongly embedded structure to a running stream of characters. The only way to afford a concomitant solution is to integrate an external non-linear database structure into a running representation of the text.

A digital edition, in conclusion, can only be adequate to its tasks if it provides a combined database and running representation of the text, enabling us to choose between several interpretative environments, or relevant structural representations of factual and conceptual knowledge concerning the information content of the text. An edition is a form of textual representation, but a textual representation of any kind is by no means an edition as long as it is thought of as a sheer duplicate of its source material⁹⁷. It has to be expedient to scholarly research. In the classical form of the printed edition, a single kind of written document was made to serve all purposes of textual scholarship, from text

⁹⁴ Id., *Il testo 'fluido'*, p. 89.

⁹⁵ Id., 'Digital Editions: Variant Readings and Interpretations', in *ALLC-ACH'96 Conference Abstracts*, University of Bergen, 1996, p. 255-56.

⁹⁶ Thaller, 'The Processing of Manuscripts', pp. 55. See above, note 70.

⁹⁷ Buzzetti, 'Image Processing', p. 151.

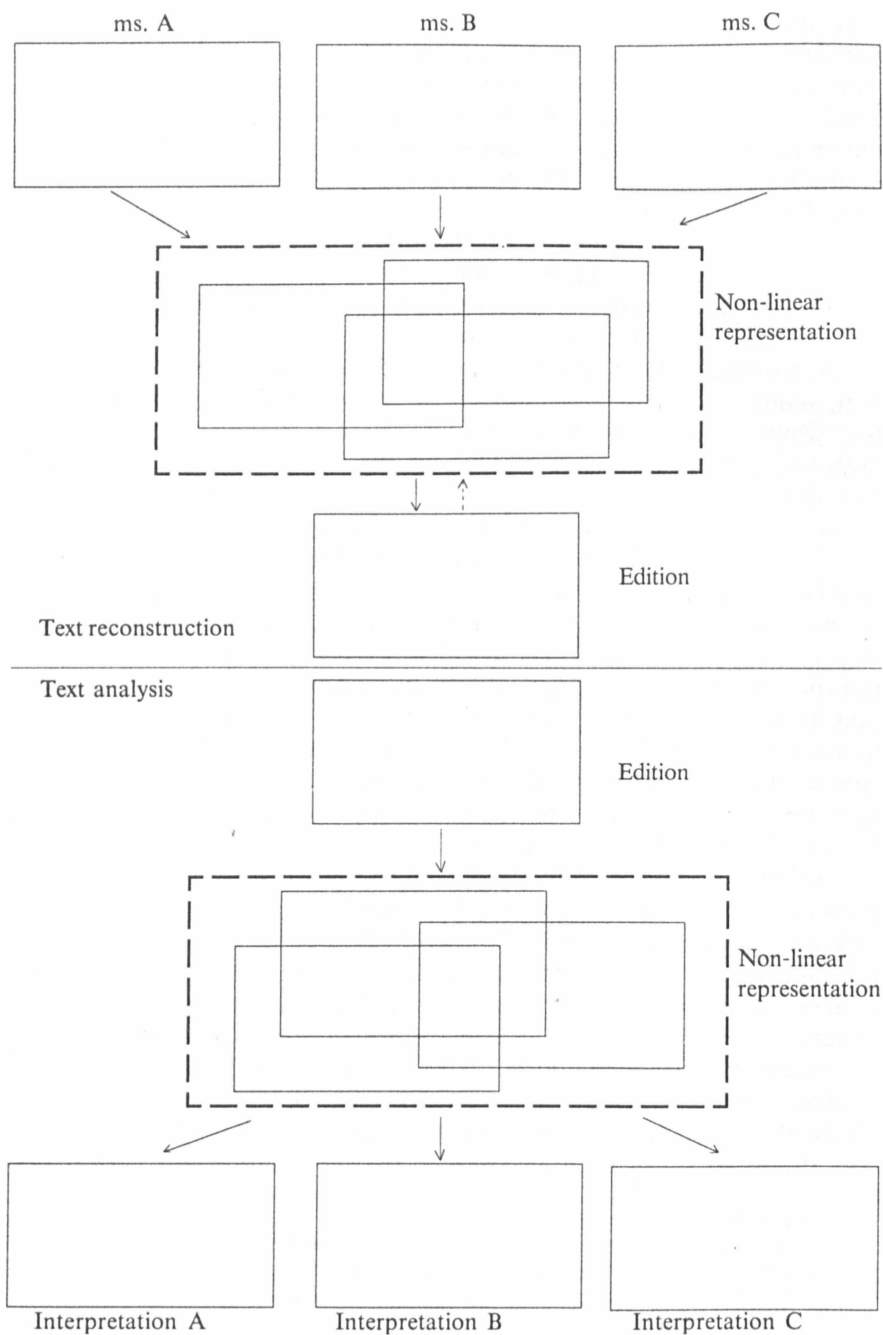


Fig. 1

reconstruction to text interpretation. In this respect, an edition in digital form can only recommend itself if it improves the quality and method of research. But what makes a digital edition essentially different from any other one is its liability to processing⁹⁸ and, in its turn, the quality of processing depends on the structural model of the information content which is embedded in the running representation of the text. The methodological significance of a digital representation is therefore inherent in its structural and logical features, which make sources available as data for further processing and analysis⁹⁹. And only a structured logical representation¹⁰⁰ of textual sources can make it possible to augment the resources open to scholars and increase their options in regard to their every day practice of text interpretation, reconstruction and analysis¹⁰¹.

⁹⁸ Id., 'Digital Editions: Variant Readings and Interpretations', in *ALLC-ACH'96 Conference Abstracts*, University of Bergen, 1996, p. 254.

⁹⁹ Id., 'Image Processing', p. 149.

¹⁰⁰ Ibid., p. 151.

¹⁰¹ Uitti, 'Old French Manuscripts', pp. 157-58.